This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
An Examination of the Problems of False Consciousness

Lilith W. Lee

Doctoral Thesis in Philosophy
The University of Edinburgh
2021
To my grandparents, 嬤嬤、公公、嬤嬤、爺爺:

我很想念你們。對不起, 請原諒我。
Lay Summary

The concept of false consciousness has been used over the years in philosophy and public discourse to characterise a prima facie puzzling social phenomenon: oppressed individuals sometimes hold sets of ignorant beliefs that motivate them to act in ways to contribute to their own oppression. These beliefs are often thought to arise out of their very situation of oppression. In this thesis, for example, I pay close attention to the false consciousness of women anti-suffragists in fin-de-siècle Britain: women who were actively campaigning against suffrage, as well as published anti-suffragist works, with a great degree of sophistication. Although my case study here is a historical one, it resonates in many ways with women anti-feminism in the US and Europe today.

Apart from the oppression of white women, the traditionally Marxian concept has been used to analyse a wide range of systems of oppression (for example, ablelist, anti-Black, class-based). Yet, the concept of false consciousness is often missing from the standard analytic social philosopher’s toolkit. This is because the concept is thought to suffer from three kinds of problems intrinsic to it. First, it is thought to be explanatorily problematic. The concept of false consciousness seems to be unable account for cases where an individual is located within multiple, or a complex of, systems of oppression. Second, it is thought to be normatively problematic. The concept requires a normative standard to judge a given consciousness as ‘false,’ but it seems as though any standard that can be guaranteed to be accepted by someone criticised for being under false consciousness cannot itself be guaranteed to be not an instance of false consciousness. Third, the concept is thought to be agentially problematic. The concept of false consciousness seems to characterise oppressed individuals in ways that constitutes victim-blaming.
As a result of these problems, philosophical attention to the concept of false consciousness has been scant. This, however, is unfortunate because the concept is still being used and abused both in academic philosophy and public discourse. The concept, for example, has been used by incels to characterise men who perpetuate their own sexual frustration in supporting feminist or queer movements. The lack of clarity that we have with regard to the concept of false consciousness, as well as its absence from the social philosopher’s toolkit, has also led to an impoverishment in understanding related concepts like epistemic standpoints.

This thesis, therefore, is an attempt to clarify, defend, and rehabilitate the concept of false consciousness, in order to expand the social philosopher’s toolkit and refine our understanding of social criticism both in philosophy and public discourse.
Abstract

The concept of false consciousness is understood to involve individuals who act in ways that contribute to their own oppression, on the basis of ignorant beliefs that resist revision when confronted with attempts at correction. And these beliefs are themselves thought to arise from the individual’s oppressed conditions. This thesis is an attempt to rehabilitate the concept back into the toolkit of analytic social philosophy against its detractors. I thus examine the common problems associated with using this concept in social philosophy, using the case of women anti-suffragists as a central case study.

I begin with a brief genealogy of the concept of false consciousness, which attends to the history of the concept as it developed within the Marxian and Critical Theory traditions, before making a few preliminary clarifications about how the concept will be defined and used here. I then explain several challenges to using the concept of false consciousness, suggesting that detractors of the concept often rely on characterisations of it that sit uneasily between our emancipatory goals and certain extant approaches to social epistemology and ontology, normativity, and moral blameworthiness.

The concept of false consciousness is often criticised as an ad hoc explanation for why someone acts a particular way that is contrary to their own interests. There is no standard account for why someone has false consciousness especially under multiple systems of oppression. I address this by proposing an approach that treats the concept as a means to regulate social research and direct attention to various looping mechanisms that sustain oppressive beliefs and practices. These mechanisms converge upon an individual to produce beliefs that may be shared with other oppressed agents when seen from the perspective of a system of oppression. Yet at the same time, such beliefs are nuanced differently at more agent-targeted levels and in relation to other systems.
Another problem with false consciousness is in determining the appropriate standard for employing the concept in critique. Standards of critique are themselves often the very things in dispute. The problem thus lies in needing to provide a standard of critique that can avoid such disagreement and also ensure that the standards themselves are not products of false consciousness. I address this by appealing to the notion of well-being, which would generate standards that are by definition the agent's own. Extant accounts of well-being, however, are inadequate for the task. I thus outline my own account, grounded in our status as sense-making organisms and second-order norms regulating our organismic processes.

False consciousness is also often thought to involve victim-blaming, since the individuals are thought to be irrational or not to know any better. Using the concept to pick out the individual as contributing to the very oppression they experience thus seems to be inappropriate as irrationality and ignorance are exculpatory in many other cases in ethics. I address this by focusing on a historical case study of several women anti-suffragists. I first examine the charges of irrationality and find that the charges do not hold for them. I then consider the reasons behind their systematic ignorance and find that they are indeed brought about by culpable mechanisms. In turn, the women are also culpable for the acts that follow from them. Hence, false consciousness does not preclude blameworthiness.

I conclude with a brief observation of a tension between moral criticism—wherein agents are criticised as individuals—and social criticism (critique)—wherein individuals are criticised in terms of social categories. I suggest that, although the concept of false consciousness can overcome the above conceptual problems, it nevertheless appears to have some limited use in relation to our blaming practices.
This thesis, therefore, is an attempt to clarify, defend, and rehabilitate the concept of false consciousness, in order to expand the social philosopher’s toolkit and refine our understanding of social criticism both in philosophy and public discourse.
## Table of Contents

Lay Summary ii  
Abstract iv  
Table of Contents vii  
Acknowledgements ix  
Preface 1  

I. The Problems of False Consciousness 4  
1. A Brief Conceptual Genealogy 12  
   1.1. Locus Classicus 16  
   1.2. Hindrance to Revolution 20  
   1.3. Distorted Relation to Reality 25  
   1.4. Socially Necessary Product 31  
2. Preliminary Clarifications 37  
   2.1. Ignorance of the Basis of One’s Beliefs 38  
   2.2. Ideology and False Consciousness 41  
   2.3. The Doxastic Form 49  
3. The Problems 54  
   3.1. The Explanatory Problem 55  
   3.2. The Normative Problem 58  
   3.3. The Agential Problem 61  
4. Approaching the Problems 62  
   4.1. Overview of Arguments 62  
   4.2. Methodological Note 65  

II. Why Do They Know Not What They Do? 69  
1. Explanatory Desiderata 71  
2. Survey of Extant Approaches 76  
   2.1. Functionalism 77  
   2.2. Psychologism 80  
   2.3. Processualism 82  
3. Social Constructionism 86  
   3.1. Interpellation 87  
   3.2. Recognition 90
Acknowledgements

This thesis is very much indebted to the support and encouragement, both academic and otherwise, from the following:

My supervisors, Matthew Chrisman, Elinor Mason, and Filipa Melo Lopes, each of whose steadfast support, patient guidance, and critical insights have often been unsparing and supererogatory, despite my stubborn apprehensions and tendency to get ensorcelled by my other research interests.

The good folk at the various conferences, workshops, and summer schools that I had attended over the course of this thesis’ development. Many of whom had unfortunately been subjected to the alpha and beta testings of earlier versions of the chapters here, but all of whom had been a wellspring of wisdom and merriment. In particular, I am grateful for the generous instruction and comments from Michael Rosen, Sally Haslanger, Gideon Rosen, Jason Stanley, and Guy Fletcher, as well as the anonymous reviewers from the *Journal of Social Ontology* and *Hypatia*.

Commenters on various parts (and embryonic versions) of this chimeric thesis for their invaluable criticisms: Dave Ward, Patrick Todd, Chong-Ming Lim, Kate Nave (who was particularly pivotal in pushing me to IRL case studies), Emma Husband, Jenny Zhang, Jodie Russell, Huzan Bharucha, Alessandro Barbieri, Angie O’Sullivan, and Joshua Tan.

My examiners, Tommy Curry and Katharine Jenkins, for their generosity in attending my work and, above all, their patient and upbuilding criticisms.

All remaining shortcomings of this thesis are mine alone.
I would also not have kept afloat amid the merciless undertows of the academic industry and the frigid, unrelenting whirlpools of doubt and depression, if not for the following lifelines (some of whom have already been mentioned above): the EPiPHENy reading group; the Discord servers, Together Isolation (especially Wen) and bored gamez; the dependable banter of BAW; the riotous PhD students here (especially da kewlest kidz, Tony & Lorenzo, Mara, Matt, and John); my fellow urban cavorters, Lea and Masha; the wild Viewcraig, Spottiswoode, and East Parkside, including Adèle, Zo, Kate, (Peggy,) and Gongzie; the feline comforts of Sparkle, Eliot, and Trixie; the assuring digital presences of Grace, Tanya, Vikki, and Rei; the patient hospitality of Aleks & Carren; the loving friendship of Jins; and the long-suffering of my family (especially my sister). Many of them may never learn how often they have saved me from drowning and provided a warmth I do not deserve.

My studies would not also have been possible without the generosity of the following funding bodies: the Tan Kah Kee Foundation, for the Tan Ean Kiam Postgraduate Scholarship in the Humanities, 2017–2018; The University of Edinburgh for the CAHSS/PPLS Research Award and Edinburgh Global Research Scholarship, 2017–2020; and the Society for Applied Philosophy, for their Doctoral Scholarship, 2020–2021.
Preface

This thesis is an attempt to rehabilitate the concept of false consciousness back into the toolkit of analytic social philosophy. Despite its prolific use in the 1970s and 80s, the concept of false consciousness has largely fallen out of fashion in the analytic mainstream since, due to a number of problems now commonly thought to be intrinsic to the concept. For this rehabilitation, I draw on resources in the emerging fields of critical social ontology and enactivism to address these problems, taking up the fin-de-siècle British anti-suffragists as a central case study.

In Chapter I, I provide a brief genealogy of the concept of false consciousness, which attends to the history of the concept as it developed within the Marxian and Critical Theory traditions, before making a few preliminary clarifications about how the concept will be defined and used here. I then enumerate and motivate the problems of false consciousness, suggesting that detractors of the concept often rely on characterisations of it that sit uneasily between our emancipatory goals and certain extant approaches to social ontology and epistemology, normativity, and moral blameworthiness. However, in Chapter II, I argue for a general theory of the concept that would allow us to avoid such characterisations. I do so by critiquing and subsuming existing attempts at a general theory of false consciousness under a social constructionist approach inspired by Louis Althusser’s discussion of ideology.

In Chapter III, I attend to considerations of well-being for the normative basis of the concept of false consciousness. I argue that, in order to avoid methodological worries about paternalism and backfire effects that alienate
those criticised for false consciousness from our emancipatory projects, we should abandon extant realist and anti-realist approaches to well-being. In their place, I argue for an immanent approach. That is, unlike these extant approaches that determine what is ultimately good or bad for any welfare subject in terms of an attitude-independent fact(s) or in terms of a fact(s) about the attitudes of some agent, respectively, the immanent approach determines what is ultimately good or bad for a particular welfare agent only after empirically-based research into the welfare subject’s constitution in social reality. The immanent approach does so by being enactive and critical: it is grounded in our shared status as self-enacting organisms that engage in participatory sense-making and it eschews relying on normative standards not already held by the agent, developing instead a critique of them from these very standards themselves.

Chapter IV looks at the assumption that moral blameworthiness is precluded under false consciousness. Agents acting under false consciousness are assumed to systematically fail the relevant rationality and epistemic conditions, due to structural distortions of reasoning or knowledge practices, undermining their status as responsible moral agents. But attending to the constitutive mechanisms and heterogeneity of false consciousness allows us to see how merely suffering false consciousness does not eo ipso render someone an inappropriate target of blame. I focus specifically on the circumstances surrounding the anti-suffragist manifesto “An Appeal Against Female Suffrage,” arguing that its signatories, despite false consciousness, satisfy both conditions for ordinary blameworthiness.

With Chapter V, I conclude by briefly suggesting an apparent methodological limitation with the use of the concept of false consciousness. While the problem is not intrinsic to the concept, it nevertheless highlights a tension in its use within our everyday practices of moral accountability. Whereas moral criticism addresses agents as individuals, social criticism (critique)
addresses agents in terms of social categories. There is an apparent incompatibility in having to adopt both the objective and participant attitude to the target in the selfsame criticism. That is, blaming an agent under false consciousness seems to involve relating to them as both an object of social theory (where certain reactive attitudes such as resentment are suspended) and, at the same time, a subject who is a fellow in one's moral community (where reactive attitudes such as resentment are characteristically constitutive of our communal relations). The application of the concept of false consciousness in moral blame by the critic also redoubles oppression in theoretically fixating the individual under an oppressed social category in an asymmetrical economy of representation. Blaming an agent under false consciousness thus seems to be self-frustrating for the aims of both moral and social criticisms.
I. The Problems of False Consciousness

CHAPTER ABSTRACT

Even though the concept of false consciousness has been used by social philosophers and commentators alike, theoretical attention to the concept itself has been scant. As such, it is not always clear what exactly the concept is supposed to involve. In this chapter, I briefly consider the uses of the concept and how it fell out of fashion in analytic social philosophy, before tracing the genealogy of the concept through its various shifts and accruals in meaning in the Marxian and Critical Theory traditions. I consider the concept as it develops from Marx and Engels, Lenin and Gramsci, Lukács, and the early Frankfurt School, before arriving at an adaptation of Geuss’ classic definition of false consciousness: false consciousness is a set of ignorant beliefs held by an individual that occurs in virtue of, and motivates them to perpetuate, structural oppression. I then argue against two prominent alternative definitions of false consciousness: namely, false consciousness as ignorance of the basis of one’s beliefs and false consciousness as being identical to ideological consciousness. These, I suggest, are both motivated by unwarranted worries of methodological individualism and the correspondence theory of truth—assumptions that are neither essential to the concept nor characteristic of the way it has been understood within the Marxian and Critical Theory traditions. I then enumerate the problems of false consciousness that are commonly thought to render it unusable as a concept, which will be the concern of the thesis as a whole. I conclude with a brief summary of the approach of the thesis, as well as a methodological note on my use of the case study of fin-de-siècle British women anti-suffragists.
Alt- or far-right politics is often associated with various forms of misogyny and anti-feminism, perhaps seen most prominently in the occasions of incel-related violence over the past few years (Nagle 2017). Still, despite this association, women were estimated in 2016 to have constituted at least 20% of the movement in the US (Hawley 2016). It has also generally been observed that women are increasingly drawn to such parties in Europe (see Köttig, Bitzan, Petö 2017): ranging from 35.7% of the gender vote share for the Alternative für Deutschland (AfD), co-led by Alice Weidel, during the 2017 German federal elections (Glaser et al. 2018, 29); to 39.7% for the Prawo i Sprawiedliwość (PiS), led by Beata Szydło, during the 2015 Polish parliamentary elections (Grzebalska and Zacharenko 2018, 87). Yet despite this trend, these movements and parties have often been deleterious for women’s rights and interests. The AfD’s political programme explicitly lists, as one of its aims, “Terminate the Promotion of Gender Research” (Alternative für Deutschland 2016, 51). In 2016, Szydło’s government put forward legislation to ban abortions under any circumstance, and, in October 2020, the PiS (under Mateusz Morawiecki) successfully banned abortions on the basis of foetal defects through judges that the party itself appointed to the Constitutional Tribunal (Walker and Strek 2020).

Of course, the recent surges towards such politics by women is not monolithic and many in these movements subscribe to such political programmes to varying degrees. Yet, from the AfD to the Honey Badger Brigade (a largely North American women’s group that publicly supports the men’s rights movement), these movements nevertheless tend to broadly

---

1 Of course, the movement, with its euro-centric associations, is also known for its racist misandry.

2 A seeming anomaly might be Marine Le Pen's leadership of the Front National that has been recently campaigning on ‘feminist’ aims, turning out almost equal numbers between men and women during the 2012 French presidential elections (Mayer 2015). However, this is often attributed to Le Pen coopting particular feminist narratives for the sake of ethno-nationalist goals—a trend common for western European states (Farris 2017).
converge on the idea that the socio-economic and political positions of women ought to be circumscribed by their natural, biological differences from men. As alt-right author Cecilia Davenport remarks in an interview with The Economist:

There are a variety of views about [women’s place within the movement] on the alt-right; it’s important to note that not everyone has exactly the same views on the value and role of women in society. I think most people on the alt-right, men and women, agree that the sexes are morally equal but materially different, and further, that we aren’t just ghosts in a machine, that our bodies are coextensive or incarnate with our identities. Men and women are made to fulfill separate yet complementary roles. Biologically speaking, just like we assert that race is real, sex is real: and is, in fact, the most fundamental distinction in humanity. I’ve never met a man on the alt-right who wanted less for their women than happiness and fulfillment—but we believe that feminism does not make women happy or fulfill them. […]

I think most men and women on the alt-right agree that a woman’s unique ability to have children and affinity for child-rearing is nature’s highest role. At the very least, alt-righters want women to have the option to stay home and raise a family if that’s what they want to do. This is not only unpopular in modern society, but often mocked as a valid choice. […]

As for female empowerment, there’s nothing that has made me feel more empowered in my life than supporting and being supported by a strong man. I think that men and women are better off when we stop fighting nature and allow our distinct identities to shine through, working together as a team. Again: just like race is real, biology is real. Why do so many fight it?

As Richard Spencer and others have explained, it’s true that women are less likely to join a vanguard political movement than men are. That said, I don’t actually think there are “so few women” in the alt-right. Personally, I know of at least a few hundred. Think of it this way: just as the polls in England couldn’t encompass the Brexit vote, and just as the polls in America couldn’t pick up Trump voters, the usual methods for looking for women on the alt-right don’t work. (Davenport 2017)
Further along this line, other women pundits associated with the alt-right, like Ann Coulter and Janet Bloomfield, have even argued that women’s suffrage was a mistake and that women today should be disenfranchised. Bloomfield, for example, argues as follows:

Recall again that women cannot be drafted. Blood and boots will be spilled to solve the refugee crisis. It is all but guaranteed, but it won’t be women’s blood, and it won’t be women’s feet in those boots, so why then should they get to decide how the future proceeds?

Women have had the vote in the West for almost 100 years, and all they have done is vote to destroy and destabilize the world men built for us, while protecting themselves from the blood consequences. They have voted selfishly, rapaciously, irrationally and quite possibly, irrevocably. Women should not vote. That’s not misogyny. It’s self-defence. (Bloomfield 2017)

These arguments for limiting the political participation of women echo those put forward by the hundred and four women anti-suffragists of the anti-suffragist manifesto “An Appeal Against Female Suffrage” (henceforth ‘the Appeal’) in fin-de-siècle Britain. The Appeal argued that, in virtue of their physical constitution, “the necessary and normal experience of women […] do not and can never provide them with such materials for sound [political] judgment as are open to men” (“An Appeal” 1889, 782). The manifesto was published in June 1889 in the Nineteenth Century, a mainstream periodical that published debates of leading intellectuals at the time, with a number of unspecified signatories who were largely upper-class women. The Appeal marked a watershed, women-led response to the purported threat of suffragism at the time and there is a general consensus among historians that it did not simply express the widespread social conservatism of the time but contributed significantly to maintaining the status quo of the

---

3 Bloomfield is often associated with the Honey Badger Brigade and had been the head of social media for the men’s rights website A Voice for Men.
parliamentary exclusion of women—until their eventual enfranchisement after the First World War, as well as the profound shifts that it brought to the British workforce (Harrison 1978, Sutherland 1990, Faraut 2003, Bush 2007).

How the effects of today’s women anti-feminism compare to fin-de-siècle women anti-suffragism remains to be seen, but cases such as these are regardless baffling at first glance for those outside of such circles, for they involve agents systematically believing themselves unqualifiedly justified in acting in ways that contribute to their own oppression (such as campaigning or voting). And, as we can clearly see, such cases have been around for centuries (see also Superson 1993). Many social philosophers and theorists have used (or are said to have used) the concept of false consciousness—whether explicitly or tacitly—in their particular analyses of such phenomena (see, for example, Mill 1869, Rowe 1975, Cudd 2006). This concept, of Marxian origins, is generally used to understand the recidivistic beliefs of self-oppressing agents in various (often overlapping and interwoven) systems of oppression. That is, these beliefs are thought to form in virtue of the agents’ oppressed status and would persist in spite of counter-evidence (in the form of either resistant discourse or the harms of their oppression).

The concept of false consciousness has been used to analyse a wide range of social phenomena by many social philosophers and theorists, from the aforementioned anti- or apolitical stance of Victorian women (Cudd 2006), to disabled people who refuse to politically identify as impaired or disabled (Shakespeare and Watson 2001), to the attributions of poverty and wealth to

---

4 Historians Brian Harrison and Julia Bush, for instance, note that then Prime Minister William Gladstone’s opposition to the Bill for Extending the Parliamentary Suffrage to Women, read in the House of Commons in 1892, cited support from “large numbers of women who have considered the matter for themselves, the most positive objection and strong disapprobation” (Gladstone 1892). Harrison and Bush also observe that the Appeal “fail[ing] to produce any continuing movement against woman suffrage” notwithstanding, sustained the status quo and was later instrumental to the formation of the Women’s National Anti-Suffrage League in 1908 (Harrison 1978, 117–118; Bush 2007).
personal irresponsibility and lack of hard work or talent, among low-income racial/ethnic minority and immigrant women in the US (Godfrey and Wolf 2016). Uses of the concept extend even beyond discussions in social and political philosophy to philosophy of science and epistemology (see, for example, van Fraassen 1996, Spurrett 2017, Srinivasan 2020), as well as beyond the arcane halls of the academe to public discourse (see Hayden 2017, Gordon 2017, Srinivasan 2018a). In 2019, Vox published a video on their YouTube channel entitled, “Why Tucker Carlson pretends to hate elites,” which explicitly employs the concept of false consciousness in its exposition of the effects of the Fox News anchor’s performance of anti-elitism on his viewers (who are mostly much farther down the socio-economic ladder than him).\(^5\) Curiously, the concept of false consciousness (specifically, as “sexual false consciousness”) has even been invoked among incels and in the broader manosphere, in relation to the alleged sexual disenfranchisement of men—although more often than not the relevant term used would be being ‘blue-pilled’ (Kshatriya 2015, 2019).\(^6\)

Perhaps as a result of such uses and abuses of the concept of false consciousness, its place in the analytic social philosopher’s toolkit has just as long been a matter of controversy, if not detraction. This thesis is an attempt to rehabilitate the concept back into this toolkit against its detractors. Within the analytic mainstream in academic philosophy, theoretical attention to the concept itself is largely restricted to discussions in the 1970s and 80s—specifically, in relation to the heyday of the Analytical Marxists (see, for examples, Cohen 1979, Elster 1983) and second-wave feminists who drew inspirations for feminist methodologies from Marxism (see, for examples, Barkty 1975, MacKinnon 1989). However, the attention

---

\(^5\) At the time of submission (November 21, 2021), the video has approximately 2.9 million views.

\(^6\) The incel metaphors of being ‘blue-pilled’ (subscribing to feminist ideas, broadly speaking) or ‘red-pilled’ (subscribing to incel ideas) comes from *The Matrix* (1999) by the Wachowski Sisters.
on false consciousness as a key concept for social philosophy, as I will elaborate later, was met with criticism not only from non-Marxist and non-feminist theorists (see, for example, McCarney 1980, Thompson 1984, Balkin 1998), but also (post-)Marxist and feminist ones (see, for example, Hirst 1979, Mouffe 1979, Stanley and Wise 1993, Pateman 1999).\(^7\) As a result of this, the concept of false consciousness fell out fashion in social philosophy.\(^8\)

Nevertheless, some of the social phenomena picked out by the concept of false consciousness from these debates have crucially informed subsequent approaches to social and political philosophy in the analytic mainstream. Perhaps most prominently, Amartya Sen’s influential criticism of preference-based economics and political philosophy (which took for granted the preferences of individuals as the normative basis for their frameworks) drew explicitly on some general notion of false consciousness (Sen 2009).\(^9\) Sen understood ‘false consciousness’ in terms of the adaptation an individual’s preferences, desires, and beliefs to their oppressed conditions. Sen’s development of the Capability Approach (alongside others like Martha Nussbaum) then, which focused instead on an individual’s capabilities to achieve well-being, was thus an attempt to provide a normative scaffold for economics and political philosophy that would avoid the theoretical sinkhole of cases of false consciousness. But as David Enoch recently observes, this approach, alongside many others in other subfields of social and political philosophy, decidedly distance themselves from the Marxian association, opting instead to understand such phenomena in broadly similar terms like “deformed desires,” “adaptive preferences,” being ‘cultural dupes,’ or simply

---

\(^7\) For a discussion of the relationship between Marxism and feminism more generally, see Haug 2016.

\(^8\) In fact, some of these issues were already in discussion before this period (see Runciman 1969).

\(^9\) Sen refers to the notion explicitly in *The Idea of Justice* (2009) as “what [Marx] called ‘false consciousness’,” even though, as I shall observe in §1.1, Marx did not once use the term (Sen 2009, 164).
‘internalised oppression,’ which “[sound] perhaps a touch more historically kosher” (Enoch 2020, 201n93; cf. Jaggar 2005, Superson 2005, Harvey 2010, Khader 2011, Chambers 2017, Mason and Wilson 2017). The presently scattered theoretical attention to the concept of false consciousness itself strongly suggests that the concept’s detraction remains the default stance for a large part of academic philosophy. These almost always either set out to defend the concept from detractors, without much success in uptake by other philosophers (see, for example, Meyerson 1991, Engelstad 2016), or are essentially throwaway, undeveloped references in marginalia (see, for example, Rawls 1996, 68–69n21; Jaeggi 2009, 82–83n17; Haslanger 2017a, 9n8).

But many today are worried about the costs incurred by the lack of philosophical attention to the concept, especially in relation to the state of contemporary politics today. One example would be a lack of clarity about what the conditions of appropriate use of the concept are, which would allow us to discriminate between, say, how feminists and the manosphere have been using it. Another example can be found in a 2018 interview with Amia Srinivasan, where she makes a passing lament about the problems with the current, popular (and caricatured) understanding of standpoint theory (where “each person has an individual authority about either the way the world is or their own personal experiences”):

[Y]ou have to figure false consciousness into this, which is politically totally out of fashion. I think the fact that false consciousness is no longer something we’re allowed to take seriously in leftist politics, and especially feminist politics, has meant that standpoint epistemology gets distorted—because

---

10 To note, the relation between false consciousness and adaptive preferences is extremely varied in the literature: some understand adaptive preferences to be a conative form of false consciousness (Deveaux 2002, Jaggar 2005, Enoch 2020), others understand it to be a possible constituent mechanism (Elster 1980, M. Rosen 1996), and still others think they are distinct concepts (Khader 2011). In this paper, I will consider adaptive preferences in terms of the first two.
then it should just follow that it’s incredibly easy for individuals to know. (Srinivasan 2018b)

There is thus a clear need, both philosophically and politically, to revisit the concept of false consciousness and examine its associated problems in light of some recent developments in the philosophical literature—beyond the narrower disciplinary confines of analytic philosophy of the 80s. This thesis is an attempt to meet this need for conceptual clarity with emancipatory practice in mind, taking into account the recent scholarship on social epistemology, critical social ontology, enactivism, and moral responsibility. To anticipate, the examination of the concept here will be a positive one—that is, I aim here to address the problems of false consciousness so as to rehabilitate the concept back into the toolkit of analytic social philosophy.

In this chapter, I will first begin with a brief genealogy of the concept of false consciousness as it is understood within the Marxian and subsequent Critical Theory traditions (§1). I will then make a few preliminary clarifications about the theoretical motivations and parameters of the present inquiry into the concept (§2), before enumerating the problems commonly thought to be intrinsic it since its inception (§3). I conclude with a brief overview of how these problems will be approached in the subsequent chapters (§4).

1. A Brief Conceptual Genealogy

False consciousness will be defined in this thesis to involve a threefold ‘falseness’: epistemically, functionally, and genetically (Geuss 1981, Williams

---

11 Standpoint theorists would hold, rather, that such knowledge is a matter of achievement rather than a given (see, e.g., Collins 1989).
That is, false consciousness is defined as a set of ignorant beliefs held by an individual, functioning to motivate actions that would perpetuate their own oppression, and caused by epistemic practices adversely conditioned by structural oppression. In this section, I provide a brief explanatory genealogy of the concept, so defined, that is neither vindicatory nor subversive (cf. Williams 2002). That is, without making evaluations, I trace the concept of false consciousness and its incipient forms through several key Marxian philosophers that have been often seen to contribute to its development, “describing the theoretical turning points in the historical shaping of [the] philosophical concept through time,” so as to “[reveal] assumptions underlying the concept in question that are often tacitly and uncritically accepted, thereby widening the space of theoretical possibilities” (Dutilh Novaes 2016, 99).

The aim of the genealogy here is ultimately to arrive at the above definition of false consciousness over and against a couple of narrower definitions scattered throughout the analytic philosophical literature, which often defines the concept only from a very restricted reading of the Marxian traditions for their own purposes (cf. Wood 1988, Barnes 1997, Shelby 2003). There are two alternative definitions that are most prominent. The first is where false consciousness is understood as merely being ignorant of the basis of one’s beliefs. That is, an individual is under false consciousness inasmuch as they assume that the origin of their beliefs is rational when it is actually irrational. The second is where false consciousness is understood as the acceptance of ideology simpliciter. That is, to be under false consciousness is simply to be under an ideology, such that no independent

---

12 This draws on Raymond Geuss’ formulation of the notion of ‘ideologically false consciousness’ of the later Frankfurt School, focusing on Jürgen Habermas (see Geuss 1981).
13 The genealogy is not straightforwardly a ‘pragmatic genealogy’ either, since it attends to the development of the concept more within how it has been used for philosophical purposes than within their immediate socio-historical context—although these are, of course, inseparable, especially for the Marxian and Critical Theory traditions (cf. Queloz 2021).
inquiry into the concept of false consciousness is necessary given the extant work on ideology. I will be attending to these alternative definitions in the following section and highlighting the important ways in which the use to which I want to put the concept exceeds the contours of these definitions (§2).

The genealogy will begin, expectedly, with Karl Marx and Friedrich Engels (§1.1), followed by several turning points in the concept: as a hindrance to revolutionary change, with Vladimir Lenin and Antonio Gramsci (§1.2); as a distorted relation to social reality, with György Lukács (§1.3); and as a necessary product of capitalist society, with the (early) Frankfurt School (in particular, Theodor Adorno, Max Horkheimer, and Herbert Marcuse) (§1.4). These philosophers are often mentioned (usually in passing) in the analytic philosophical attention to false consciousness (see, for example, Runciman 1969, Geuss 1981, Meyerson 1991, M. Rosen 1996, Williams 2002, Medina 2013).

For the purposes of merely arriving at the aforementioned working definition for this thesis, my brief genealogy will not be evaluating the claims of these philosophers, much less be engaging much with the historical scholarship or various reconstructive projects on them (such as attempts to reconstruct X's concept of false consciousness, where lexical or functional equivalents are lacking).

---

14 Similar genealogies related to the concept of false consciousness have been attempted by Ron Eyerman (1981) and Michael Rosen (1996), but their focus is on distinguishing the concepts of ideology and false consciousness, as well as how the former is meant to be explanatory for the latter.

15 There are other Marxian philosophers whose contributions have been unfortunately (and, to my mind, unjustly) sidelined, especially in the anglophone academy, such as Henri Lefebvre's notion of mystified consciousness [conscience mystifiée] (Guterman and Lefebvre 1936). To note, as a result of Louis Althusser's lack of engagement with Lefebvre's works, the notion of mystified consciousness has also received minimal attention in the non-anglophone attention to the concept of false consciousness.

16 For such evaluative and reconstructive projects pertaining to false consciousness, see, for example, Pines 1993, M. Rosen 1996. For examples of ‘reconstructive projects’ per se outside of Marxian scholarship, such as the concept of rights in Aristotle or the concept of democracy in Confucianism, see Miller 1995, Tan 2004, Chan 2014.
As a historical preamble, some early traces of the concept in the Anglo-European philosophical traditions has been said to be found all the way back in Plato's *Republic*, with Socrates’ less-observed suspicion of the transactional morality of the rich and powerful and Thrasymachus’ subsequent, infamous challenge, as well as the notorious Noble Lie (see Myers 2002, Luco 2016).\(^{17}\) Philosophers and social theorists have also attributed a Renaissance formulation of the problem of false consciousness to Étienne de la Boétie’s question on voluntary servitude: “Why, when the balance of force is so obviously on their side, do the mass of men submit even to the rule of tyrants?” (M. Rosen 1996, 62; Lukes 2011). This more explicitly political lineage has been seen to thread through the early modern period as “the fundamental problem of political philosophy” (Deleuze and Guattari 1972, 29; cf. M. Rosen 1996). However, the concept of false consciousness, as we understand it today, really crystallised in the Marxian and subsequent Critical Theory traditions.

It should also be noted that many social philosophers acknowledged in the Anglo-European philosophical traditions—from Charlotte Perkins Gilman and W. E. B. Du Bois, to Simone de Beauvoir and Frantz Fanon—have developed similar concepts for their own analyses and critiques (for example, ‘sex-consciousness’ or ‘double consciousness’).\(^{18}\) But these have nevertheless drawn from, if not developed alongside, the Marxian traditions—and the boundaries of traditions are after all often blurry, if not arbitrary. For the

\(^{17}\) If we are to take on board Mario Wenning (2011) and Dorothy Kwek’s (2019) readings of the *Zhuangzi* (c. 3rd century BCE) as an instance of critical theory, then we could also include its indictment of the man-made daos of the Confucians as an ancient instance of a critique of false consciousness. There are, doubtless, other instances outside of the confines of the Anglo-European canons in academic philosophy (such as the concept of ‘indolence’ in José Rizal’s (1890) *The Indolence of the Filipino*).

\(^{18}\) Gilman, for example, attends to how fashion contributed to a “sex-consciousness” or “clothes-consciousness” in girls that “hamper her all her later life” (Gilman 1915, 135). For discussions on the relation between the concepts of false consciousness and double consciousness, see, e.g. Gooding-Williams 2009, McWeeny 2016.
purposes of this introduction, then, I will restrict the focus of the genealogy to the exclusively Marxian traditions here.

1.1. *Locus Classicus*

It might come as a surprise to some that Marx himself never used the term ‘false consciousness [falsches Bewusstsein]’ throughout his writings. Rather, in his works, the concepts that more recent scholars use to triangulate some tacit concept of false consciousness are ‘ideology,’ ‘superstructure,’ and ‘fetishism’ (see Wood 1988, Pines 1993, M. Rosen 1996). Marx’s remarks on ideology are often sourced from the unpublished manuscripts of *The German Ideology* (1846), in which he and Engels appropriated the term to disparage idealist philosophers, taking the term in turn from Napoleon’s own disparaging appropriation from the Condillacian *idéologistes*, who simply understood it to mean “a theory of ideas” (see Mannheim 1936, 63n1). With respect to the concept of superstructure, as in the base-superstructure relation (see Cohen 1979, Mills 1992), discussions often centre on *A Contribution to the Critique of Political Economy* (1859), where the term refers to legal and political institutions and practices “to which correspond definite forms of social consciousness” (MECW 29: 263). And the concept of fetishism is found in *Capital, Volume 1* (1867), which more narrowly concerns how the social character of a commodity and its producers’ relations to each other are masked under the capitalist mode of production. Marx’s shifts in emphasising these concepts tracked corresponding shifts in emphases in his thought in relation to his interlocutors: from the Young Hegelians (such as Bruno Bauer and Max Stirner) to the physiocrats and the classical political economists (such as Adam Smith and David Ricardo). Yet throughout these shifts, Marx’s attention in using the related concepts was not on the beliefs

---

19 I say more about the relationship between the concepts of ideology and false consciousness in §2.
of the proletariat specifically, so much as on those of the bourgeoisie or capitalist society more generally.

Even when, in the entire canon of Marx and Engels, the term ‘false consciousness’ eventually did appear once, when Engels wrote to Franz Mehring in 1893, it did not pertain to the proletariat specifically but to the figure of the individual “ideologist” (MECW 50: 164). Particularly, in the letter, Engels refers to Paul Barth as an instance of such an ideologist who ‘thinks with a false consciousness.’ And it is crucial to attend to the context of the correspondence in order to understand what Engels means in using the term here. Earlier, Barth had published a critique of Marx, in which he mischaracterised Marx for denying any efficacy to ideas (whether political, juridical, religious, etc.) in both their own development and the development of material relations (see Gustafsson 1966). Barth had instead advanced that ideas were the real driving force behind historical change. Hence, Engels’ characterisation of him as ‘thinking with a false consciousness’ was a sarcastic one: as having “carried out” ideology “consciously,” but with a false consciousness [mit einem falschen Bewußtsein], inasmuch as “[t]he actual motives by which he is impelled remain hidden from him” (MECW 50: 164).

That said, the following passage in The German Ideology (attributed to both Marx and Engels) about the division of mental and physical labour, is often cited as the conceptual origins of false consciousness as we understand it today (see, for example, Meyerson 1991, Pines 1993):

The ideas of the ruling class are in every epoch the ruling ideas: i.e., the class which is the ruling material force of society

---

20 This was ten years after Marx’s death in 1883.
21 Die Geschichtsphilosophie Hegels und der Hegelianer bis auf Marx und Hartmann [The Philosophy of History of Hegel and the Hegelians up to Marx and Hartmann] (1890).
22 This was, of course, a poor caricature of Marx. As Engels puts it in the letter to Mehring, the Marxian denial of independent historical development of ideas did not eo ipso deny their historical efficacy.
is at the same time its ruling *intellectual* force. The class which has the means of material production at its disposal, consequently also controls the means of mental production, so that the ideas of those who lack the means of mental production are on the whole subject to it. (MECW 5: 59)

The trouble, however, is that *how exactly* the ruling ideas are meant to have emerged from the material relations and how they subjected the oppressed were left unarticulated. With respect to this lack of articulation, Engels admitted, in the aforementioned letter to Mehring, the following:

> Otherwise only one point has been omitted, a point which, however, was never given sufficient weight by Marx and myself in our work, and in regard to which we are all equally at fault. For we all of us began, as we were bound to do, by placing the main emphasis on the derivation of political, legal and other ideological conceptions, as of the actions induced by those conceptions, from economic fundamentals. In so doing we neglected the formal in favour of the substantial aspect, i.e. the manner in which the said conceptions, etc., arise. (MECW 50: 164) [emphasis mine]

That is, although Marx and Engels had keenly emphasised *that* political and legal ideas derived from economic bases (for example, the idea of equality, as in Marx's *Critique of the Gotha Programme*), they had not specified *how exactly* such ideas were derived as a matter of practical constraints.23

---

23 Engels made a similar lament in an earlier letter to Joseph Bloch:

> If some younger writers attribute more importance to the economic aspect than is its due, Marx and I are to some extent to blame. We had to stress this leading principle in the face of opponents who denied it, and we did not always have the time, space or opportunity to do justice to the other factors that interacted upon each other. But it was a different matter when it came to depicting a section of history, i.e. to applying the theory in practice, and here there was no possibility of error. Unfortunately people all too frequently believe they have mastered a new theory and can do just what they like with it as soon as they have grasped—not always correctly—its main propositions. Nor can I exempt from this reproach many of the more recent 'Marxists' who have, indeed, been responsible for some pretty peculiar stuff. (MECW 49: 36)
Moreover, the subjection of the ruled to the ruling ideas was *not* understood by Marx and Engels to be a stable state, given the self-destructive trend of the capitalist mode of production—in relation to the infamous “Law of the Tendency of the Rate of Profit to Fall” (MECW 37: 209)—and the crises that arise thereupon as a result of countervailing tendencies.\(^{24}\) As these occur, for Marx and Engels, the common situation of the workers of lacking property, having familial relations dissimilar to the bourgeoisie’s, and the cosmopolitan character of increasing proletarianisation make it such that “[l]aw, morality, religion, are to [the worker] so many bourgeois prejudices, behind which lurk in ambush just as many bourgeois interests” (MECW 6: 494–495). The rule of bourgeois ideas over the proletariat, then, seem to be less a matter of the proletariat believing these ideas and more of being indirectly subject to them. That is, the division of labour and the competition between proletarians therein are the ways in which the mass of proletarians remain distantly subjected to bourgeois ideas of, say, classical political economy through the structural organisation of capital. So, for Marx and Engels, it is in the escalating class struggle between the bourgeoisie and the proletariat that this mass, initially merely sharing a common plight due to capitalist determinations, would through ever greater unionising across localities eventually become centralised and “[constitute] itself as a class for itself” (MECW 6: 211). That is, the proletariat would become self-determining and thus acquire revolutionary force.

To sum up, the term ‘false consciousness,’ at this point, was only used to refer to the thought processes of the bourgeois historian—for whom history takes place in the realm of ideas independently of economic facts—rather than the working class. And where some incipient notion of our understanding of false consciousness today might be located in *The German Ideology*, this is yet underspecified and merely states a relation between

\(^{24}\) The phenomenon of the tendency of the rate of profit to fall, as Marx himself notes in *Capital, Volume 3*, had already been observed by the classical political economists such as Smith (MECW 37: 211).
asymmetrical economic arrangements and an asymmetry in the production of ideas. Attempts to turn to only Marx and Engels to arrive at a definition of false consciousness would therefore be inadequate for understanding how it is generally used and understood today (as will be seen in §2).

1.2. Hindrance to Revolution

It is with Lenin and Gramsci’s concerns about how the proletariat revolution was to be brought about in Tsarist Russia and Fascist Italy, respectively, that we find greater attention on the ways in which bourgeois ideas hindered revolutionary praxis through their uptake among the working class. Lenin and Gramsci’s concerns were primarily the *implementational challenges* presented by bourgeois ideology for the proletariat to ‘constitute itself as a class for itself.’ Although they did not use the term ‘false consciousness’ in any significant way, we can again find closely related concepts that help us shed light on how the concept of false consciousness has been understood today since Engels’ letter.

In *What Is To Be Done?* (1902), for example, we find Lenin particularly concerned with the problem of “trade union consciousness” and how it might be prevented from developing into “Social-Democratic consciousness”:

> We have said that *there could not have been* Social-Democratic consciousness among the workers. It would have to be brought to them from without. The history of all countries shows that the working class, exclusively by its own effort, is able to develop only trade union consciousness, i.e., the conviction that it is necessary to combine in unions, fight the employers, and strive to compel the government to pass

---

25 This text is briefly attended to by Geuss (1981), in considering ideology in a descriptive, even positive, sense. I touch on the distinction between ideology in the descriptive and pejorative sense later in Note 42.
necessary labour legislation, etc. The theory of socialism, however, grew out of the philosophic, historical, and economic theories elaborated by educated representatives of the propertied classes, by intellectuals. By their social status the founders of modern scientific socialism, Marx and Engels, themselves belonged to the bourgeois intelligentsia. (LCW 5: 375)²⁶

Lenin here was specifically responding to an issue of the political newspaper Rabocheeye Dyelo, which had been satisfied with relying on the ‘spontaneity’ of the working class to organise themselves by themselves and thereby bring about change in their exploitative circumstances (for example, the 1896 textile worker strikes in St Petersburg demanded reliable wages and a reduction of working hours). But this position of the journal, for Lenin, was rooted in “a profound mistake” of “[imagining] that the labour movement pure and simple can elaborate, and will elaborate, an independent ideology for itself” (spontaneity) and thereby thinking that a specific focus on how economic asymmetry affected the beliefs of the proletariat was ‘overrated’ (LCW 5: 383). The problem was that the lack of attention to the bourgeois ideology circumscribing the content of their beliefs meant that the working class could only ever clamour for goals that were defined within a problematic mode of production. As such, they would develop ‘trade union consciousness,’ an intermediary stage which, left undeveloped, would serve to provide an additional hindrance to revolutionary change and thus played into the hands of the bourgeoisie.

What is particularly curious, however, is that Lenin finds the means to go beyond mere ‘trade union consciousness’ necessarily from among the ideas of bourgeois individuals. His solution to this famously thus involves a vanguard who drew their ideas from the bourgeois intelligentsia, with whom

²⁶ cf. The Communist Manifesto’s claim that “[t]he real fruit of [trade union] battles lies, not in the immediate result, but in the ever-expanding union of the workers,” aiming to “centralise the numerous local struggles, all of the same character, into one national struggle between classes” (MECW 6: 493).
“modern socialism originated, and it was they who communicated it to the more intellectually developed proletarians who, in their turn, introduce it into the proletarian class struggle where conditions allow that to be done.” (LCW 5: 383–384).27

In his Prison Notebooks (1929–1935),28 Gramsci shared a similar worry about how certain ideas could be a hindrance to the proletariat, yet understanding it more broadly in terms of how there was a ‘social hegemony’ directing them in ways that frustrate revolutionary activity:

Every social group, coming into existence on the original terrain of an essential function in the world of economic production, creates together with itself, organically, one or more strata of intellectuals which give it homogeneity and an awareness of its own function not only in the economic but also in the social and political fields.

[...]

The most typical of these categories of intellectuals is that of the ecclesiastics, who for a long time (for a whole phase of history, which is partly characterised by this very monopoly) held a monopoly of a number of important services: religious ideology, that is the philosophy and science of the age, together with schools, education, morality, justice, charity, good works, etc.

[...] The intellectuals are the dominant group’s ‘deputies’ exercising the subaltern functions of social hegemony and political government. These comprise:

27 cf. The Communist Manifesto’s claim that in times when the class struggle nears the decisive hour, the process of dissolution going on within the ruling class, in fact within the whole range of old society, assumes such a violent, glaring character, that a small section of the ruling class cuts itself adrift, and joins the revolutionary class, the class that holds the future in its hands. Just as, therefore, at an earlier period, a section of the nobility went over to the bourgeoisie, so now a portion of the bourgeoisie goes over to the proletariat, and in particular, a portion of the bourgeois ideologists, who have raised themselves to the level of comprehending theoretically the historical movement as a whole. [emphasis mine] (MECW 6: 494)

28 Gramsci is briefly considered in Meyerson 1991, Shelby 2003 and Stanley 2015, the latter two in their terms of “ideological consciousness” and “ideological beliefs,” respectively. I say more about how these terms relate to false consciousness later in §2.
1. The ‘spontaneous’ consent given by the great masses of the population to the general direction imposed on social life by the dominant fundamental group; this consent is ‘historically’ caused by the prestige (and consequent confidence) which the dominant group enjoys because of its position and function in the world of production. […] (Gramsci 1948, 5–12)

Rather than seeing a stunted development of working-class consciousness being a hindrance to revolutionary change, Gramsci saw a status quo hegemony directing the working class (who gives ‘spontaneous consent’) as the obstacle to change—that is, contra Marx and Engels’ picture of the worker being increasingly unfazed by bourgeois ideas as capitalism develops. More than being subject to ruling ideas through the extant economic organisation of capital, or even to bourgeois ideology qua worker demanding socio-economic change, false consciousness now involved the internalisation of these ideas into the worker’s own thought processes “outside of his professional activity” (Gramsci 1948, 9).

Furthermore, such a hegemony was no longer the sole purview of the economic ruling class but any social group. Gramsci’s own solution against anti-revolutionary hegemony was not for the proletariat to turn to the bourgeois intelligentsia for more direction, but to turn to the “organic intellectuals” within its own social group to create and sustain a counter-hegemony by developing its own ideas of social life in order to overtake and assimilate bourgeois ideas (including those of Marx and Engels), forming a ‘philosophy of praxis’:

One of the most important characteristics of any group that is developing towards dominance is its struggle to assimilate and to conquer ‘ideologically’ the traditional intellectuals, but this assimilation and conquest is made quicker and more efficacious the more the group in question succeeds in simultaneously elaborating its own organic intellectuals. (Gramsci 1948, 10)
In sum, with both Lenin and Gramsci, the concept of false consciousness shifted from an emphasis on the (illusorily) independent thought process of the individual bourgeois ideologists in relation to economic asymmetry, to how the bourgeois content of certain beliefs served to hinder revolutionary ideation among the proletariat class in relation to the same asymmetry. False consciousness was also now explicitly seen as a direct contributor to the continued exploitation of workers beyond how bourgeois ideology structured their working conditions. What is also significant, further, is that the concept no longer specifically pertained to individual (bourgeois) thinkers, but now characterised the beliefs and attitudes of the proletariat as a class when they are directed by any non-proletarian ideology. In other words, at this moment of the development of the concept of false consciousness throughout the tradition, it acquires the meaning of functional ‘falsity,’ being beliefs and attitudes that motivate them to act in ways that would perpetuate their own oppression.

But there are also two new questions that arise between Lenin and Gramsci: whether false consciousness is pervasive to such an extent as to become the default condition of the working class (as in under hegemony) or is merely a development of it (as in trade union consciousness), and whether the sources of emancipatory consciousness are internal or external to the working class.29 And, as we will see in §3, these questions, while altered in the subsequent genealogical shifts, nevertheless still contribute to the problems of false consciousness that remain with us today.30

29 The political value of spontaneity is a matter of longstanding controversy. Rosa Luxemburg, for example, famously contested Lenin’s denigration of spontaneity (Luxemburg 1904, cf. Lukács 1923).
30 Specifically, these both contribute to the Explanatory and Normative Problems, which I deal with in Chapters II and III, respectively.
1.3. Distorted Relation to Reality

The epistemological dimension to the concept of false consciousness was further explored with Lukács’ *History and Class Consciousness* (1923), particularly in his widely-cited essay “Class Consciousness,” with his attention on the more philosophical (read: Hegelian) strands of Marx’s early writings. Lukács in particular saw himself as drawing on these writings to fill in theoretical gaps in Lenin’s revolutionary method. What was distinctive here about Lukács’ discussion of the concept of false consciousness was an attention to the conditions under which false consciousness is said to be false—that is, how it is the case that the beliefs of the proletariat are problematically related to social reality, i.e. how they are ignorant (epistemic ‘falsity’).

Lukács was perhaps the first to explicitly return to, and elaborate, the term ‘false consciousness’ from Engels’ letter, as referring to the thought process of the bourgeois ideologist. Lukács makes explicit that the thought process of the bourgeois ideologist (such as Barth) is here considered only inasmuch as it expresses “the appropriate and rational reactions ‘imputed’ to a particular typical position in the process of production,” which is “neither the sum nor the average of what is thought or felt by the single individuals who make up the class”—i.e., “class consciousness” (Lukács 1923, 51). That is, in Marx’s words, the individual is considered “only in so far as they are the personifications of economic categories, embodiments of particular class relations and class interests” (MECW 35: 10). But Lukács also takes the term

---

31 While Lukács’ book has been an important source text for the concept of false consciousness and understanding Marxian views on oppression more generally by analytic philosophers (particularly, Geuss 1981, Meyerson 1991, Medina 2013), he in fact repudiated this work twice. First, for political reasons, when he was forced in 1924 by its condemnation at the Fifth Congress of the Communist International (Axelos 1960), and, second, in a retrospective in a new preface to the 1967 republication of the book (Lukács 1967). These repudiations, however, largely concern the Hegelian and Marxian concepts of alienation and do not affect the genealogy here. For more on how Lukács understands ideology to be distinct from false consciousness, see Eagleton 1991.
‘false consciousness’ to apply more broadly to any class and extends ‘falseness,’ from the ignorance of the ‘real motives of ones’ actions,’ to ignorance specifically in relation to “the actions they need to perform in order to obtain and organise power” (Lukács 1923, 53).

It is useful to understand how Lukács contributes to the concept of false consciousness by way of a juxtaposition with his understanding of non-proletarian ‘false consciousness.’ For the non-proletarian classes (such as the bourgeoisie or feudal lords), knowledge of the actions ‘they need to perform in order to obtain and organise power’ is not a condition on performing them. Such non-proletarian ‘false consciousness,’ in fact, is “crucial for the activity of the classes in society.” (Lukács 1923, 55). That is, while such classes (the bourgeoisie or feudal lords) do obtain and organise power, it is necessary for them to remain ignorant about the ways in which they obtain and organise power in order to retain their position. This is because Lukács understands the relevant kind of knowledge of p to specifically pertain to how p relates to “society as a concrete totality, the system of production at a given point in history and the resulting division of society into classes” (ibid., 50). And such knowledge would be destabilising for these classes.

Illustrating this with the particular case of the bourgeoisie, Lukács understands their ignorance to be a result of an inherent tension “between the function of capital as private property and its objective economic function” (ibid., 63). That is, capital is believed by the bourgeoisie to be a matter of private ownership and an exercise of their individual freedom and rights in the context of immutable laws of the economy, as it were. Such beliefs are necessary for the bourgeois individual inasmuch as they see themselves as participants within civil society. Yet, understood in relation to ‘the system of (capitalist) production’ more broadly, such activities are anything but ‘private’ and have profound implications for the functioning of
the system, at the expense of the proletariat and other classes: the ‘laws,’ in fact, are not immutable but brought about by their very activity. It is in this sense that the bourgeoisie who come to this knowledge can no longer remain insulated in ‘false consciousness’ about their role in the system of production. As the self-destructive trends of capitalism continue to present crises and manifestations of class struggle that portend the impermanence of the capitalist mode of production, the bourgeoisie as a class are forced to confront both this and their role in it, such that this ‘false consciousness’ develops into what Lukács calls ‘mendacious consciousness’: where the bourgeoisie have to justify “hegemony […] exercised not merely by a minority but in the interest of that minority,” such that “[e]ither they must consciously ignore insights which become increasingly urgent or else they must suppress their own moral instincts in order to be able to support with a good conscience an economic system that serves only their own interests” (ibid., 65). ‘False consciousness,’ then, would not simply be a passive description of the bourgeoisie’s situation but be actively brought about by them.32

Lukács’ understanding of ‘false consciousness’ is different in the case of the proletariat. And what is crucial to understanding his contribution to the concept of false consciousness as we understand it today to pertain to the oppressed is that, unlike the other classes, Lukács sees the proletariat as the only class “able to see society from the centre, as a coherent whole” and, according to him, is thereby the only class “able to act in such a way as to change reality” (Lukács 1923, 69). That is, knowledge of how their activities relate to the system of production as a whole is a condition on obtaining and organising power. This is because while the beliefs of the other classes do not and need not relate to the capitalist mode of production

32 As such, Lukács understands the moral development of the bourgeoisie in mendacious consciousness not as a matter of hegemony per se but defensive concessions that acknowledges the fragility of their class situation. This is similar to a kind of motivated, and even wilful, ignorance (Wieland 2016, cf. Mills 2007)
comprehensively, the proletariat are forced by class warfare “to advance beyond what was immediately given” (ibid., 72). That is, the proletariat are compelled by crises and escalating class struggle towards realising that, as participants in capitalist civil society, its beliefs about itself, other classes, and the world (for example, the idea of ‘fair working conditions’), are ultimately inadequate. The proletariat comes to the realisation that the world and its ‘laws of nature,’ rather than being as immutable as the proletariat and the bourgeoisie had thought them to be, need to be—and that they also can be—changed by it. The extent of the requisite change would involve changing the extant social order as such: in other words, the proletariat “cannot liberate itself as a class without simultaneously abolishing class society as such” (ibid., 70).

Yet, for Lukács, such a realisation is not yet the case and this is also precisely what makes false consciousness uniquely problematic for the proletariat: that the overcoming of false consciousness and the necessary step for political liberation (i.e. revolution) is in tension with the more immediate economic struggles (such as fights for ‘fairer’ wages and better working conditions). Concern for the latter, not only presupposes the maintenance of a proletariat class, but also distracts from the ultimate goal of revolution. Lenin’s ‘trade union consciousness,’ then, can be seen as an instance of such a development of false consciousness driven by economic
struggles.\textsuperscript{33} The false consciousness of the proletariat, then, is not so much a problem of a passive situation (due to the escalating crises and struggles) or active self-deception (since this would not resolve even the immediate issue). Rather, it is specifically a problem of an ignorance of the systematicity of their oppression. Awareness of this systematicity is not achieved all at once, but through the unification of local struggles and continual criticism of insufficiently comprehensive beliefs. As such,

\begin{quote}
[t]he struggle for [a classless] society […] is not just a battle waged against an external enemy, the bourgeoisie. It is equally the struggle of the proletariat \textit{against itself}; against the devastating and degrading effects of the capitalist system upon its class consciousness. The proletariat only have won the real victory when it has overcome these effects within itself. […] The proletariat must not shy away from self-criticism, for victory can only be gained by the truth and self-criticism must, therefore, be its natural element. (Lukács 1923, 80–81)
\end{quote}

To sum up, Lukács provides an epistemological basis for the concept of false consciousness and how it hindered revolution. It was, for him, the default position of the proletariat, insofar as the content of their beliefs do not pertain to society ‘as a concrete totality.’ As such, these beliefs were not effective in bringing about solidarity among proletarians and thereby consolidating revolutionary power. Furthermore, false consciousness is further exacerbated by the immediacy of local economic struggles: these

\textsuperscript{33} Cf. Lenin’s description of the ‘trade union consciousness’ of the 1896 textile worker strikes:

\begin{quote}
The [local] revolts were simply the resistance of the oppressed, whereas the systematic strikes represented the class struggle in embryo, but only in embryo. Taken by themselves, these strikes were simply trade union struggles, not yet Social Democratic struggles. They marked the awakening antagonisms between workers and employers; but the workers, were not, and could not be, conscious of the irreconcilable antagonism of their interests to the whole of the modern political and social system, i.e., theirs was not yet Social-Democratic consciousness. In this sense, the strikes of the nineties, despite the enormous progress they represented as compared with the ‘revolts,’ remained a purely spontaneous movement. (LCW 5: 375)
\end{quote}
symptomatic struggles distract the proletariat from realising and organising against the underlying malady (i.e. class society itself). Lastly, Lukács’ analysis of false consciousness pertained neither to the beliefs and attitudes of individual proletarians nor that of the proletariat as a whole: what was the focus was attention to “the appropriate and rational reactions ‘imputed’ to a particular typical position in the process of production” (Lukács 1923, 51). As such, under such an analysis, an individual or an idea was no longer considered ‘bourgeois’ once it ceases to be circumscribed within the bourgeois position in the capitalist system of production and is related to the system as a whole. The source of emancipatory consciousness was thus entirely located within the proletariat class.

But, from this consideration, there nevertheless remains the question of how it is possible for the proletariat to identify within local struggles particular beliefs of theirs as problematically still circumscribed within the capitalistic worldview, extrapolating from the content of these specific beliefs to develop beliefs with wider content that pertain to the underlying social order in its totality. There is also the question of whether there is a bootstrapping problem in expecting a class which continually experiences the ‘devastating and degrading effects of the capitalist system upon its class consciousness’ to overcome immediate struggles and reach beyond a worldview thoroughly structured by capitalism. These questions are particularly important because, according to Lukács, “the fate of the revolution (and with it the fate of mankind) will depend on the ideological maturity of the proletariat, i.e. on its class consciousness” (Lukács 1923, 70). Both questions, however, receive only pessimistic answers with the early Frankfurt School.

---

34 These both contribute to the Explanatory Problem and the Agential Problem (which I sketch more fully in §3), respectively—which are in turn dealt with in Chapters II and IV.
1.4. Socially Necessary Product

The final significant shift in the meaning of the concept of false consciousness occurred with the philosophers associated with the Institute for Social Research at Frankfurt, i.e. the Frankfurt School (whose founding Lukács himself was instrumental to).\(^{35}\) Attention here is paid to how it is that false consciousness is caused by epistemic practices adversely conditioned by structural oppression, i.e. *genetic* ‘falsity.’ The Frankfurt School’s development of Marxism as critical theory was very much a response to the totalitarian social orders of the Second World War, as well as the enduring stability of the capitalist society in Europe and the US despite the upheavals of the war. As Jürgen Habermas puts it, “Critical Theory was initially developed in Horkheimer’s circle to think through political disappointments at the absence of revolution in the West, the development of Stalinism in Soviet Russia, and the victory of fascism in Germany. It was supposed to explain mistaken Marxist prognoses, but without breaking Marxist intentions” (Habermas 1985, 116).

It is with the Frankfurt School—specifically Adorno, Horkheimer, and Marcuse—that the concept of false consciousness as we understand it today is removed from the context of revolutionary praxis, pertaining no longer to the proletariat specifically, and becomes characteristic of living under ‘late’ capitalist society as such.\(^{36}\) In fact, for the Frankfurt School, “the reality of the laboring classes in advanced industrial society makes the Marxian ‘proletariat’ a mythological concept” (Marcuse 1964, 193). With this, revolutionary consciousness is also no longer seen as the converse of false

---

\(^{35}\) Lukács could not officially be involved with the Institute because of the restrictions imposed by his membership in the Communist Party.

\(^{36}\) The attention to Adorno, Horkheimer, and Marcuse in analytic philosophy are largely due to Geuss 1981, Meyerson 1991, Pippin 1994, M. Rosen 1996. For the different nuances in how other members of the Frankfurt School approached the concept of false consciousness (such as Eric Fromm, Habermas, and Walter Benjamin), see Eyerman 1981, Geuss 1981, and M. Rosen 1996.
consciousness. The converse of false consciousness is now simply to “fashion” perspectives “that displace and estrange the world [...] as indigent and distorted,” and “beside [this] demand thus placed on thought, the question of the reality or unreality of redemption itself hardly matters” (Adorno 1951, 247). As Marcuse puts it:

The critical theory of society possess no concepts which could bridge the gap between the present and its future; holding no promise and showing success, it remains negative. Thus it wants to remain loyal to those who, without hope, have given and give their life to the Great Refusal [i.e. “the protest against that which is”]. (Marcuse 1964, 261 [66f.])

Attending to how the Frankfurt School understands the disappearance of the proletariat class, along with the disappearance of its accompanying revolutionary hope, is instructive for us in understanding their contribution to the concept of false consciousness today.

In “Puzzle-picture” of *Minima Moralia* (1951), Adorno provides an explicit answer to “the grimly comic riddle: where is the proletariat?”: economic development has advanced to the extent that control over capital demands technical expertise, while “[the] dissection of [technical processes] into minute operations largely independent of education and experience, makes the expertise of these new-style managers to a large degree illusory, a pretence concealing the privilege of being appointed” (Adorno 1951, 194). As such, “[w]hile objectively the relation of owners and producers to the productive apparatus grows ever more rigid, subjective class membership becomes all the more fluctuating” (ibid., 193). This fluctuation is a result of both “the [seeming] possibility that anyone can be appointed” and “an egalitarian threat” of “an opaque hierarchical structure” that threatens even the elites: there is thus a promise of becoming an elite as long as one maintains a “healthy technocratic outlook” (ibid., 194). Thus, as a result of the increased technicisation of capital (specifically in mental labour), the
immediate economic situation presents an appearance of justice and equality which goes unchallenged, since there is no longer a crisis (apparently) or class antagonism to impel workers ‘beyond what is immediately given’ and constitute themselves as a class unto their own.

Moreover, as Marcuse observes in *One-Dimensional Man* (1964), these beliefs are self-reinforcing in such a society:

> [The way in which men have organized their societal labor] is natural only to a mode of thought and behavior which is unwilling and perhaps even incapable of comprehending what is happening and why it is happening, a mode of thought and behavior which is immune against any other than the established rationality. To the degree to which they correspond to the given reality, thought and behavior express a false consciousness, responding to and contributing to the preservation of a false order of facts. And this false consciousness has become embodied in the prevailing technical apparatus which in turn reproduces it. (Marcuse 1964, 149)

In this way, beliefs about the ‘justice’ and ‘equality’ of late capitalist society are reproduced inasmuch as the status quo situation that gave rise to them is maintained.

Additionally, in the “Elements of Anti-Semitism” in their *Dialectic of Enlightenment* (1947), Adorno and Horkheimer describe how this is not merely a matter of the relation between how individual workers relate to capital, but also (similar to Gramsci) how these workers relate to themselves in the broader culture whose production is subject to the same technicisation of capital:

> The subjects of the economy are psychologically expropriated, and the economy is more rationally operated by society itself. The individual no longer has to decide what he himself is to do in a painful inner dialectic of conscience, self-preservation and drives. Decisions for men as active workers are taken by the
hierarchy ranging from the trade associations to the national administration, and in the private sphere by the system of mass culture which takes over the last inward impulses of individuals, who are forced to consume what is offered to them. The committees and stars serve as the ego and super-ego, and the masses, who have lost the last semblance of personality, shape themselves more easily according to the models presented to them than the instincts ever could by the mechanism of inner censorship. […] Since thought has become a mere sector of the system of division of labor, the plans of the competent experts and leaders have made it superfluous for individuals to plan their own happiness. The irrationality of the unresisting and busy adaptation to reality becomes more reasonable than reason for the individual. […] As industrial society progresses and is supposed to have overcome its own law of impoverishment, the notion which justified the whole system, that of man as a person, a bearer of reason, is destroyed. The dialectic of Enlightenment is transformed objectively into delusion. (Adorno and Horkheimer 1947, 203–204)

Under late capitalist society, the stability of false consciousness has to do not just with how the interaction between the worker and their local production process(es) effectively maintains status quo beliefs about the organisation of the production process(es). The technicisation of capital has developed to the extent that this lack of antagonism with the management of the production process(es) is in a large part also due to the absorption of the wider culture and society into the capitalist system. Individual workers’ belief-forming processes are extensively compromised beyond matters pertaining to the production process, due to an extension of the technical division of mental and physical labour in the workplace to every other aspect of culture and life—such that the epistemic practices that emerge tend to remain circumscribed within such a society’s parameters and antagonistic beliefs thus lack room to develop.

To sum up, following Lukács, the Frankfurt School sought to analyse false consciousness within the context of capital civil society qua a “system of production at a given point in history and the resulting division of society into
classes” (Lukács 1923, 50). But what is distinctive about the Frankfurt School’s understanding was that false consciousness—rather than being a result of an underdeveloped set of beliefs about how the oppressed’s socio-economic situation should be ameliorated or a social hegemony of a dominant social group’s ideas—is now seen a self-reproducing, emergent feature of late capitalist society as a whole: a “necessarily false consciousness” (Frankfurt School for Social Research 1956, 190). Further, since the same false consciousness extends to all within such a society, there being no longer a distinction between the beliefs of the workers and the elites about their socio-economic situation, the criticism of false consciousness (which pertains to the ideas of a given oppressed group) converges with the criticism of ideology (which pertains to the ideas of a given society). As we will see in the next section (particularly §2.2), this convergence contributed to why social theorists and philosophers who work on ideology today tend to also assume to a fortiori have addressed issues of false consciousness.

With the Frankfurt School’s modification of the concept of false consciousness, two questions in particular thus arise: If we are all in the grip of false consciousness, how can there be any way for anyone to see what an unadulterated, or ‘true’ consciousness would look like? And, even if it were conceptually possible, who would be able to make such a critique? These questions, taken together, are hallmarks of the Critical Theory tradition as it has survived since and constitute one of the key problems of false consciousness today.37

From the genealogy above, we can see how false consciousness might be generally understood in contemporary analytic social philosophy—even outside Marxian and Critical Theory contexts—to involve how it might be

37 These both contribute to the Normative Problem, which I sketch more fully in §3 and deal with in Ch. III.
systematically the case that an oppressed agent’s set of beliefs is epistemically, functionally, and genetically ‘false.’ False consciousness pertaining fundamentally to an individual agent (rather than a social group) maps onto Engels’ use of the term (as we saw in §1.1); its epistemic character maps onto Lukács’ concern with false consciousness’ ignorance of social reality (§1.3); its functional character maps onto Lenin and Gramsci’s attention to how bourgeois ideas hinder successful social change (§1.2); and its genetic character maps onto the Frankfurt School’s focus on how false consciousness is socially necessitated under late capitalism (§1.4). It is thus this general definition that will be operative throughout the rest of this thesis, for it best resonates with the accrual of meanings that the concept has had throughout key shifts in the Marxian and Critical Theory traditions, before its more recent fall from grace.

We saw also from the above genealogy that the term ‘false consciousness’ was sometimes applied to cases of similar phenomena with respect to the bourgeoisie. This has thus understandably occurred at points in the analytic philosophical literature (see, for example, Geuss 1981, Meyerson 1991). However, the one key difference between false consciousness as it pertains to the bourgeoisie (and, more generally, oppressors), as compared to the proletariat (the oppressed), is that “the bourgeoisie, paradoxically enough, has an interest in being self-deceived” (Geuss 1981, 24). That is, it is to the benefit of oppressors that they remain in false consciousness—however narrow these benefits might be (as we had seen with Lukács). This, of course, is also crucial a line of inquiry to be pursued in coming to an understanding of oppression and how it can be resisted. Recent analytic social philosophy has devoted much attention to this (as well as neighbouring phenomena) in terms of concepts such as white ignorance (see, for example Mills 2007, 2015, Martín 2020) and epistemic oppression (Fricker 2007, Dotson 2014), as well as in the revived discussions on the concept of ideology by the “New Ideology Critics” such as Rahel Jaeggi,
Robin Celikates, Sally Haslanger, and Jason Stanley (Sankaran 2020; see Jaeggi 2009, Celikates, Haslanger, Stanley forthcoming). For the purposes of clarity, however, I will use the term ‘false consciousness’ to pertain specifically to cases where an individual is a member of an oppressed group. I will reserve ‘ideological consciousness’ as an umbrella-term that includes oppressors, designating the more general phenomena of individuals holding a certain set of ignorant beliefs that arise in virtue of structural oppression and that serve to perpetuate structural oppression, regardless of whether they are oppressed. That is, while all false consciousnesses would be ideological, not all ideological consciousnesses would be false consciousnesses (I say more about this relation between false consciousness and ideology in §2.2).

2. Preliminary Clarifications

I will now consider how false consciousness, as understood here, differs from the alternative definitions in the philosophical literature and, in the process, clarify some of the parameters of the present inquiry. To recall, these alternatives are: false consciousness as mere ignorance of the basis of one’s beliefs (§2.1) and false consciousness as the acceptance of ideology simpliciter (§2.2). As I will now go on to show, the turn to these definitions are unwarranted and have tended to obfuscate the utility of the concept of false consciousness.

38 For others who have argued that these are conceptually distinct within the Marxian and Critical Theory traditions, see Eyerman 1981, Eagleton 1991, and M. Rosen 1996.
2.1. Ignorance of the Basis of One’s Beliefs

Let us now first turn to the trend that defines ‘false consciousness’ as merely ignorance about the irrational reasons for which one holds onto a given set of beliefs. Purportedly taking their cue from Engels’ 1893 letter to Mehring, some philosophers decouple the concept of false consciousness from the concept of ideology and its structural implications, holding that the former is a cognitive matter of merely mistaking the basis of one’s beliefs (see, for examples, Runciman 1969, Wood 1988, Barnes 1997, van Fraassen 2002, Shelby 2014). For example, a ‘raunch feminist’ might think that her beliefs about the purported feminist nature of her ‘sex-positive’ parties are based on her university education, when the belief is really based on a “fear of being branded as ‘bad women’ by their misogynistic environment” (Melo Lopes 2019, 2528). False consciousness, so understood, is simply “the misrecognition of its own presuppositions, of its own effective conditions, a distance, a divergence between so-called social reality and our distorted representation” (Žižek 1989, 24). And this is often understood to refer to beliefs thought to be held rationally when they are really held irrationally. After all, the basis of many ordinary beliefs—albeit perfectly rational—may be unknown to the subject. As we saw in the genealogy of the previous section, however, this narrow definition does not on its own adequately characterise the ways in which the concept of false consciousness has been understood in the broader Marxian and Critical Theory contexts.

And if we are to limit the definition of the concept solely to the context of Engels’ letter, the interpretation misses the mark even more. Consider the

---

39 To be clear, Filipa Melo Lopes argues that these beliefs are explained by a ‘meaning vertigo’ such women experience, a “perception of a vertiginous and unsettling emptiness at the level of social meaning” (Melo Lopes 2019, 2531). As I will suggest later in Ch. II, meaning vertigo can be seen as a mechanism that contributes to the maintenance of false consciousness (as I understand it).
oft-cited passage from Engels’ letter that such philosophers draw on for support:

Ideology is a process which is, it is true, carried out consciously by what we call a thinker, but with a [false consciousness]. The actual motives by which he is impelled remain hidden from him, for otherwise it would not be an ideological process. Hence the motives he supposes himself to have are either spurious or illusory. Because it is a mental process, he sees both its substance and its form as deriving solely from thought—either his own or that of his predecessors. He works solely with conceptual material which he automatically assumes to have been engendered by thought without inquiring whether it might not have some more remote origin unconnected therewith; indeed, he takes this for granted since, to him, all action is induced by thought, and therefore appears in the final analysis, to be motivated, by thought. (MECW 50: 164)

While this passage on its own might seem to suggest textual support for this narrow definition, it is crucial to contextualise it with the paragraph just preceding it in the letter. In that paragraph, Engels emphasises to his fellow historical materialist, Mehring, “the derivation of political, legal and other ideological conceptions, as of the actions induced by those conceptions, from economic fundamentals,” before he describes the mental process of the ‘so-called [sogenannten] thinker’ as being a case of ‘false consciousness’ here (MECW 50: 164). And, as we saw in §1.1, this ‘so-called thinker,’ against which the charge of ‘false consciousness’ had been made, was in fact the bourgeois ideologist (specifically, Barth):

The [ideologists] forget, often almost deliberately, that an historical element [of theology, philosophy and political science], once it is ushered into the world by other, ultimately economic, causes, will react in its turn, and may exert a

---

40 The translation for ‘mit einem falschen Bewußtsein’ here was “with a consciousness that is spurious.”
41 See also Mills 1985, which also shares this interpretation of Engels by means of a detour through the latter’s Ludwig Feuerbach and the End of Classical German Philosophy (1886) and Anti-Dühring (1878)—but focusing specifically on the “tendency of intellectuals in class society to ignorance the social determinants of their thought” (Mills 1985, 341).
reciprocal influence on its environment and even upon its own causes. Cf. Barth, for example, on the priestly caste and religion, your p. 475. I was delighted by the way you dismissed this quite incredibly superficial Johnnie. And they go and make the chap professor of history at Leipzig! Old Wachsmuth used also to be there; he too was not a little shallow-pated but he had a tremendous feeling for facts—a very different sort of chap. (MECW 50: 165)

More than ignorance of the basis of one’s irrational beliefs, then, the term ‘false consciousness’ was used here with the background understanding that not only was this a characterisation of a thinker (namely, Barth) carrying out his function as a bourgeois ideologist, but also such a characterisation would not be in principle extendable to the proletariat, who do not have the means of mental production to perform such a mental process (as Marx and Engels argue in The German Ideology).

That said, perhaps one might be tempted to make the move to claim that the concept of false consciousness can be defined independently of its Marxian origins. But understanding it merely as being ignorant of the irrational basis of one’s beliefs is still unhelpfully unspecific for the purposes of emancipatory praxis. Such a characterisation could apply to a wide range of beliefs that we would not ordinarily regard as problematic, even if they were held recidivistically: for example, “that I have hands, that 2 + 2 = 4, that my mother loves me” (Srinivasan 2016, 374). Additionally there are other problematic kinds of recidivistic beliefs that would all fall under such a broad definition: for example, delusions, self-deception, epistemic vice, or implicit bias. Of course, it could well be the case that one is satisfied with taking ‘false consciousness’ to involve such a sweeping definition. But then these various kinds of beliefs differ so much from each other that, apart from flouting the genealogy I have sketched above, there would be little, if any, emancipatory utility in using the term ‘false consciousness’ as such a broad umbrella term. That a given set of beliefs could be either delusional or a result of epistemic laziness does not on its own tell us anything specific
about the nature of the social structures in which the epistemic agent is embedded: one may be delusional or epistemically lazy in a utopia.

2.2. Ideology and False Consciousness

At the end of the previous section (§1), I distinguished the term ‘false consciousness’ from ‘ideological consciousness,’ reserving the latter as a term that would include both oppressed and oppressors. There is a trend, however, for philosophers to nonetheless use these two terms interchangeably (see, for example, Geuss 1981, Meyerson 1991, Pines 1993, Stanley 2015). This trend allows for the concept of false consciousness to play a greater role in social analysis than with the definition considered in the previous section. False consciousness here is understood to have some use in critique (as opposed to mere criticism): that is, to draw attention to the neglected, oppressive aspects of an otherwise seemingly benign social structure and, in doing so, orient us towards amelioration or emancipation. The thought behind treating ‘false consciousness’ and ‘ideological consciousness’ as synonyms is that the concept of false consciousness is so bound up with the critical concept of ideology (which has historically and even now received much more attention). This is so much so that ideology critique would be thought to sufficiently exhaust any critique of false consciousness and more—especially since it would also include the beliefs of oppressors (recall the convergence that we noted with the Frankfurt School in §1.4). Attending to how the concept of ideology is generally
understood by such philosophers would allow us a better sense of how one might have such an impression.42

Recent reflections on the concept of ideology tend to disambiguate between general or pejorative senses of the term (see Haslanger 2014, Celikates 2017): the general (or non-evaluative/descriptive) sense remains agnostic about the problematic nature of the social practices undergirded by ideology (this sense was used by both Lenin and Lukács, as we saw; cf. Geuss 1981); whereas the pejorative (or evaluative/normative) sense has that the social practices undergirded by ideology are always problematic.

To note, this distinction between the two senses can already be found in the locus classicus of the concept of ideology, which, much like the concept of false consciousness, is the works of Marx and Engels. As I observed in §1.1, ‘Ideology’ in Marxian scholarship is often understood as referring to those “definite forms of social consciousness” corresponding to the superstructure of legal and political institutions and practices (MECW 29: 263), which are themselves distinct from the base of economic/material entities and operations (see Cohen 1979, cf. Rosen 1996). At the same time, I argued that ideology was understood by Engels specifically in his 1893 letter to Mehring to include, in a pejorative sense, higher-order institutions and practices that attempt to understand the former without an appreciation of

42 An example of yet another critical concept might be found in Mark Fisher’s Capitalist Realism, where he proposes the concept of “depressive anhedonia,” which is defined as “an inability to do anything else except pursue pleasure”:

There is a sense that ‘something is missing’—but no appreciation that this mysterious, missing enjoyment can only be accessed beyond the pleasure principle. In large part this is a consequence of students’ ambiguous structural position, stranded between their old role as subjects of disciplinary institutions and their new status as consumers of services. (Fisher 2009, 21–22).

Fisher suggests that the concept leads us to realise that the rampant depression among teenagers in the UK is a result of “privatizing these problems—treating them as if they were caused only by chemical imbalances in the individual’s neurology and/or by their family background—any question of social systemic causation is ruled out” (ibid., 21).
the causal powers of the base (e.g. idealist theories about the origins and function of legal and political institution and practices, such as Barth’s; cf. MECW 26).

Importantly, this implicit two-fold interpretation of Marxian ‘ideology’ has been anticipated and made explicit by Charles Mills (1985, 1992; cf. Mannheim 1936), who points out that these two senses of ‘ideology’ in Marx and Engels can and should be disambiguated because it would allow for: first, more precise investigations that distinguishes those social aetiologies of beliefs that are neutral (either merely superstructural or materialist) from those that are problematic (idealist, in the Marxian case); and also, second, “multidimensional conceptualization[s] of ideologies” beyond a narrow class-reductionist conception, since the underlying problem with ideology is not so much its being specifically tied to partisan class ideas but concerns how ideas are ignorant of the superstructure’s base more generally (Mills 1992, 323; see also 1994, 2017).

As such, recent reflections on ideology (such as those by the aforementioned New Ideology Critics) might thus be read as following the general direction of Mills’ distinction in their use of the concept beyond its Marxian context. That is, these philosophers generally take agents, oppressed or otherwise, to be organised under schemas of “public meanings, scripts, etc.” that confer upon them a range of constraints and enablements salient for action-guidance (Haslanger 2019, 5; cf. Connolly 1981, Ásta 2018). ‘Ideology,’ in particular then, refers to a schema that serves in some way to undergird, and also arises out of, problematic social practices. For example, the public meaning of marriage as a necessary institution in Victorian gender ideology undergirds, and also arises out of, romantic practices and legal institutions that involve interrelated public

---

43 I thank my examiners for their advice to highlight the relevance of Mills’ understanding of Marxian ideology.
meanings such as ‘not wanting to marry means being unfeminine’ for
women and scripts such as ‘propose immediately when you take an interest
and liking’ for men (Caird 1888). A network of such practices, involving
resources (such as tools and materials) and agents, organised by schemas,
is said to constitute a social structure (Haslanger 2011). So these romantic
practices and legal institutions constitute the social structure that is British
patriarchy, organised by Victorian gender ideology broadly understood.

These schemas or ideologies are usually taken for granted by ordinary
agents within the social structure, left in the background and unthematised,
such that they seem to be ‘just the way things are.’ Allowing us to thematise,
problematisate, and reorient, our consideration of social practices is precisely
what the concept of ideology is meant to do: with it, we would be able to
identify problems in schemas that manifest in particular social practices. For
instance, British suffragists like Harriet Taylor and John Stuart Mill drew on
some tacit theory of ideology when they campaigned against the gender
inequality manifest in asymmetrical marital laws and parliamentary
exclusion. Such practices, they argued, were for a long time organised
around the ideological principle that “it is the duty of women, […] that it is
their nature, to live for others; to make complete abnegation of themselves,
and to have no life but in their affections […] to the men with whom they are
connected, or to the children who constitute an additional and indefeasible
tie between them and a man” (Mill 1869, 132–133; cf. Cudd 2006). The
assumption here, then, is that if ‘false consciousness’ simply designates
how such schemas are internalised by an oppressed agent at a given node
within the network, then an analysis of an instance of false consciousness
would already have to involve the more comprehensive analysis of its
corresponding ideology. That is, an analysis of false consciousness would
already involve structural explanations, attending to “the behavior of the
individual given their place in a structure” and also understanding “the
individual as the instance of a type—a type defined by the conditions for
existing at that node” (Haslanger 2016, 129). Separate attention to the concept of false consciousness, therefore, would seem to be redundant.

Yet, while I agree with the general picture of a conceptual intimacy between ideology and false consciousness here, it is nevertheless important to distinguish between the questions concerning either. The concept of false consciousness attends to that which a focus on ideology alone overlooks: one’s capacity for self-knowledge qua agent. That is, while a focus on ideology is more concerned with problematic representations in relation to structures as a whole, a focus on false consciousness is more concerned with the formation and maintenance of these problematic representations in relation to the lived experiences of agents in these structures (cf. Eyerman 1981, Eagleton 1991, M. Rosen 1996). Such explicit focus on how and why oppressed agents at these structural nodes “effectively confer from the bottom up the relations of subordination that affect their lives” is relatively lacking in the recent philosophical literature that is focused on ideology or epistemic injustice enacted by oppressors (Thompson 2013, 185). But the bottom-up focus of false consciousness is just as essential for critique as the top-down focus of ideology. False consciousness reveals an important aspect of oppression: that agents may be active participants in their own oppression. Strategies of resistance against oppression must take this into account in order to mitigate backlash and insure the possibility of social change.

More specifically, this ‘bottom up’ approach, on the one hand, better attends to how overlapping structures of oppression and ideologies present distinct—not just additive—instances of false consciousness that can also be dynamic and interactive. Agents, after all, may be located within of overlapping and interwoven systems of oppression (cf. Crenshaw 1991, Hutchinson 2001), thereby developing belief-sets involving multiple
ideologies. A theory of ideology alone is ill-suited for this task. To illustrate this, consider how, in Britain, the women anti-suffragists' widely circulated justifications for rejecting the vote reflected their varied social positions, where gender oppression overlapped with economic oppression as well as colonialism (much of the discourse also concerned women’s imperial responsibilities). Both anti-suffragists Mary Ward, who was a best-selling author, and Maud Ellen Simkins, a wage-earner, shared the ignorant belief that “women are domestic creatures” by nature and can never be political ones (Simkins 1909a, 14). Yet, for Ward, ‘domestic’ activities included “care of the sick and the insane; the treatment of the poor; the education of children,” confined by national borders (“An Appeal” 1889, 782). While, for Simkins, ‘domesticity’ was restricted to the upkeep of the household. Ward campaigned against suffrage for reasons such as the superfluity of the vote partly since “all the principal injustices of the law towards women have been amended by means of the existing constitutional machinery” (“An Appeal” 1889, 787). Conversely, Simkins publicly opposed suffrage because it would disproportionally affect women workers like herself, who would be saddled with “the triple burden of wage-earning, housekeeping, and political responsibility,” believing that the initial double burden of wage-earning and housekeeping was “because the wealthier women have been neglecting the first elementary duties of women that the poor have come to [squalor and discomfort]” (Simkins 1909b, 790–793). Both Ward and Simkins’ false consciousnesses did not result from just adding patriarchal and bourgeois ideologies together—there is mutual reinforcement between these ideologies in each specific overlap, presenting unique types of oppression and corollary false consciousnesses depending on the overlap. Our understanding of how gender ideology worked in Britain thus cannot just involve a ‘top down’

44 Overlapping and interwoven systems of oppression are often approached, depending largely on one’s disciplinary orientations, in terms such as, say, intersectionality (Crenshaw 1991) or multidimensionality (Hutchinson 2001). The basic idea is that oppression often compounds non-additively and uniquely. For the purposes of this thesis, I do not take any position on which is the best approach to theorise these overlaps. For more recent discussions see Cho, Crenshaw, McCall 2013, Mutua 2013, Curry 2021.
accommodation of these diverging sets of beliefs. While both sets of beliefs might follow from the principle of feminine self-abnegation, this alone is insufficient to account for the conflict between the false consciousnesses of Simkins and Ward.

The need for the ‘bottom up’ analyses of false consciousness also comes from another growing trend in analytic social philosophy that (following Lukács) approaches the possibility of emancipation from how collectivising epistemic resistance may arise from among the oppressed (see, for examples, Goodin and Spiekermann 2015, Jugov and Ypi 2019, Toole 2019, Pohlhaus Jr. 2020). This is because, as standpoint theorists and epistemologists of resistance are keen to emphasise, “any account seeking to remedy structural injustice in ignorance of the standpoint of the oppressed would be both paternalistic and more likely to err,” even reinforcing the oppressive relations (Jugov and Ypi 2019, 11, cf. Collins 1989, Wylie 2003). Some such philosophers in this trend see this as even involving how structural oppression is to be resisted and challenged under “ideological distortion and false consciousness” (Jugov and Ypi 2019, 9). If the goal of critique is to conduce the emancipation of the oppressed, for them to overcome their alienation in the social world and in relation to themselves, critique must not only attend to the problems of ideology but also attend to how exactly individual epistemic agencies is to be restored for agents under false consciousness.46

It is thus for these reasons of overlapping and interacting systems of oppression, as well as attending to the particular agencies of oppressed individuals, that I do not treat ‘false consciousness’ and ‘ideological

45 Tamara Jugov and Lea Ypi explicitly set out to determine political responsibility among oppressed agents that would hold even in cases of false consciousness (or ‘epistemic opacity,’ in their preferred words). I attend to a moral version of this when I deal with the Agential Problem in Ch. IV.
46 For more on the notion of ideology critique as such, see, for example, Shelby 2003, Celikates 2018, Jaeggi 2019, and various essays in Christ et al. 2020.
consciousness’ as interchangeable terms. I understand the concept of false consciousness (and corresponding critiques of false consciousness) as not a corollary of the concept of ideology (and ideology critiques) but an ancillary. In other words, in order to meet Mills’ aforementioned aspiration for the concept of ideology to allow for precision in attending to the social aetiologies of beliefs and multidimensional conceptualisations of ideologies, the concept of ideology in the social philosopher's toolkit must be accompanied by a concept of false consciousness. As to Michael Rosen puts it:

The theory of ideology presents an explanation of how, nevertheless, these societies can survive. In its classic form, that explanation has two parts. In the first place, the theory of ideology asserts that the acceptance of oppressive or otherwise illegitimate social orders depends upon (even if it is not exclusively the result of) a certain state of mind (a set of beliefs, attitudes, values or whatever) on the part of the oppressed. The second element is that this state of mind is neither a mere accident nor the intended outcome of conscious propaganda or manipulation on the part of the powerful. It is, in a way that obviously requires explaining, held to be an intrinsic (perhaps functional) characteristic of the social system in question that it should produce such states of mind. (Rosen and Wolff 1996, 210)

I understand the first element of ‘the theory of ideology’ to be the concept of ideology as focused on by recent analytic social philosophers (such as those of the New Ideology Critics). However, it is the ‘second element’ described by Rosen that is the concept of false consciousness, which has been neglected by these philosophers, and accordingly the focus of this thesis. At the same time, in attending specifically to how oppressed individuals might themselves be implicated in this ‘state of mind,’ the agentic idea of false consciousness advanced in this thesis may also be read as a complementary theory to contemporary theories of epistemologies of ignorance (see Mills 1997, Alcoff 2007, Martín 2020). That is, the concept of false consciousness offers systematic explanations as to how and why
particular oppressed individuals come to acquire and perpetuate such a ‘state of mind’ in their individual lived experiences—articulating the “[appropriate] causal route[s]” between “power relations and patterns of ideological hegemony” and the beliefs of particular individuals, which is often otherwise (but understandably) glossed over in the more general attention to the epistemic ignorance of agents qua members of oppressing groups or to the epistemic resistance of agents qua members of the oppressed groups (Mills 2007, 22; cf. Medina 2013, Pohlhaus Jr. 2020).

2.3. The Doxastic Form

Even though I have just shown that the alternative definitions are both unwarranted and obfuscate the critical utility of the concept of false consciousness, something should be said about an issue which I suspect underlies the motivation behind the adoption of either definition.

While I have defined the concept of false consciousness in terms of beliefs, uses of the concept of false consciousness have tended to include non-doxastic forms (as could be seen in §1) such as: perceptual, emotional, and volitional forms (see M. Rosen 1996; cf. Deleuze and Guattari 1972, Bourdieu and Eagleton 1992). Some or all these may no doubt overlap in various concrete cases, but in this thesis I have focused (and will be continuing to focus) only on doxastic false consciousness, given that most users of the concept largely understand them to minimally involve beliefs (see, for example, Meyerson 1991, Jaeggi 2009, Stanley 2015). In addition, much more recently has been said about non-doxastic forms of false consciousness, if the term is to be used at all (see, for examples, Jaggar 2005, Khader 2011). That said, the doxastic form of false consciousness is

---

47 I would like to thank my examiners for pressing me on this so as to clarify my contribution.
precisely a matter of controversy and often the reason for the concept’s dismissal *in limine*. This was very much the case for a large part of the post-Marxist traditions after the early Frankfurt School, with some in analytic social philosophy having also done the same.

With respect to these post-Marxist traditions, Jaeggi writes:

> The phrasing ‘false consciousness’ has been under attack from various sides (from Althusser to Zizek, theorists tried to overcome this exact aspect of ideology critique). But the reproaches that object to localizing the ‘wrongness’ of ideology ‘in consciousness,’ and those who see an outdated (viz. representationalist) epistemology at work, overlook that in this phrase ‘consciousness’ is quite ambivalent. It is almost misleading to speak of consciousness, since this consciousness, constituted by society and itself having a practical effect, is no longer consciousness in the traditional opposition to ‘being’ (or ‘superstructure’ in contrast to ‘base’) but is instead a structure in which the two are intertwined: a complicated network of norms, ideas, and practices affecting one another. Furthermore, ideology’s odd status between truth and untruth suggests that the concept of ideology introduces an understanding of true and false that does not conform to the traditional representationalist model. Thus, it is not at all understood that there is a true reality in the sense of a nonconstructed, not conceptually constituted reality waiting behind the false and distorting one. (Jaeggi 2009, 82–83n17)

I gather from Jaeggi’s remarks here that the post-Marxists’ worries about the concept of false consciousness—specifically when doxastically formulated—has to do with two assumptions: *methodological individualism*, where determining the problematic nature of false consciousness lies fundamentally in attending to an individual’s cognitions; and a *correspondence theory of truth*, where the truth and falsity of a given belief is determined by whether its attempt to represent reality accords with the
actual state of affairs. These worries are echoed by Sally Haslanger in her criticism of Tommie Shelby:48

Shelby suggests that ideologies are sets of beliefs held with false consciousness; therefore, the same set of beliefs held by one person may not be ideological, but when held by another might well be. This requires an individualistic approach: whether a set of beliefs is ideological is partly a matter of content, partly a matter of social function/effects, and partly a matter of how the individual(s) in question hold(s) them. (Haslanger 2017a, 9n8)

In the case of ideology, it isn’t just a ‘matter of chance’ that Smith and Jones (et al.) share their (ideological) beliefs. Note, however, that because their ideological beliefs are false or distorted, we can’t explain their convergence by reference to the truth. What explains why we are so systematically and enduringly mistaken? (ibid., 9)

Part of the worry is that if such assumptions are indeed true of the concept of false consciousness, attending to it would have little to no critical utility, since it can be addressed independently of social structures. Moreover, addressing in such a way “requires that we not question why the structural conditions are what they are but that we limit our questioning to states of affairs consistent with the structure”—and this indifference to the structure “ends up, in practice, supporting a proinstitutional bias” (Garfinkel 1981, 152). If correcting false consciousness simply involves attending to the ways in which the relevant individual’s particular cognitions fail to track truth without reference to their structural position, then the structure would not itself come under critical view and be left unproblematised. Another part of the worry is put by Serene Khader as follows, which points to a reason why some (though not all) philosophers opt to distinguish between ‘adaptive preferences’ and ‘false consciousness’ (cf. M. Rosen 1996, Jaggar 2005, Enoch 2020):

48 As will be clear later in Ch. II, I am in broad agreement with Haslanger’s criticism of a cognitivist approach and only am picking out her understanding of false consciousness here.
A number of philosophers describe adaptive preferences as a version of Marxist ‘false consciousness.’ To liken [“inappropriately adaptive preferences”] to ‘false consciousness’ has the effect of linking the notion of [adaptive preferences] to condescension to deprived people—given that the term ‘false consciousness’ has fallen out of favor among feminists. We tend to avoid the term ‘false consciousness’ because it suggests that deprived people are victims of an inability to distinguish between illusion and reality; calling [adaptive preferences] a version of ‘false consciousness’ rhetorically ties the notion of [adaptive preferences] to the belief that oppressed people are simple dupes of injustice. (Khader 2011, 136) [emphasis mine]

Yet, as Jaeggi notes in her earlier remarks, the worries pertaining to methodological individualism and the correspondence theory of truth are in fact inessential to the concept. As we have already seen in §1.3, Lukács had been keen to avoid a ‘psychologistic’ approach to false consciousness, instead attending to its constituent beliefs only insofar as it belonged to class consciousness: “the appropriate and rational reactions ‘imputed’ to a particular typical position in the process of production” (Lukács 1923, 51). In other words, Lukács’ analysis of false consciousness is already a structural one and not a methodologically individualistic one. Furthermore, even as he explicitly invokes terms such as ‘truth’ and ‘falsity,’ these had not been understood in terms of a correspondence theory of truth, but had “transferred its search for the criterion of reality to the realm of practice and particularly political practice” (Mannheim 1936, 84; cf. Cunningham 1983). That is, again as we saw, Lukács understood a given consciousness to be ‘true’ or ‘false’ in terms of whether a given class had knowledge of “the actions they need to perform in order to obtain and organise power,” inasmuch as they could relate immediate economic struggles to the ultimate revolutionary goal (Lukács 1923, 53). With the Frankfurt School, we also saw that the falsity of false consciousness was not understood in terms of correspondence with reality, since for them it was (social) reality itself that
was ‘false,’ as it were, due to how the extensive culture industry of late capitalist society forecloses the imaginative possibilities of how the extant social order could be otherwise.\footnote{Adorno, ever so cheerfully, puts this as follows:}

In a similar vein, while I remain focused on doxastic false consciousness, I do not adopt a methodologically individualist approach or a correspondence theory of truth. My analyses in this thesis, following the methodology of §2.2 broadly speaking, are \textit{structural}. At the same time, the reader might notice that I have understood the epistemic ‘falsity’ of false consciousness in terms of ignorant beliefs.\footnote{I thank Mona Simion for suggesting this articulation to me at the Glasgow MAP Workshop 2019 “Us and Them: Violence, Discrimination and Minorities.”} I proceed this way so as to allow that false consciousness may in fact include a belief that \( p \), where \( p \) is true (cf. Cunningham 1983, Shelby 2003). But such a belief, importantly, would still be understood as ignorant insofar as it falls short of knowledge: the belief would still be held if \( p \) were false and if it is based on unreliable belief-forming processes. In other words, a belief that \( p \) would still be an ignorant one because it would be epistemically \textit{insensitive} and \textit{unsafe}.\footnote{For a discussion of the differences between the view of ignorance as presented here, i.e. the absence of knowledge, and a view of ignorance that requires the absence of true belief, see Le Morvan and Peels 2016.} Insisting that false consciousness can only be constituted by false beliefs, and not include ignorant yet true beliefs, would mischaracterise the belief-sets of individuals under false consciousness both within the genealogy presented and the way the concept of false consciousness has been used in social philosophy today. For example, Lenin’s critique of trade union consciousness was not that it was epistemically false that textile workers needed better working hours when they decided to strike in 1896, but that it was based on the unreliable spontaneity of the workers.

---

\footnote{Adorno, ever so cheerfully, puts this as follows:}

\begin{quote}
We do not live in a revolutionary situation, and actually things are worse than ever. The horror is that for the first time we live in a world in which we can no longer imagine a better one. (Adorno and Horkheimer 1956, 72)
\end{quote}
I suspect that the lack of clarity with regard to the space of theoretical possibilities of the concept of false consciousness has mistakenly lead those philosophers who have nevertheless found the concept appealing to the alternate definitions. Defining ‘false consciousness’ as ignorance of the basis of one’s beliefs, for instance, does not require a correspondence theory of truth. And subsuming questions concerning false consciousness under those of ideology ensures that the approach to the phenomena avoids methodological individualism. However, if these definitions are indeed motivated by the worries I have suggested, this would be an overreaction that would be unfortunate in two ways: as I have suggested, the alternative definitions obfuscate the critical utility of the concept of false consciousness, and the aforementioned worries about the doxastic form are hardly substantial worries in light of those I will consider in the next section.52

3. The Problems

I will now go on to consider the more substantial problems that we have seen arising in the course of the conceptual genealogy and which had also been picked up on by the detractors of the concept of false consciousness. In §1, I highlighted a number of issues at each historical shift in the meaning of the concept of false consciousness. These, as I will go on to show, all fall under three distinct problems commonly thought to plague the concept: the Explanatory Problem, the Normative Problem, and the Agential Problem. These problems must be addressed if the concept of false consciousness is to perform the work that social philosophers want for it to do in their

52 The same suspicion, I think, can be had about the abandonment of the concept of false consciousness in favour of, say, adaptive preferences (e.g. Khader 2011). Although it is likely that the Marxian associations of the former (and political prejudices) are partially responsible too.
analyses of structural oppression. I will now give a brief sketch of them, before I attend to them in greater detail in the subsequent chapters.

3.1. The Explanatory Problem

The first and most fundamental problem, which I term ‘the Explanatory Problem,’ is that the concept of false consciousness lacks an adequate general theory and bottoms out as an ad hoc explanation for particular behaviour. Specifically, there is a lack of a satisfactory standard explanation for how individual cognition and social schemas generally interact within oppressive practices, such that agents systematically believe propositions against their own interests. As such, the concept is generally deemed to be theoretically ill-suited for systematic research and emancipatory praxis beyond particular, one-off cases (at best). To recall Rosen’s formulation:

[False consciousness] is neither a mere accident nor the intended outcome of conscious propaganda or manipulation on the part of the powerful. It is, in a way that obviously requires explaining, held to be an intrinsic (perhaps functional) characteristic of the social system in question that it should produce such states of mind. (Rosen and Wolff 1996, 210)

The concern here is not just the content of a given ideology and how it manifests in social practices (which is a concern shared by the concept of ideology), but also the processes rendering an individual agent’s lived ideological experience ‘natural’ and how they might be avoided or overcome by such an agent. This subsumes: the worry arising from The German Ideology, in terms of the underspecification of the relation between the economic asymmetry and the asymmetry in the production of ideas; the first worry arising from our consideration of Lenin and Gramsci, in terms of how to account for the extent of false consciousness; and the first worry arising from our consideration of Lukács, in terms of how immediate beliefs relate to the social totality. It is precisely in the presence of a lack of an explanation
which addresses all these worries that the concept of false consciousness has been haunted by the worries of methodological individualism and the correspondence theory of truth.\footnote{Cf. Lukács’ comment that the dialectical method does not permit us simply to proclaim the ‘falseness’ of this consciousness and to persist in an inflexible confrontation of true and false. On the contrary, it requires us to investigate this ‘false consciousness’ concretely as an aspect of the historical totality and as a stage in the historical process. (Lukács 1923, 50) [emphasis mine]}

There are, to my mind, two main reasons for why it is generally held that there has not been an adequate explanation for false consciousness, within the context of analytic social philosophy, even now. First, the most prominent attempts to address this problem by the Analytical Marxists have been found to be dissatisfactory in two ways. The attempts (such as Cohen’s \textit{Karl Marx’s Theory of History}) were either ontologically excessive in positing society as a self-maintaining entity over and above individuals, so as to explain why false consciousness occurs (as a means to maintain such a society)—thereby conflating the questions of ideology and false consciousness. Or the attempts were excessively individualistic (such as Elster’s \textit{Sour Grapes}), being too focused on individual psychology (such as wishful thinking) to reveal anything informative about structural oppression on their own—thereby realising the aforementioned worry about methodological individualism (see M. Rosen 1996).\footnote{In a personal correspondence with Rosen in 2016, he admits that this is a view he still holds.} This dissatisfaction was rooted in the assumption that the concept of ideology either exhaustively explains false consciousness (cf. Eyerman 1981), or, failing that, false consciousness “is not systematically explained by the nature of the society in question” (Rosen and Wolff 1996, 225). But inasmuch as we are concerned with false consciousness as a phenomenon that can sometimes be resisted (even with the Frankfurt School’s pessimism, at least in thought),
we cannot be satisfied with an appeal to ideology alone as an all-or-nothing explanation.

The second reason that extant attempts (especially those from coming from ideology critique) have tended to only focus on static, seemingly self-contained, singular types of oppressive schemas: for example, bourgeois ideology in Marxism, or patriarchal ideology in feminism. However, as J. M. Balkin argues:

[T]he notion of false consciousness is problematic because it is a holdover from the bivalent oppressor-oppressed model of hegemony that I have just criticized. This model makes little sense in a world in which people have multiple and cross-cutting identities. Even assuming that African Americans and women have objective interests as a class, surely these interests can sometimes conflict. When they do, how can an African-American woman avoid a charge of false consciousness, regardless of the position she takes? Indeed, accusations of false consciousness are often attempts by one portion of a social group to assert unitary and objective interests that disadvantages or ignores the claims of another portion or subgroup. Working-class women may be accused of false consciousness by middle-class women when in fact their interests differ because of their class position. Similarly, the interests of African-American women may diverge in important respects from those white women. Once internal divisions and cross-cutting identities are recognized, the notion of false consciousness threatens to become incoherent or at best self-serving. (Balkin 1981, 117–118)

That is, for a general theory of false consciousness to be coherent, it has to account for how dynamic and interacting systems of oppression affect agents at their overlaps differently (Curry 2017b, Ásta 2019). It is crucial to recognise both the practical and theoretical inseparability of social categories (for example, ‘Black man’) and the various injustices that are occasioned by one-sided analyses (cf. Curry 2017a, Finlayson 2018, Jenkins 2020). With regard to the latter, as Sylvia Wynter warns, “if all modes of
resistance are defined as a struggle against exploitation one form of struggle becomes hegemonic, i.e., the labor mode of resistance. This mode then defines the targets, methods, places, and instruments of confrontation” (Wynter 1982, 34).

3.2. The Normative Problem

The second problem is a Normative Problem (which is particularly exacerbated by the Explanatory Problem). A condition on a successful critique of false consciousness—indeed, of the very coherence of the concept to begin with—is the critic’s ability to answer to their addressee what exactly makes instances of false consciousness problematic (cf. Hall 1986, MacKinnon 1989, Stanley and Wise 1993, Balkin 1998). The Normative Problem, however, raises the question as to which norms a critic may appeal to justify the ‘falseness’ of false consciousness. As we saw, the concept is often analysed as a set of ignorant attitudes that an agent acquires and maintains (epistemic ‘falsity’) in virtue of their oppressed position in a given system of oppression (genetic ‘falsity’), such that they are also thus motivated to perpetuate such a system (functional ‘falsity’). From this, we might observe false consciousness to be straightforwardly bad inasmuch as these three constituent features are problematic.

But agents under false consciousness are precisely such that they do not find their beliefs and experiences of the world, others, and themselves to be ignorant or oppressive. With varying degrees of sophistication, they may even understand and experience their (actual) oppression positively as well as empowering—sometimes regarding their critics as themselves under false consciousness. Women anti-suffragists, for example, argued that such participation in parliamentary politics was antithetical to the feminine virtues that ennobled them.
Steven Lukes observes that the concept of false consciousness is closely associated with others, notably that of ‘real’ or ‘true’ or ‘objective’ interests, that is, of interests that false consciousness supposedly conceals from those whose interests they are. Those who object to this answer generally do so on two distinct, even opposite, grounds. The first, more traditional objection is that these concepts suggest an arrogant assumption of superior knowledge, an assumption notably embedded in the Marxist tradition—a claim to privileged access to what is ‘correct,’ a claim theorized by Georg Lukacs and well exemplified by Leninists, Trotskyists, Stalinists, and Communist Party apparatchiks across the decades of the twentieth century—and a corresponding disposition to treat people as cultural dupes. The second objection is more recent. The suggestion, commonly associated with postmodernist thinking, is that there cannot be false consciousness since there are multiple true consciousnesses—socially constructed ‘regimes of truth,’ generated and sustained by power. On this view, to impute false consciousness is mistakenly to believe that there even could be a correct view that is not itself imposed by power. (Lukes 2011, 19)

I interpret these objections to mean that there are the following two methodological worries for any potential answer to the Normative Problem. First, the concept of false consciousness seems to impose evaluative standards that are alien and alienating to the agents being criticised. Relying on such standards in critique are counter-productive for the solidarity-building which is necessary for emancipatory praxis and would often lead to backfire effects from the criticised. This can be seen in the anti-suffragist reactions to suffragist criticisms of the Appeal in the Nineteenth Century: thousands more signatures were made in support of the Appeal (which had already been undersigned by a hundred and four women) in the issue following that of the criticisms (Sutherland 1990, Bush 2007). Related psychological studies have also been carried out with respect to failures of
corrections of political misperceptions (see, for example, Nyhan and Reifler 2010).

Second, there is a sceptical worry that the critic’s normative standards themselves are a function of another ruling ideology or themselves belong to another type of false consciousness (Mannheim 1936, 57). Both philosophers of epistemic oppression and resistance have observed that critics themselves do not always have access to an adequate body of knowledge in virtue of their social position (see, Mills 2007, Dotson 2014). For example, the arguments of the leading suffragists often reflected their positions as largely upper- and middle-class white women. Both the Women’s Social and Political Union (WSPU) and the National Union of Women’s Suffrage in Britain, for various reasons, campaigned to have the same voting rights as men prior to the Representation of the People Act 1918—that is, voting rights on the basis of property qualification. These reasons given were often, while not always, classist and imperialist. As Sylvia Pankhurst dejectedly records, the East London suffragettes were expelled by Emmeline and Christabel Pankhurst from the WSPU because they thought that “a working women’s movement was of no value” (Pankhurst 1931, 180).

This subsumes: the second worry arising from our consideration of Lenin and Gramsci, in terms of whether the source of emancipatory consciousness was internal or external to the oppressed; and the worries arising from our consideration of the Frankfurt School, in terms of how there is conceptual space for critique and who is to make it. Indeed, this very concern regarding the norms of critique in relation to the social position of the critic is often thought to be definitive of the Critical Theory tradition, which is animated by the problem of how to “[overcome] the one-sidedness that necessarily arises when limited intellectual processes are detached from their matrix in the total activity of society” (Horkheimer 1986, 199).
3.3. The Agential Problem

The third problem is an Agential Problem. Considerations of how to theorise about false consciousness have always avoided the question of the moral responsibility of oppressed agents with some unease (see, for example, Shelby 2007). This is because the use of the concept has been often seen as itself amounting to victim-blaming by conceptual fiat. Here, victim-blaming is understood to refer to blaming practices that focus attention inappropriately on victims in accounting for the relevant harms (Harvey 1999; cf. Eister 1985, 22n2, Matthews 2014). As a critical concept, false consciousness is in fact a means to account for the relevant harms of oppression by taking (at least some) oppressed agents themselves into consideration. Part of why the concept of false consciousness might be thought to bring inappropriate attention to oppressed agents is because it would, allegedly, detract attention from underlying unjust structural conditions and the oppressors:

[A]ppeals to the idea of false consciousness, popular a few years ago, suggested that the problem about capitalism was a problem about workers. To reduce the question of capitalism to deficiencies in workers’ consciousness diverts attention from the capitalist, the other participant in the employment contract. Similarly, the patriarchal assumption that prostitution is a problem about women ensures that the other participant in the prostitution contract escapes scrutiny. (Pateman 1999, 56)

I take the worry here to be that, within a context of asymmetrical attention, the concept of false consciousness detracts from the participation of the oppressors within the relevant social structure. But given the aforementioned attention to the ideological consciousness of oppressors in recent analytic social philosophy, and that the discussion to the concept of false

---

55 As opposed to using ‘victim-blaming’ as a charge against someone having problematic motivations in their criticism of an oppressed agent (see Manne 2018).
consciousness is ancillary to these, this is an insubstantial albeit fair concern.

But another part of why the attention brought by the concept of false consciousness is thought to be inappropriate—and the Agential Problem more precisely formulated—is the general worry that agents under false consciousness have an otherwise absent or diminished moral agency (cf. Jost 1995). This subsumes the second worry arising from the bootstrapping problem we found in Lukács, in terms of how we can be justified in expecting emancipatory consciousness to develop among the oppressed, given the deleterious effects of oppression on them. Despite disagreement over whether oppressed agents may be blameworthy for acting in ways that contribute to their own oppression, most feminist philosophers are in agreement that false consciousness would exculpate simpliciter (see Superson 1993, Cudd 2006, Hay 2013). The assumption behind this, put in Strawsonian terms, is that such an agent systematically fails to meet the relevant rationality and epistemic conditions of moral blameworthiness due to ideological distortions of reasoning or knowledge practices. That is, being reasons-unresponsive or ignorant, the status of someone under false consciousness as a responsible moral agent is thereby undermined.

4. Approaching the Problems

4.1. Overview of Arguments

Having enumerated the problems above, the overall structure of the thesis is modular and corresponds to an attempt to address of each of them in turn. I briefly outline my approach to each below.
Chapter II will prove foundational in addressing the subsequent concerns inasmuch as it provides a workable general theory of false consciousness. I argue that the Explanatory Problem persists in the literature because of a lack of clarity about the explanatory desiderata involved for any account of false consciousness. I identify four explanatory desiderata: belief acquisition, content prevalence, resistance, and systematicity. Lack of attention to all these leads to the recurring inability to provide a satisfactorily general account of the systematic nature of false and emancipatory consciousness—which are conceptually inextricable from each other. To resolve this, I turn to critical social ontology to provide a social constructionist account of false consciousness as a regulative concept for social research. Such an approach can be shown to satisfy these by understanding the concept of false consciousness as regulating social research, rather than as determining the exact mechanisms for all instances: the concept attunes us to a complex of mechanisms conducing oppressed agents to mistake social understandings of themselves as natural self-understandings—the possibility of resistance lies where these overlap, or are entirely absent. These mechanisms converge upon an individual to produce beliefs that may be shared with other oppressed agents when seen from the perspective of a single system of oppression. Yet at the same time, such beliefs are nuanced differently at more agent-targeted levels and in relation to other systems.

In Chapter III, I provide a prudence-based account of the Normative Problem, arguing that the Normative Problem persists because extant approaches to well-being thus far are not sufficiently attentive to its fundamentally social dimensions and how it is embedded within forms of life. These approaches are either realist or anti-realist: respectively, they either determine what is ultimately good or bad for any welfare subject in terms of an attitude-independent fact(s), or in terms of a fact(s) about the attitudes of the subject. But, as I will argue, such approaches to providing standards of critique for false consciousness either fail to provide internal
motivation for emancipation or reproduce epistemic oppression by imposing alien, external standards. An immanent approach is thus advanced as a possible way forward: to articulate the relevant prudential standards of critique in terms of the adaptational nature of organisms within their forms of life. I explain how this is grounded in an enactivist understanding of the participatory sense-making that goes on between an organism, other organisms it is coupled with, and the dynamical systems they form. That is, an organism, as itself a dynamical system, issues its own prudential standards of critique in terms of what does not facilitate the coordination of both its parts and the systematic whole.

In Chapter IV, I argue that the Agential Problem persists because it is simply assumed that an individual under false consciousness systematically fails to meet the relevant rationality and epistemic conditions due to structural distortions of reasoning or knowledge practices. However, I attend specifically to the case of women anti-suffragism, and argue that some of them, despite false consciousness, satisfy both rationality and epistemic conditions for ordinary blameworthiness. I contend that the irrationality charge only holds at the group-level and that these women are culpably ignorant on account of culpable belief-forming mechanisms. Hence, false consciousness does not _eo ipso_ render agents inappropriate targets of blame.

In Chapter V, I conclude with a remaining puzzle about the relation between moral blame and a critique of false consciousness. Blaming an individual under false consciousness involves incompatible modes of cognising others: social criticism cognises agents socio-theoretically, as objects of investigation; while moral criticism cognises agents interpersonally, as fellow members of the same community. The application of the concept of false consciousness in moral blame by the critic also redoubles oppression in theoretically fixating the individual under an oppressed social category in an
asymmetrical economy of representation. Blaming an agent under false consciousness thus seems to be self-frustrating for the aims of both moral and social criticisms.

4.2. Methodological Note

As a methodological note, although I draw from a wide range of oppressive contexts (with hopefully sufficient care not to conflate them) for a general theory, I return constantly to the historical case of women anti-suffragists as a relatively uncontroversial instance of false consciousness. And a caveat is in order with regard to my approach to the archive of women anti-suffragism (especially since I am engaging in some of the archival work myself). Important historical work has been done to show how arduous, complex, and often disharmonious it was for these anti-suffragists at the time to formulate and publicise their opposition to the suffrage campaign over the decades (as it also was for the suffragists). As such historians have cautioned, our various understandings of the feminist/anti-feminist distinction today also do not easily map onto the suffragist/anti-suffragist distinction, not least because the term ‘feminism’ was inconsistently deployed and could also be applicable to many anti-suffragists. Indeed, a significant number of anti-suffragists did not see their opposition to suffrage as a maintenance of a status quo in gender relations, but rather as directing women’s progress out of the household towards more appropriate social, yet non-political, roles befitting of their gender—for example, education and healthcare (see, for example, Delap 2005, Heilmann and Sanders 2006, Crozier-De Rosa 2018).

56 Such expansion of the domestic also attracted criticisms from within their working-class ranks (I consider such a case, that of Simkins, in Ch. III).
My approach in this thesis is not inconsistent with these important observations, which deserve careful attention in their own right, but my interests are nonetheless primarily philosophical. I mostly attend to the discourses and circumstances of the Appeal and any surrounding developments only to the extent that they help us elucidate the nature of false consciousness. The narrow focus here on how we may understand particular anti-suffragists to be under false consciousness in the context of particular actions, such as undersigning the 1889 Appeal, is thus distinct from the broader and far more complex questions about how we should judge them (or women anti-suffragists in general) for hindering or promoting women’s, class, able-bodied, or imperial interests on the whole.

Despite its historical and localised nature, the case of the women anti-suffragists provides a useful heuristic for clarifying the concept of false consciousness more generally across other systems of oppression and where they overlap and interact (for example, ableist, economic, gendered, racial systems). There are a few reasons why this case study is particularly amenable to this end.

First, the reasons of both anti-suffragists and suffragists, as educated and influential women, were well-articulated and documented in fictional and non-fictional works and correspondences. In this thesis, I especially emphasise self-authored documents as much as possible.

Second, the principled, planned, and protracted efforts of a significant number of the anti-suffragists against suffragism unfortunately echoes among a significant number of women anti-feminists arguing for the political exclusion of women on the basis of natural differences even today (such as, as we saw at the start of this chapter, the cases of alt-right women like Davenport and Bloomfield).
Third, there is near-universal agreement today that women’s suffrage is normatively desirable, irrespective of the difficulties that many women might nevertheless still face in exercising it. This is particularly important for us as ethicists and social philosophers for purposes of having a common starting point to inquire into the concept of false consciousness.

Fourth, for better or worse, the dominant discourse of the 1889 Appeal was largely confined to one to two systems of oppression: gender and economic. This has, of course, been a longstanding problem with British suffragism, as well as its historiography, which has recently come under greater scrutiny within and without academic circles (Delap et al. 2006, Caroll 2015, Crozier-De Rosa 2018, cf. McConnaughy 2013): for example, the historiographical exclusion of Sophia Duleep Singh from accounts of the suffragette movement (Anand 2015), the overlooked recruitment of many suffragettes to the British Union of Fascists (Caldicott 2017), or the little attention paid to how the New Woman novel emerged from the context of imperial South Africa (Free 2016 cf. Crozier-De Rosa 2010).

Yet, the focus of this thesis on the gender and economic dimensions of the 1889 Appeal in particular is not to say that other (often overlapping and interrelated) systems of oppression (for example, colonialism) should be neglected in general (see Beisel and Kay 2004, Curry 2017b), but only that they are not as salient for the very particular instances of false consciousness under consideration here. That is, as I will go on to argue, even while circumscribed by the British (white) women anti-suffragists’ own (limited) imaginaries in the 1889 Appeal, a conception of false consciousness can be coherently articulated as well as critiqued—and they can even be blameworthy under it. Nevertheless, where relevant, I will attempt to provide some analysis of the racial, imperial, and evolutionary
aspects of race and womanhood in situating the claims deployed by the
suffragists and anti-suffragists over the right to vote in Britain.\textsuperscript{57}

Thus, with the historical example of anti-suffragists, I hope to provide a
relatively tidy basis upon which to examine the problems of false
consciousness as such—which social philosophers may then apply to more
overlapping, complex systems of oppression (for example, women anti-
feminism in the colonies, or racism within queer communities in postcolonial
Singapore).

\textsuperscript{57} I thank my examiners for pressing me to expand a little more on this, in addition to the
extant abstract analysis in Chapter III of how such aspects would come together.
II. Why Do They Know Not What They Do?  
A Social Constructionist Approach to the Explanatory Problem of False Consciousness

CHAPTER ABSTRACT

False consciousness requires a general explanation for why, and how, oppressed individuals believe propositions against, as opposed to aligned with, their own interests, in virtue of their oppressed status. This involves four explanatory desiderata: Belief acquisition, Content, Scope limitation, and Systematicity. Previous approaches to providing an explanation for false consciousness—namely, functionalism, psychologism, and processualism—have failed to take all the desiderata into account. In this chapter, I argue for a social constructionist approach that satisfies the desiderata by understanding the concept of false consciousness as regulating social research, rather than as determining the exact mechanisms for all instances. The concept attunes us to a complex of mechanisms conducing oppressed individuals to mistake social understandings of themselves as natural self-understandings—the limits lie where these sometimes overlap, or are entirely absent. The social constructionist approach here also subsumes the previous approaches and supplements their weaknesses.

I will now focus on the first and most fundamental of the problems outlined in the previous chapter (Ch. I, §3): the Explanatory Problem. The Explanatory Problem of false consciousness, to recall, is the lack of a satisfactory standard explanation in social ontology and epistemology for how individual cognition and social schemas generally interact in (sometimes overlapping and interacting) oppressive practices, such that agents systematically
believe propositions against their own interests. What is required of such a
general theory is also for it to account for how dynamic and interacting types
of oppression affect agents at their overlaps differently.

To address these concerns, I propose a social constructionist approach here
to the Explanatory Problem as an exercise in critical social ontology (cf.
Renault 2016). The approach brings together social construction and Critical
Theory to better satisfy the theoretical and emancipatory requirements of
social analyses hitherto overlooked in previous approaches—especially in
cases of overlapping oppression. Following recent directions taken by social
creationists like Sally Haslanger, I draw on Louis Althusser’s discussion
of ideology in On the Reproduction of Capitalism (1995) to understand the
concept of false consciousness as allowing us to attend to the various
stabilising mechanisms (sometimes called ‘distortion mechanisms’) in
structural oppression that conduce the oppressed to mistake socially
constructed understandings of themselves as natural self-understandings.58

In what follows, I first lay out the explanatory desiderata that any such theory
is required to satisfy for critique (§1). I evaluate three theoretical approaches
against how they satisfy these desiderata—approaches taken as providing
explanations for false consciousness in analytic social philosophy (§2):
functionalism, which explains it in terms of its overall societal function;59
psychologism, which explains it in terms of oppressed agents’ psychological
failures; and processualism, which explains it in terms of the oppressed’s
epistemic deference to oppressors. While these are shown to be inadequate
for satisfying the desiderata on their own, as opposed to the social

58 That is, in spite of both Haslanger and Althusser’s aversion to the concept of false
consciousness—which is unwarranted, as I have argued in Ch. I, §2.
59 ‘Functionalism,’ as pertaining to structural oppression, is used here in the prevailing
sense in social philosophy, following G. A. Cohen (Cohen 1979); for a different use of the
term not discussed here, see Thompson 2017.
constructionist approach, I ultimately suggest that they can all be subsumed under the latter (§3).

It should be underscored that social constructionism is not meant to be a new approach as particular accounts of false consciousness go: historical precedences include David Hume’s *Natural History of Religion* (1757), José Rizal’s *The Indolence of the Filipino* (1890), and Simone de Beauvoir’s *The Second Sex* (1949). What is advanced here, rather, is a general approach to theorising about false consciousness that meets the explanatory desiderata needed by the social philosopher, whichever type(s) of oppression is analysed. In Althusser’s words, the project we are concerned with here is ancillary to “the project of a theory of ideology in general, as opposed to a theory of particular ideologies, considered either with respect to their regional contents (religious, moral, legal, or political ideology, and so on) or class orientation (bourgeois, petty-bourgeois, proletarian ideology, and so on)” (Althusser 1995, 174). Resolving the Explanatory Problem would thus be a fundamental step for rehabilitating the concept of false consciousness as a whole, on the basis of which the subsequent Normative and Agential Problems may be addressed later.

1. **Explanatory Desiderata**

In setting out a general theory of false consciousness, we must first clarify the necessary and sufficient conditions for any explanation of any instance of the phenomenon. I will now lay out the explanatory desiderata that such a theory must satisfy, along with why and how it should do this. An adequate explanation for any instance of false consciousness must satisfy the following:

---

60 Interestingly, Michael Rosen (1996) points out that Hume was an indirect influence on Marx via Charles de Brosses’ *Du culte des dieux fétiches* (1760).
**Acquisition:** How affected oppressed agents come to form and maintain certain ignorant beliefs as (putative) knowledge.

**Content:** How the content of such beliefs are formed and maintained in the structural ambits within which such agents are embedded;

**Scope:** How other oppressed agents may nevertheless avoid, or even resist, forming or maintaining such beliefs.

Any explanation of false consciousness must not only attend to the processes resulting in the acquisition of ignorant beliefs at the individual level in a way that resists counter-evidence (Acquisition) but also processes that promote ignorant content at the structural level (Content). And given the ameliorative aim of critique, resistant epistemologies among oppressed groups, and how infrequently we notice false consciousness among the oppressed, such an explanation must also involve accounting for how an absence of false consciousness is nevertheless possible despite the aforementioned processes (Scope).

Moreover, there is a ‘meta-requirement’ for any explanation for false consciousness (following Lukács):

**Systematicity:** An account of false consciousness must explain the previous three desiderata as a systematic unity.

In a systematic unity, components are interconnected in a manner grounded by a single principle to form a whole. Within the context of critique, the limited acquisition and content of false consciousness should be identifiable
as belonging to a social structure (or overlapping and interacting structures).\textsuperscript{61} This would mean being able to map out mutually implicating beliefs within a particular oppressive schema (or schemas), allowing us to characterise them as ideological and needing amelioration. For example, an explanation of a belief-set involving ‘women are naturally deficient in political deliberation’ or ‘women should be married’ would be interconnected within the patriarchal schema of Victorian Britain’s social structure, grounded by a principle of feminine self-abnegation (as suggested by Mill). Whereas a belief like ‘women inhabit the second planet from the Sun’ might be connected with the previous beliefs by the principle ‘$x$ pertains to women,’ but it is neither mutually implicating nor relevant for critique.

There cannot be any critique with explanatory patchworks that satisfy all three conditions but have them obtaining together only by sheer happenstance. That is, it is not enough to just say that \textit{Acquisition} and \textit{Content} often coincidentally operate together and that \textit{Scope} can be explained insofar as such operations are absent. To illustrate this, consider the particular case of Mary Ward’s false consciousness:

\textbf{WARD:} Mary Ward has the ignorant beliefs that women are naturally deficient at political deliberation and that suffrage is unnecessary for addressing injustices specific to women [\textit{epistemic ‘falsity’}]. These beliefs motivate her to pen and publish the influential manifesto, “An Appeal Against Female Suffrage” [\textit{functional ‘falsity’}]. Ward has these beliefs owing to a general absence of women in Victorian politics and also her being unfamiliar with how legislative practices more than education still constrained working-class women thus far [\textit{genetic ‘falsity’}].

\textsuperscript{61} cf. Hannah Arendt’s understanding of ideology as “quite literally what its name indicates: it is the ‘logic of an idea’” (Arendt 1976, 469).
A possible (fictional) patchwork explanation for this, satisfying all the explanatory desiderata but Systematicity, might be:

*EXPLANATION*: Ward has these beliefs because, as a child, she picked up a note in school stating that ‘women are naturally deficient in political deliberation’ and that ‘suffrage is unnecessary for addressing injustices specific to women.’ She had no opinion on the matter before this encounter because these were not in the school’s curriculum; she maintains them because her childhood memories of school becomes sacrosanct for her in face of rapid urbanisation [Acquisition]. The note was in the school because someone scribbled it in a sudden religious epiphany and it is maintained because everyone who picked up the note just tosses it on the ground again [Content]. And those who did not attend the school did not pick up the note, thereby not forming and subsequently maintaining the beliefs [Scope].

It is clear from this that, even if *EXPLANATION* were true, it would fall short of the social philosopher’s aim of bringing one’s attention to the processes that systematically operated among similar anti-suffragists: the explanation, while it might be satisfactory for explaining *WARD* in the context of such a biography, is irrelevant to structural oppression, as the details attended to do not highlight structural factors that contribute to the epistemic, functional, and genetic ‘falsity’ of her beliefs: for example, the reason for her school’s curriculum design, the reason for the epiphany’s anti-suffragist content, or the reason why suffragists were not generally in her school. *EXPLANATION* could also apply to a wage-earner like Maud Ellen Simkins in a way that provides no insight into how Ward’s bourgeois position is related to her position under patriarchy. It also does not account for how her beliefs and actions converged with the other bourgeois anti-suffragists. Over 420,000 mainly upper- and middle-class women signed an anti-suffrage
parliamentary petition organised by the Women’s National Anti-Suffrage League. And modern historians gather that most working-class women then “had little interest in the vote before the war, and only limited interest once it had been finally placed in [their hands]” (Bush 2007, 189, 311). How Acquisition and Content are related in the cases of these two classes of women anti-suffragists, and how their cases are related to each other, thus cannot be a matter of mere statistical coincidence. Such operations were not just regular but schematic: affected agents regularly accept or reject evidence according to these schemas governing false consciousnesses. For a theory of false consciousness to contribute to the diagnostic and ameliorative tasks of critique, it must therefore show how these sets of beliefs are interconnected within ideologies of structures of oppression.

Furthermore, for an explanation to be critical, it is expected to also involve Scope systematically for at least two reasons. First, critical theorists often hold that motivating and normative reasons for ideological resistance are continuous with, and particular to, a given instance of ideology (Jaeggi 2019). The content of these reasons are to arise from analyses of the corresponding social structure, instead of it being just a matter of applying universal normative criteria in evaluating the ideology. Second, irrespective of this, if we hold that emancipatory hope is to be found largely in solidarity among the oppressed, we need to account for how such cases diverge for the sake of differential emancipatory strategies. This was reflected in the suffragist arguments directed to women, which had varying degrees of efficacy depending on their audience’s class: they emphasised how extending suffrage would ameliorate working-class conditions and how many upper- and middle-class women already had one foot in politics via organisations like the Primrose League.

---

62 I say more about what this continuity means for false consciousness in the next chapter (Ch. III, particularly in §2.1).
2. Survey of Extant Approaches

As we saw in Chapter I (specifically §1.1), the go-to, Marxian starting point for theorising about false consciousness is found in *The German Ideology*:

\[ \text{EXPLANATION} \text{M:} \] The ideas of the ruling class are in every epoch the ruling ideas: i.e., the class which is the ruling material force of society is at the same time its ruling intellectual force. The class which has the means of material production at its disposal, consequently also controls the means of mental production, so that the ideas of those who lack the means of mental production are on the whole subject to it. The ruling ideas are nothing more than the ideal expression of the dominant material relations, the dominant material relations grasped as ideas; hence of the relations which make the one class the ruling one, therefore, the ideas of its dominance.
(MECW 5: 59)

That is, the relevant beliefs are formed and maintained by oppressed agents because they are subject to the ruling intellectual force (*Acquisition*) and that the content of these beliefs prevail because the ruling class control the means of mental production (*Content*). These are systematically related inasmuch as they are grounded in a material asymmetry between the ruling class and the ruled class (*Systematicity*).

\text{EXPLANATION} \text{M alone, however, is under-elaborated (as Engels himself admitted) and insufficient for critical work (cf. M. Rosen 1996, Leiter 2007). Much more elaboration would be needed if it were to adequately explain} \text{WARD. Not only does it miss out on Scope but it also does not sufficiently address Acquisition and Content. In the case of Acquisition, what does it mean for someone to be ‘subject’ to ideas in a way that they would accept them as true? Just because the ideas of the ruling group are ruling does not on its own guarantee that the ruled would positively internalise them on a systematic basis—frustration and resignation are often responses to being confronted with ideas opposing one’s interests. But in some cases, we even} \]
find oppressed agents themselves taking up the task of generating and propagating highly nuanced and abstract beliefs (for example, women anti-suffragists like Ward were at the forefront of the Victorian literati). With respect to Content, what exactly are the means of mental production and why should the ideas produced correspond to the asymmetry in material relations? It is not clear why it should be the case that oppressive ideas (for example, ‘women are naturally deficient at political deliberation’) rule instead of non-oppressive ones (for example, ‘women have equal capacity for political deliberation’). Furthermore, it cannot account for cases where there are no clear material inequalities between the groups: Ward, apart from being a member of the bourgeois class, received enormous royalties from her novels; Robert Elsmere, the best-selling novel of the 19th century, made her a millionaire-equivalent for her time.

*Functionalism, psychologism, and processualism* are often seen as candidate elaborations of *EXPLANATION* and, in effect, constitute different approaches to answering the Explanatory Problem. Criticisms of functionalist and psychologistic approaches specific to the concept of ideology have had a long run since the heyday of Analytical Marxism (see M. Rosen 1996). Instead of revisiting them, I only focus here on how these approaches, alongside processualism, fail to fully satisfy the desiderata for an adequate explanation of false consciousness.

### 2.1. Functionalism

Functionalism is an approach to explaining false consciousness, and ideology more broadly, that is often associated with the controversial ontological assumption that society is a distinct entity that maintains itself through the false consciousness of the members of its constituent social groups (Cohen 1979, Shelby 2003). The approach mainly arose out of
attempts by Analytical Marxists—specifically, G. A. Cohen—to not only provide a scientific basis for the theory of ideology but also maintain as much fidelity to Marx’s works as possible. Still, the reifying assumption is unnecessary. What is crucial to this approach, as paradigmatically put forward by Cohen, is that false consciousness is functionally explained by the maintenance of the social arrangement of a society that is stratified across oppressive lines. By ‘functional explanation,’ the functionalist means that the function of an attribute of an object is cited as an explanation of the object having it—akin to the kinds of explanations that are found in the natural sciences. For example, to explain the melanism of peppered moths, it would be sufficient to cite melanism’s function of promoting species survival. This is not because it directly explains the attribute-possession but because giving the statement ‘A is functional for B’ as an explanation is adequate as a working hypothesis for why a particular object is A, without yet needing to provide an underlying, elaborating mechanism (like natural selection, in the peppered moths example). Functionalists thus think it sufficiently explanatory for false consciousness to cite its function of maintaining a given oppressive society.

The functionalist would explain WARD as follows:

\[ \text{EXPLANATION}^{FN} : \text{Ward has the ignorant beliefs that women are naturally deficient at political deliberation and that suffrage is unnecessary for addressing injustices specific to women, motivating her to campaign against women’s suffrage, because this is functional for the patriarchal and bourgeois organisation of Victorian Britain.} \]

\[ \text{EXPLANATION}^{FN} \text{ seems to be able to secure Content: the oppressive content of ideas (that ‘women are naturally deficient at political deliberation’) would be explained as those content that conduce the maintenance of an oppressive society (keeping women out of politics). It also secures} \]
Systematicity: the relation of Content to the social structure is purportedly secured by the principle of societal self-maintenance.

There is, however, no concern for Acquisition and Scope: no account is provided for the mechanisms that lead Ward to acquire the relevant ignorant beliefs given her position in gender and economic oppression, nor is there any account of why there should be oppressed agents, like working-class suffragists, who systematically avoid false consciousness. Further, the explanations for Content and Systematicity can only be partial, since EXPLANATION does not account for why there should be a difference between the content of Ward and Simkins' belief-sets. In considering the overlap of gender and economic oppression, both their false consciousesses are simply ‘functional for the patriarchal and bourgeois organisation of Victorian Britain.'

It is unclear then, how much EXPLANATION is an explanation for false consciousness so much as a restatement of the Explanatory Problem, like citing the virtus dormitiva of opium as an explanation for it causing sleep. That is, that false consciousness occurs because it sustains a given oppressive society is uninformative for the general project set out here: it is in fact just rehashing the functional ‘falsity’ in the definition of the concept as an explanation. This is because a functional explanation is, after all, meant as a hypothetical explanation that would guide particular social research into underlying mechanisms (Cohen 1982, Paprzycka 1998). At the level of abstraction we require, it is ill-suited for satisfying the explanatory desiderata.
2.2. Psychologism

‘Psychologism’ refers to various approaches that attempt to explain false consciousness by psychological mechanisms at the individual level, coming out of trends in social psychology to treat structural failures as fundamentally individual failures. These mechanisms that compromise one’s epistemic capacities range from adaptive preferences, where ignorant beliefs are indirectly acquired and maintained through an adjustment of the preferability of an initially preferable object (Elster 1983); the just-world hypothesis, where in virtue of one’s bias that the world is just, oppression is understood as just retribution (Lerner 1980); to aliefs, which are automatic proto-doaxastic responses to stimuli (Gendler 2011). Further, many note that the recidivism of false consciousness occurs not just in virtue of an agent’s group membership in such a structure but also in virtue of their self-identification as a group member (Meyerson 1991, Cudd 2006). Roughly, the maintenance of the relevant ignorant beliefs is motivated by a desire to avoid the truth about a given proposition, where this desire is generated by an agent’s higher-order desire to understand themselves as a certain kind of (efficacious and competent) moral agent and where this self-understanding is in turn derived from their membership in a social group.

A psychologistic explanation of WARD would be:

**EXPLANATION**: Ward has the ignorant beliefs that women are naturally deficient at political deliberation and that suffrage is unnecessary for addressing injustices, motivating her to campaign against women’s suffrage, which is formed and maintained owing to psychological

---

64 An earlier evaluation of this approach is Rosen's insightful criticism of adaptive preferences, wishful thinking, and inferential failures as explanations for false consciousness (M. Rosen 1996).
mechanisms that inhibit her ability to appropriately respond to suffragist reasons.

The variety of mechanisms identified by those who adopt this approach ensures that Acquisition would be satisfied. However, on its own, EXPLANATION cannot satisfy Content, Scope, and Systematicity. As Michael Rosen observes, a focus on content-insensitive failures of individual psychology is inadequate for why the content of the ignorant beliefs that are accepted and maintained through these mechanisms are precisely the kind that support structural oppression (M. Rosen 1996). That is, the question remains as to why ‘women are naturally deficient at political deliberation’ should be the content of Ward’s belief in psychological failure, instead of ‘women have equal capacity for political deliberation’ or say, ‘Uranus is the furthest planet from the Sun.’

Further, why these failures do or do not occur (and continually so) in a systematic way remains unaddressed. This is precisely the issue that is found in the worry about methodological individualism in the previous chapter (Ch. I, §2.3). Unsurprisingly, explicit attempts to explain false consciousness with psychologism do so by hybridising the approach with functionalism (Meyerson 1991), historicism (Elster 1982), or processualism (Shelby 2003).

These hybrid approaches are nevertheless still inadequate for critique. Psychologism-functionalism and -processualism hybrids (as we will see in the next section, §2.3) would still fail to account for Scope and at least Systematicity, since no explanation is provided for systematic resistance (like why such mechanisms are not also in operation for suffragists). And while a psychologism-historicism hybrid, wherein psychological mechanisms are contextualised within historically specific social practices (like the asymmetrical upbringing and education of Victorian women), may well account for Acquisition, Content, and Scope in particular cases of false

---

65 Neptune was discovered in 1846, 43 years before the Appeal.
consciousness, it remains too uninformative for our concern here for a general theory—failing Systematicity.

2.3. Processualism

Processualism is gaining attention recently due to ongoing work on epistemic oppression (as pioneered by Kristie Dotson), largely inspired by Gramsci’s understanding of hegemony which we saw in Chapter I (particularly §1.2; see also Williams 2002, Medina 2013, Mills 2017). The approach attempts to explain false consciousness by articulating the epistemological implications of the hegemonic relations between the oppressing group and the oppressed. That is, instead of understanding individual epistemic failure among the oppressed to be fundamentally explanatory as is the case in psychologism, processualism understands epistemic failure to be a relational failure. That is, processualism approaches false consciousness as a result of an active exclusion of the oppressed from participating in the larger epistemic community by the oppressing group at various orders (Dotson 2014). As such, the oppressed have deflated credibility with regard to themselves and epistemically defer to an credibility-inflated oppressing group with respect to the relevant ignorant beliefs.

66 For example, Charles Mills, in referring to his concept of ‘white ignorance,’ emphasises that “white” in “white ignorance” does not mean that it has to be confined to white people. Indeed, as the earlier Du Bois discussion emphasized, it will often be shared by nonwhites to a greater or lesser extent because of the power relations and patterns of ideological hegemony involved. (This is a familiar point from the Marxist and feminist traditions—working-class conservatives, “male-identified” women, endorsing right-wing and sexist ideologies against their interests.) Providing that the causal route is appropriate, [Black individuals] can manifest white ignorance also. (Mills 2007, 22)

67 This is often discussed in terms of hermeneutical and testimonial injustices, although there are also other kinds of epistemic injustices (cf. Hookway 2010).
According to the most recent and detailed processualist account by Jason Stanley (2015), it is because the oppressed are regularly placed in high-stakes situations (due to less material security) that they lose knowledge. That is, as a result of pragmatic encroachment (whether in virtue of interest-relativity or confidence-shaking), self-confidence in oneself as an epistemic agent decreases. And given the epistemic advantage afforded by their material advantage, members of the oppressing group overextend epistemic and practical authority over to claims pertaining to the oppressed. Consequently, the oppressed defer to the oppressing group’s claims, which are systematically detrimental because their content derive from legitimation myths developed by the oppressing groups. Such myths arise out of the latter’s desire to preserve the material inequality from which they benefit (like through propaganda).  

The processualist would explain *WARD* as follows:

*EXPLANATION*$_{pr}$: Ward has the ignorant beliefs that women are naturally deficient at political deliberation and that suffrage is unnecessary for addressing injustices, motivating her to campaign against women’s suffrage, *due to her deference to the epistemic authority of men, who propagate the myth of the political inaptitude of women to maintain their position in the gender hierarchy.*

In this way, *EXPLANATION*$_{pr}$ aims to satisfy *Acquisition* with epistemic deference, *Content* with legitimation myths, and *Systematicity* with how these relational practices presuppose the hegemonic organisation of social groups.

However, this account is unable to account for *Scope in all* instances of false consciousness. Specifically with Stanley’s account in mind, Amia Srinivasan

---

68 Denise Meyerson also alludes to legitimation myths (Meyerson 1991).
illustrates this problem with the case of a woman whom we would want to say does retain knowledge that she has been sexually harassed, despite being in a high-stakes situation (for example, where she is encouraged to take legal action against her boss for the act) and despite the “scepticism or gaslighting she encounters,” such that “she continues to act rationally when she pursues her protest” (Srinivasan 2016, 378). Being unable to say that the oppressed retains knowledge in such situations, Stanley is unable to account for the (systematic) epistemological advantage required for emancipatory consciousness in the oppressed, such as with the case of the suffragists. The account itself might thus be said to enact an overextension of epistemic and practical authority (stoking the fears of paternalism).

Some address of this might be found in the idea of practices where oppressed agents direct their “epistemic energy toward and in connection with other non-dominantly situated subjects,” attending to affinities and distinctions among them (Pohlhaus Jr. 2020, 245). These would accordingly form new relations and alliances independent of oppressors and illuminate “how epistemic systems function in the non-ideal world” (ibid., 247; cf. Medina 2013). However, the idea of such practices does not suffice for how some of the oppressed avoid false consciousness to begin with.

Yet, even if Scope could be addressed, under any processualist approach, Content is not always satisfied. A conspiratorial account of the oppressing group is not always plausible. It is often the case that the maintenance of legitimation myths are unintentional (perhaps as a matter of ideological consciousness in the oppressed) or even only tacit in practice (like implicit bias). Especially for myths maintained by the same group across long time scales, such deception would require an accurate grasp of the actual state of affairs, which is not only unnecessary for propagating these myths but also largely unempirical (Shelby 2003). Admittedly, it may be the case that legitimation myths are sometimes initially formed by a select portion of the
oppressing group intentionally seeking to preserve their material inequality. But the crucial point is that it would be implausible to see this in all occasions of false consciousness. It seems more plausible and common that many legitimisation myths that supposedly reinforce structural oppression are not perniciously created. Examples of this include the notions of meritocracy and racial colour-blindness. While these notions may have since been used to legitimise certain structures of oppression, this does not imply that their formation were similarly motivated: meritocracy has its earliest, and most influential, articulations in Plato’s *Republic* and Confucius’ *Analects*; and racial colour-blindness in the US civil rights movement.\(^{69}\)

Further, the necessary material basis of this approach cannot accommodate the persistence of cases of false consciousness when there is *material equality* between groups. (Recall how *EXPLICATION* \(_M\) falters in explaining the false consciousness of Ward as a millionaire-equivalent.) We also see this in cases of ‘Pinkerton Syndrome,’ where Singaporean Chinese would tend to provide preferential treatment for white people (Chew, Young, and Tan 2019). And more generally, processualism alone is still insufficient to account for the *positive internalisation* of the relevant content by some significant portion of the oppressed, especially among those that further produce highly abstract and nuanced elaborations of the relevant ignorant beliefs. Many women anti-suffragists were at the forefront of the Victorian literati. It is not clear why they should not just be defeatists, sceptics, or ironists about such beliefs (cf. Cunningham 1983, Tessman 2009, Wilderson III 2020). Both these explanatory gaps apply to *EXPLANATION*\(_{PR}\). *Systematicity* and *Acquisition* thus do not fully obtain under processualism either.

\(^{69}\) This criticism of (Confucian) meritocracy, as initially benign but subsequently becoming problematic, has been made even as early as the *Han Feizi* (c. 3rd century BCE). For a discussion of this, see Wilson forthcoming.
3. Social Constructionism

The above approaches to false consciousness are, however, not necessarily in direct tension with each other. Instead of displacing them, the social constructionist approach may perhaps be seen as grounding a hybridisation of all three. In this section, I develop such an approach based on Althusser’s analysis of ideology that meets all the explanatory desiderata. Althusser’s analysis of ideology differs from Gramsci’s (as favoured by processualists) insofar as the focus is on an agent’s consciousness as structured rather than structural hegemony. Specifically, Althusser understands ideology (or schemas, more generally) as a three-part system that involves the “interpellation of individuals as subjects,” “mutual recognition between subjects and Subject and among subjects themselves, as well as the recognition of the subject by himself,” and “the absolute guarantee that everything really is so” for the subject (Althusser 1995, 197). That is, ideology hails an individual as such-and-such a participant in its structure, within which they are affirmed and affirm themselves as such a participant, and this state of affairs is naturalised. I will conclude this chapter by considering how functionalism, psychologism, and processualism are subsumed and supplemented by social constructionism. For consistency, Althusser’s own notion of ‘ideology’ is simply referred to as ‘schema’ here, as he uses the term in a neutral way that also applies to unproblematic practices.\(^\text{70}\) For example, he refers to revolutionary organisations as operating on “proletarian ideology” (Althusser 1995, 181).

\(^{70}\) As I have observed, so did Lenin and Lukács at times (see Note 42).
3.1. Interpellation

The ‘interpellation of individuals as subjects’ is a constitutive process by which an individual becomes a *kind* of subject under a representational schema. In other words, social properties are constructed upon a base property (aspects of the individual human being) to create a subject (an interpellated individual). Whether under constitutional (Searle 2010), response-dependent (Pettit 1991), or conferralist frameworks (Ásta 2018), the relevant type of social construction here minimally involves a joint acceptance condition:

Members of a group individually accept fact(s) $F$ about object $O$ and there is mutual knowledge that other members accept $F$.

That is, kinds of objects are socially constructed when the kind of object comes to possess a certain social property, where such a property arises out of a joint acceptance (that $O$ is $F$) by members of a group of people whether at the level of formal institutions or informal communities.

This type of social kinds, notably, is distinct from the type of social kinds that can exist even if no one has any attitude towards them (see Khalidi 2015). Ideologies and false consciousnesses themselves are cases of this latter type of social kinds: they are dependent on certain attitudes of people for their existences, but not on people having attitudes towards them qua ideologies or false consciousnesses (such as in critiques of them). Ideologies and false consciousnesses are, in fact, often dependent on people not having attitudes towards them for their existences. To return to the example of Victorian gender ideology: while the existences of such an ideology and corresponding false consciousnesses depend on attitudes

---

71 I like to thank my examiners for pushing me to clarify this point.
such as the belief that ‘women who do not want to marry are unfeminine,’ they do not depend on anyone having attitudes towards them, such as in Harriet Taylor and John Stuart Mill's thematisations and critiques of Victorian gender ideology and corresponding false consciousnesses. On the other hand, the Content of false consciousness would involve the former type of social kinds that depend on people having attitudes towards them: a constituent belief such as ‘women (O) who do not want to marry are unfeminine (F)’ itself depends on having attitudes towards women qua social kind.

Crucially for interpellation, we see that groups of people—that is, members within them—are also constructed as subjects, insofar as they have properties jointly accepted of them within a network of social practices. For example, an intersex individual with ovotesticular disorder of sex development—i.e. born with both ovarian and testicular tissues—would be assigned exclusively either female or male by most medical institutions. These interpellating social categories and corollary properties come with, if not just are, constraints and enablements within social contexts that would have significant effects on the attitudes and behaviours of individuals. Judith Butler points out, for example, that the categories embedded in sexuality norms “produce and constrain at once” and “play [themselves] out in the securing for every body a sex, a sexed position within language, a sexed position which is in some sense presumed by any body who comes to speak as a subject, an ‘I’” (Butler 1993, 95, emphasis mine). Put another way, individuals are interpellated as subjects insofar as they are always already constructed under some schema that relate them in social practices—as schematised objects—to other resources, other schematised agents, and to themselves. Such schematisation is a precondition of being considered a participant in a given social practice (consider the plight of non-binary individuals in countries that have highly binarised customs and laws regarding gender).
Importantly, apart from institutional regulation like incarceration or even the threat of negative communal sanctions (which both apply to the already-constituted subject), Althusser notes a “[strange] obligation to respond to every hailing [by interpellation]”: an underlying norm for an individual to acknowledge themselves as being addressed by the social schemas they find themselves in (Althusser 1995, 191). This is irrespective of repressive state apparatuses like police or military intervention, or what Searle calls “background power,” i.e. the “set of presuppositions, attitudes, dispositions, capacities, and practices of any community that set normative constraints on the members of that community in such a way that violations of those constraints are subject to the negative imposition of sanctions by any member of the community” (Searle 2010, 160). This norm of acknowledging ‘being a subject who is addressed as such-and-such’ asserts a pressure on individuals to participate in joint acceptances, whose source is thus neither institutional nor communal but simply an individual's receptivity to sociality—even if they do not subsequently recognise themselves in the schema or only do so in defeatism, scepticism, or irony. Non-binary individuals, for example, still acknowledge binary gender schemas as at least addressing them—albeit problematically. Therefore, the normative force of joint acceptances is not found merely coming from the confines of the oppressed group (cf. Thompson 2017). Both oppressing and oppressed groups partake in the same, wider social structure and ideology, within which they partake in the joint acceptance of themselves as instantiating certain social facts particular to the groups they belong to.72

So while schemas may be found to be functionally problematic in critique, the constructive acts that give rise to certain attitudes and self-understandings that support structural oppression need not be—that is,

72 Cf. the Frankfurt School’s description of homogeneity in a society of technological rationality under advanced capitalism (see Ch. I, §1.4).
both the initial positing of an object as possessing certain facts and the subsequent joint acceptances. *Pace* conspiratorial accounts, constructive acts may also just be defective or unwitting, with the effect of the oppressive relation resulting unintentionally.

### 3.2. Recognition

As a result of interpellation, a subject is able to have standing under a schema or ideology—be *recognised*—with respect to the social world, other subjects, and, most importantly for the concept of false consciousness, oneself. That is, a subject acquires social standing within a network of practices, within which they are related to—reinforcing joint acceptance—as such-and-such a subject, whether practically (in sexual harassment) or representationally (in TV scripts). And this forms the basis of their self-recognition and hence self-understanding. This is what Ian Hacking (1995) observes as ‘the looping effect,’ feedback loops between social *classifications* and how the *classified* understand themselves and behave (see Figure 1):

![Figure 1: Looping Illustration 1](image)

That this feedback loop occurs might be said to be simply a feature of our cognition in general. Seen from the predictive processing paradigm (I turn to an enactivist articulation of this in Chapter III), cognition aims to track the world by modelling sensory input with generative models, which “capture the statistical structure of some set of observed inputs by inferring a causal matrix able to give rise to that very structure” (Clark 2015, 21). Our cognition not only models objects and their relationship to each other but also agents—including, crucially, our selves. Unlike cognitions of objects and other
agents, which involve causes that pertain to sensory data coming from outside the body, our self-cognition involves modelling endogenous causes (Seth, Suzuki, and Critchley 2012). Such self-cognition begins in infancy, when we start to apply the ‘agent-models’ generated prior, and the agent-models ascribed to us, to understand the endogenous sensory input of the body as caused by such-and-such an agent—on the basis of which we then act accordingly, confirming the ascriptions (Hohwy and Michael 2017). Such application of agent-models continues into adulthood. This is thus how interpellation is internalised and one recognises oneself as such-and-such a subject under a given schema: in this feedback loop between the agent and the ascribed classification, the corollary social properties and schemas are more than jointly accepted of a classification—they become one’s own properties and self-understanding.73

Pertinent to false consciousness is how such a complex of feedback loops leads to certain joint acceptances taking precedence over others. We need to attend to how, in cases of structural oppression, certain looping mechanisms systematically function to promote specific ignorant beliefs in spite of counter-evidence.74 That is, we are concerned with how the concept can guide attention to what these epistemically unresponsive mechanisms are, as well as how they work to naturalise ideological experience of the social world, such that beliefs pertaining to these social objectifications (and

---

73 This is the other side of Lukács’ observation of the phenomenon of reification, wherein one’s “own activity […] becomes something objective and independent of [them], something that controls [them] by virtue of an autonomy alien to man” (Lukács 1923, 87). While Michael Thompson insightfully notes that reification [Verdinglichung] draws on Kantian distinctions between Ding, Gegenstand, and Objekt as levels of being cognised, this alone nevertheless fails to account for how ideology also concerns the Subjekt, the cognising empirical unity of apperception [empirische Einheit der Apperzeption] i.e. the interpellated subject (Thompson 2017; cf. Althusser 1965, 230n7).

74 While Hacking sees the looping effect as destabilising human kinds and thereby social objects, many other social constructionists like Haslanger and Ron Mallon have noted that it can also be stabilising (Haslanger 2011, Mallon 2016). But this is merely a verbal disagreement: Hacking largely uses the looping effect to differentiate human kinds from natural kinds, which is supposed to be more stable, whereas Haslanger and Mallon are concerned with such kinds having enough stability to “support our epistemic projects over periods of time that are interesting to us” (Mallon 2016, 181).
not others) are ‘guaranteed’—that is, taken for granted as (putative) knowledge in the agent’s lived experience.

### 3.3. Guarantee

Social objects are guaranteed by the stability they acquire within social structures over time. There need not be any general reason for why these objects naturalise over others, or why these objects and not others are constructed to begin with—the circumstances and content of the constructions, just as the reasons for their propagation, are of course socio-historically specific. At the same time, this feedback loop always occurs against a formal background system of other loops that occur in the broader social world and among other subjects. Hacking (2017) observes that looping also involves more than the arcs between classifications and classified at the individual level, including also, at the structural level, knowledge involving such classifications, experts who classify, and institutions that uphold classifications:

![Figure 2: Looping Illustration 2](image)

---

75 For criticisms of a single-loop caricature, see, e.g., Celikates 2017.
Consider, for example, how the link between disabled people and disability classifications are understood to stabilise and be stabilised by how disability is understood in medical and non-medical texts, medical professionals and disability activists, as well as hospitals and legislative councils (Lim 2018).

Our concern here is explicitly for looping mechanisms that are epistemically unresponsive (unlike, say, epistemic virtues). Remaining pluralistic about the mechanisms, then, the following is a suggested, inexhaustive list of such mechanisms at both individual and structural levels, noting that they need not all occur and that not all of them are problematic in themselves:

**Individual:** Nash equilibria, stereotype threat/boost, implicit bias, wishful thinking, confidence-shaking, speech act accommodation, epistemic laziness, testimonial injustice, meaning vertigo, etc.

**Structural:** market equilibration, culture industry, consumer production, hermeneutical injustice, epistemic bubbles, echo chambers, recommender systems etc.

An example of a looping mechanism functioning at the individual level to perpetuate structural oppression may be found in the stereotype boost for East Asians in North America. Joint acceptance of East Asians as possessing properties of being highly competent but cold and submissive constructs the classification of the ‘Asian-American’ as a model minority and interpellates relevant individuals (Berdahl and Min 2012). This undergirds the recognition of the Asian-American-as-model-minority that manifests in practices like hiring decisions or deferring the calculation of splitting of the bill to them. With these practices, many such individuals also recognise

---


themselves as possessing these properties due to the stereotype boost, contributing to their relative absence in leadership positions (Sy et al. 2010).

Consider also how such individual mechanisms could also work where structures of oppression overlap and interact, allowing us to attend to variations in false consciousnesses and the corollary epistemic and moral harms suffered by affected agents. In YouTuber Natalie Wynn’s retrospective “Shame” video, she illustrates this in her reflection on her prior ignorant belief that she, as a trans woman, was not also a lesbian. She analyses how this evidence-resistant belief, which motivated her to play the role of a “Stepford wife” as an influential representative of the trans community (thus reinforcing both gender role and sexual conformity), was partly grounded in avoiding the stereotype threat of “trans women [being] men who transition to creep on women in bathrooms” (Wynn 2020). Such a stereotype threat does not result from merely adding “compulsory heterosexuality” and gender binarism together (Rich 1980). Trans men, for example, are not stereotyped as transitioning to creep on men in bathrooms.

An example of a looping mechanism functioning at the structural level may be found in Haslanger’s example of market equilibration in the global food industry:

In a less-globalized world than ours, food crops were grown to support the local cuisine and the local tastes and culinary techniques evolved in ways that took advantage of the crops. In more complex and broadly social changes we can watch consumer taste develop so that certain products become ‘must haves’ in a particular milieu. Trends in cuisine can become trends in production which, in turn, affect trends in labor, and this affects schemas of class and taste, and so on. […]

The reliance on, say, wheat in a particular cuisine may seem inevitable, natural, ‘given.’ Wheat is what is available; wheat just is what we eat. But the wheat is available because

78 I attend to Wynn’s retrospective more closely in Ch. III (cf. Tagonist 2019).
of the impact of schemas on resources that establish farming practices, food distribution, and such. Given the stability of such structures, culinary taste conforms. In this context quinoa, or soy, or spelt tastes bad and has a funny texture too; so who would want to plant it? (Haslanger 2011, 197–198)

This provides us with a looping mechanism that stabilises, say, Orientalist attitudes towards tofu-based dishes as ‘a Far East exoticism’ as it overlaps with global capitalism.

Looping effects thus allow us to say that the ignorant beliefs in false consciousness both supports (functional ‘falsity’) and is supported by structural oppression (genetic ‘falsity’), though we avoid vacuous circularity since the relevant ideological schematisations of the oppressed are constructed and naturalised for an agent over time. This also means that the formation of a given false consciousness in the agent neither is causally prior to, nor emerges simultaneously with, the relevant oppressive relation. Rather, the relevant oppressive relation exists, at least initially, independently of the interpellation and hence also prior to the development of the particular false consciousness—guaranteeing it. That is, looping mechanisms normalise, or render ‘natural,’ existing oppressive relations. This is not to say that such oppressive relations do not also depend on false consciousness later on, only that it is impossible to exhaust the reasons for why a given instance of false consciousness is problematic without also attending to why the relations of its corresponding social structure(s) are problematic.  

This thus allows us to elaborate Explanation, wherein material force grounds intellectual force: “the ruling ideas are nothing more than the ideal expression of the dominant material relations, the dominant material relations grasped as ideas; hence of the relations which make the one class the ruling one, therefore, the ideas of its dominance” (MECW 5: 59). That is,

---

79 This leaves open the question of the appropriate norms of evaluation, i.e. the Normative Problem, which will be the focus of Ch. IV.
the asymmetry of material interests exists *prior* to the formation of ideology and corresponding false consciousnesses. We see a clear example of this priority in the history of slavery in the US, where such a hierarchical consciousness arose out of an initial oppression. European colonialists constructed a ‘lesser race’: it was, initially, mainly Black and white indentured servants that were relied upon for the growing plantations in the New World. This was before a series of revolts by both Black and white servants against plantation elites made slaves imported from Africa better favoured by plantation elites for labour, due to their abundance and unlikelihood to form alliances with poor whites (Alexander 2010). Both white and Black populations later developed an ideology, and corresponding economic and racial false consciousnesses, of a hierarchy of white over Black to be the ‘natural’ order.80

### 3.4. Dynamics

The dependence of false consciousnesses on an oppressive relation means that they are *domain-based*: indexed to the range of resources and agents schematised by the relevant social structure. A given social structure would delineate a stabilised complex of loops, but a given domain may at the same time be schematised, partially or wholly, in other structures (oppressive or otherwise). Importantly, this mean that a given instance of false consciousness is limited both in its content and mechanisms. An instance pertaining to suffrage, for example, would not involve beliefs about Pluto. Being domain-based means that false consciousness might not develop, or

80 It would also address Robin Celikates’ (2017) worries that a Haslangerian approach to ideology cannot explain why certain schemas loop better than others and lack top-down, non-contingent ways that fundamental structural mechanisms might develop a logic of their own.
could erode, for at least two reasons (independent of critique, which I will
discuss in the next chapter): radical marginalisation and epistemic friction.81

*Radical marginalisation* comes from Marx’s own description of the proletariat
as the locus of the possibility of German emancipation: the violent revolt of
the *fully* oppressed, who have nothing to lose but their ‘radical chains,’
against the existing structure as a whole.82 It is precisely the insignificance of
the structure for such agents who are completely marginalised that allows
for the possibility of emancipatory consciousness. This is because the
stabilising mechanisms have no force on them since they are not *subjects*
under the domain of the ideological schemas (for example, there would be
no stakes at all for confidence-shaking): they would be able to form beliefs
independent of ideology.

*Epistemic friction* comes from the fact that different social structures may
schematise over the same resources and agents. That is, an agent likely has
multiple group memberships within a given society and, as such, may be
under multiple schemas in cases of overlapping and interacting systems of
oppression (whether purely oppressor, oppressed, or a mix). As Althusser
notes:

> Ideologies never stop interpellating subjects as subjects, never
stop ‘recruiting’ individuals who are always-already subjects.
The play of ideologies is superposed, criss-crossed,
contradicts itself on the same subject: the same individual

From this multiplicity, ideologies can either be *compounded* or *neutralised*
for an agent. When they are compounded, schemas combine such that a

81 Yet another means to erode the self-understanding implicit in false consciousness might be through ego-dissolution brought about by psychedelics (Letheby and Gerrans 2017). But alas this is not systematic.
82 Also Frantz Fanon’s turning to the *Lumpenproletariat* for decolonial hopes (Fanon 1961).
given false consciousness can gain not only content but further stabilising mechanisms that reinforce evidence-resistance. At the same time, these overlaps can also condition the prevention or erosion of false consciousness in cases where ideologies are neutralised. For such a multiplicity of self-understandings may generate what José Medina calls ‘epistemic friction,’ a discordance resulting from incompatible schemas and stabilising mechanisms running up against each other, which may lead to ‘meta-lucidity.’ An agent would thus have “insights in to the functioning of perspectives that makes it possible to redraw [one’s] cognitive maps, to redescribe [one’s] experiences, and to reconceptualize [one’s] ways of relating to others” (Medina 2013, 47; cf. Cunningham 1983). Such reconceptualisation would not take place in abstracto, but rather would be responsive to the various extant self-understandings. In Butler’s discussion of Althusser, she similarly observes that such reflective distance from one’s subjecthood, wherein the “discursive possibilities for existence exceed the reprimand voiced by the law [of interpellation],” conditions the necessary space for developing emancipatory consciousness, involving “a willingness not to be—a critical desubjectivation—in order to expose the law as less powerful than it seems” (Butler 1997, 129–130).

In both cases of a lack of development or erosion, it is by participating in alternate social constructions of objects—and oneself as a subject—in non-ideological or resistant communities that one might come under social schemas stabilised by epistemically responsive mechanisms, and thus non-ideological or emancipatory knowledge (see Goodin and Spiekermann 2015). This supports the processualists’ insight of resistant practices, while also accounting for how agents might avoid false consciousness in the first place.

However, such a lack of development or erosions—which may take a while—render emancipatory consciousness only possible and cannot secure
actualisation. For resistant communities to be successful, alternate schemas must overcome the interpellation of the old ideologies.\textsuperscript{83} This must be done to the extent of diminishing the joint acceptances that issue from the norm of acknowledging that one is interpellated by a particular ideology. For example, no one would acknowledge being hailed as a Roman citizen in Victorian Britain. Such diminishment requires sustained competition and control over the schematised resources and looping mechanisms, and may be unsustainable in the long run without disrupting looping at the structural level, as well as the corollary reconfiguration of the internal relations of the relevant social structures (cf. Lugones 2003). The massive changes to British labour markets during the First World War and the consequent shifts in the public consciousness of women’s role were crucial for securing suffragism’s success. It is thus perhaps unsurprising that false consciousness is just as commonly, if not more often, reinforced by epistemic friction (Nyhan and Reifler 2010). Or, erosion sometimes even results in a new type of false consciousness developing in moments of meta-lucidity as a flight from ‘meaning vertigo’ back to the interpellation of the old ideology (Melo Lopes 2019).\textsuperscript{84} Similarly, it remains a controversial empirical question whether a revolt by the radically marginalised has ever been or will ever be more than a possibility.

4. Summary

I will now summarise the foregoing and how social constructionism accommodates the previous three approaches. The social constructionist approach to false consciousness understands the formation and

\textsuperscript{83} This leaves open how representative such new schemas and communities would be, if other oppressed agents do not share the same overlaps of structural oppression (as we will see in the Normative Problem in the next chapter), cf. Celikates 2017.
\textsuperscript{84} Filipa Melo Lopes finds this in raunch culture, a purportedly ‘feminist’ movement in which women were encouraged to ‘own’ their sexual objectification as a means to ‘empower’ themselves—largely in heterosexual contexts (Melo Lopes 2019).
maintenance of ignorant beliefs to be grounded in agents taking a certain socially objectified understanding of themselves as natural self-understanding, within variable complexes of epistemically unresponsive looping mechanisms that are embedded in oppressive social structures over time. A given complex could overlap with another to form either a non-additive set of beliefs that includes both ideologies, or erode either sets of beliefs—this depends on the compatibility of looping mechanisms salient for the agent in question.

The social constructionist would explain the particular case of \textit{WARD} as follows:

\textit{EXPLANATION}_{sc}: Ward has the ignorant beliefs that women are naturally deficient at political deliberation and that suffrage is unnecessary for addressing injustices, motivating her to campaign against women’s suffrage, because the contents of those beliefs are being sustained by a complex of looping mechanisms at the individual level and structural level. These mechanisms are patriarchal and bourgeois.

\textit{At the individual level, this involves a loop between Ward’s self-understanding and patriarchal and bourgeois classifications: the classification of ‘domestic creature’ under patriarchal ideology and the classification of ‘subject who is capable of activities that secure economic independence’ under bourgeois ideology. These then merge as ‘women are capable of national domestic interests’ and is sustained by mechanisms like testimonial injustice and epistemic arrogance.}

\textit{At the structural level, these would further involve an organisation of loops involving: the contemporary gender exclusion in political}
institutions but gender inclusion in literary institutions; the expert opinions of members of the Houses of Parliament, literary figures, and critics; and the limited public knowledge of the political capacity of women. These are sustained by mechanisms like hermeneutical injustice and market equilibrium.

Acquisition and Content are thus met in EXPLANATIONsc and shown to be interacting in the complex of looping mechanisms at both individual and structural levels. Scope is met in attending to how suffragists might lack similar testimonial injustice or might not have similar assurances with the ‘existing constitutional machinery,’ thereby coming to know of women’s capacity to perform activities that would but do not secure economic independence, sustained by the growing resistant community of the suffrage movement. And lastly, Systematicity is met, since Acquisition, Content, and Scope are interconnected by the complexes of looping mechanisms in both gender and economic structures whose domains significantly overlap. And, while the general form of this explanation may be applied to cases like Simkins’, the particular explanation (EXPLANATIONsc) only extends to anti-suffragists in a similar social position to Ward. The looping mechanisms pertaining to the latter involve a different overlap of gender and economic oppression to the former: the resources of upper- and middle-class British women then, like charitable funds and chambermaids, are not shared by working-class women.

Like functionalism then, social constructionism is pluralistic in regarding which underlying mechanisms are at play in general: it could be one or a combination of them in a particular case of false consciousness. Instead of a specific mental state that agents are in, false consciousness is to be understood as a conceptual tool that regulates systematic social research.
(without predetermining the particular mechanisms for every instance). What is further provided by the social constructionist framework is a means to identify and unify these mechanisms under a given ideology or, importantly, multiple ideologies (Acquisition), as well as account for their absence (Scope).

Social constructionism also accommodates the various possible psychological mechanisms that might factor into particular cases of false consciousness, without committing itself to any one (or set) of them as essential to the concept. Moreover, it provides an account for why ideal psychological conditions are so difficult to achieve and sustain, in supplementing the psychologistic approach with Content, Scope, and Systematicity.

Finally, like processualism, social constructionism is attentive to epistemic oppression and the epistemic failure of false consciousness as relational failure. It also provides an account for the possibility of the success and failure of resistant epistemic practices. At the same time, social constructionism is also able to address cases that are not conspiratorial in origin or that do not depend on material inequality (Scope), as well as the positive internalisation of ideological self-understanding (Acquisition).

I have thus argued that social constructionism is the most comprehensive answer to the Explanatory Problem of false consciousness. It provides a general theory that meets all four explanatory desiderata of Acquisition, Content, Scope, and Systematicity needed for the concept in social research and emancipatory praxis. And, instead of supplanting functionalism, psychologism, and processualism, it subsumes their explanatory strengths and supplements their weaknesses. This thus places us on firm footing to

---

85 Here, it might be useful to note the analogy of these tasks with the regulative and constitutive operations of reason for Kant, as laid out in his Critique of Judgment (5: 401–404).
address the other problems of false consciousness, the Normative and Agential Problems, which I will go on to address in the subsequent chapters of this thesis.
III. What’s So Bad about False Consciousness?
Or, Towards an Immanent Approach to Well-Being

CHAPTER ABSTRACT

The concept of false consciousness requires a normative standard on the basis of which it is said to be problematic. This standard, however, faces a twofold challenge: it needs to not be alien or alienating to the agent being criticised, and it needs to avoid being itself ideological. In this chapter, I turn to conceptions of well-being for such a standard, due to the various appeals made to them by social philosophers and how they may condition other kinds of standards (e.g. moral or political). I set out two criteria that any normative standard to be used in the concept of false consciousness must satisfy, in order to meet the challenge: Continuity and Endorsement. I find, however, that the kinds of prudential theories that have been used for the Normative Problem —namely, realist and anti-realist approaches—cannot satisfy both Continuity and Endorsement. In particular, I will be considering these conceptions in relation to the case of the working-class anti-suffragist Maud Ellen Simkins. Given the failures of realist and anti-realist approaches, I turn instead to outline a possible immanent approach to well-being that would meet the demands of an adequate normative standard for the concept of false consciousness. Such an approach draws on enactivism and recent Critical Theory for a normative standard that is grounded in the need for us as sense-making organisms to adapt to systemic problems in a way that ensures the well-functioning of future adaptations.
In the previous chapter, I argued that the concept of false consciousness is best understood as a regulative concept under a social constructionist approach. The application of the concept draws attention to the heterogenous mechanisms that work at both individual and structural levels, in order to conduce agents to hold ignorant beliefs that motivate them to act in self-oppressive ways. This approach allows us to better appreciate the various mechanisms by which particular systems of oppression work to systematically co-opt ‘bottom-up’ support from the oppressed—and, as such, also allow us to strategise ways to disrupt these mechanisms and form resistant loops within such systems. With regard to the latter, it is crucial to understand how the use of the concept of false consciousness in critique might effectively enjoin some of the relevant agents of social change to resistance: traditionally and often thought to be the oppressed themselves, who may be under varying false consciousnesses to varying extents.

A successful critique of false consciousness would involve such agents coming to understand false consciousness as problematic for them qua oppressed individuals. Yet, under false consciousness, agents often tend to see their (oppressed) situation as having positive valence, sometimes with high degrees of sophistication—to the extent that the agent may regard the critic as themselves under false consciousness. It is within this context that the question of why an agent under false consciousness should seek to overcome it becomes problematic. In this chapter, I will take up this question within the constraints of the Normative Problem of false consciousness. The problem, put generally, is as follows:

To which normative standards should a critique of false consciousness appeal in accounting for why false consciousness is bad, in order to not alienate the agent criticised and also ensure that the critic does not reproduce any ideology?
To answer this, I turn to a prudence-based account of the Normative Problem. The underlying idea in such an account is that false consciousness is, at least partially, bad because it compromises the well-being (‘the human good’ or ‘the good life’) of the agent under it. That is, it is bad for the agent qua welfare subject, irrespective of the relevant ignorance about their well-being. It seems, after all, rather intuitive to characterise false consciousness as problematic by understanding its ‘acting against one’s own interests’ to involve an agent ignorantly compromising their own well-being—being ignorant of what it is like to ‘be the kinds of being that we are well.’ In this way, the critic may rely on certain subject-relative standards to explain to agents under false consciousness how they are indeed worse-off.

To anticipate, I do not ultimately turn to a substantive theory of well-being for the relevant prudential norm(s) here. Substantive theories set out to specify in advance what is ultimately good or bad for any welfare subject and they tend to be of either realist or anti-realist kinds (Fletcher 2021). For a realist theory, what is ultimately good or bad for an agent would be an attitude-independent fact (such as autonomy). For an anti-realist theory, what is ultimately good or bad for an agent would be a fact about the attitudes of some agent (such as a preference). Versions of these two kinds of theories have had particular prominence in extant attempts to provide prudential

---

86 Philosophers have often attempted to ground their answers to the normative problem of ideology critique in political, social, epistemic, and moral norms thought to be universal (or at least sufficiently shared). For example, norms of recognition (Honneth 2007), justification (Forst 2017), and first-order “moral truths” such as “slavery or rape is wrong” (Haslanger 2021, 7). I will not be revisiting and assessing these attempts here, which have had rich and extensive discussions in various literatures (see, e.g., Jaeggi 2019, Saar 2020).

87 This distinction is usually drawn in terms of ‘externalism’ and ‘internalism,’ or ‘subject-transcendence’ and ‘subject-dependence’ (Hall and Tiberius 2015). I follow Fletcher’s distinction here since it is the least disputed, and to my mind the most convincing, taxonomy of extant theories of well-being.

88 Autonomy itself may be understood substantively (e.g. Nussbaum 2011) or non-substantively (e.g. Friedman 2003). I briefly discuss this distinction and consider non-substantive (sometimes ‘content-neutral’ or ‘procedural’) conceptions of autonomy in greater detail in §3.2.
standards for the concept of false consciousness—as well as for its spin-offs, deformed desires and adaptive preferences (for examples, Geuss 1981, Superson 2005, Terlazzo 2016, Enoch 2020). But I will argue in this chapter that neither such realist nor anti-realist theories of well-being—focusing on the commonly used autonomy and preference-satisfaction accounts, respectively—can adequately answer the Normative Problem, suggesting that there might be no role for substantive theories of well-being in a critique of false consciousness.

In their place, I outline a formal, non-substantive approach to the relevant prudential standards in critique—which are not determinable in advance of empirically-based research into a particular welfare subject’s constitution in social reality. Unlike contextualist approaches to the concept of well-being, which hold that “the semantic content of sentences in which ‘well-being’ and its cognates occur depends, at least in part, on the context in which it is uttered” (Alexandrova 2017, 5; cf. Campbell 2015), the contextual element in my approach pertains to the conception of well-being. More specifically, the account I offer here only articulates the formal conditions under which a critic may judge something to be detrimental to an agent’s well-being, on the basis of the form(s) of life that both the critic and the criticised participates in. I thus call this an immanent approach to well-being, since the articulation and criticism of well-being remains within any welfare subject’s conception of it.

My immanent approach draws on a largely neglected kinship between enactivism in the biological-cognitive sciences (inspired by Sylvia Wynter) and recent Critical Theory (specifically, Rahel Jaeggi’s). Such an approach is

89 This is distinct from an indirectly substantive account of well-being, where substantivity is built into some procedural condition. For example, in the case of an indirectly substantive account of autonomy, it requires that “some of a person’s live alternatives must be valuable ones” (Terlazzo 2016, 224).

90 For an argument against the contextualism about the concept of well-being, see Fletcher 2019, 2021.
enactive, inasmuch as it understands that we are self-enacting organisms that engage in participatory sense-making and, accordingly, the relevant prudential standards are conceptually inextricable from the form(s) of life in which we find ourselves.\textsuperscript{91} The approach is also critical, inasmuch as it does not rely on normative standards not already held by the agent, but rather develops a critique of them from these very standards themselves.\textsuperscript{92} The approach would be, in virtue of these two aspects, neither alienating nor ideological.

In what follows, I begin with a brief consideration of how many philosophers implicitly understand the badness of false consciousness as (at least partly) fundamentally a prudential matter (§1). I then consider the Normative Problem in the context of a public exchange in 1889 between the then anti-suffragist Louise Creighton and the suffragist Margaret Dilke—specifically in terms of how Dilke’s critique failed to enjoin Creighton to the suffragist cause because it was concerned with ideology rather than false consciousness. This reveals that any answer to the Normative Problem will involve two necessary requirements particular to a critique of false consciousness: first, an account of how agents under false consciousness might have normative reasons to resist, ones which are continuous with their extant concepts; and, second, an account of how such reasons would garner their endorsement (§2).\textsuperscript{93} Taking up the false consciousness of working-class anti-suffragist Maud Ellen Simkins as a case study here, I argue that extant realist and anti-realist approaches to the Normative Problem are conceptually inadequate for meeting either or both requirements, thus failing to solve the Problem

\textsuperscript{91} Socio-political extensions of autopoietic theory and enactivism—and 4E cognition more broadly—have otherwise been scarce (see Wynter 2001, Engelstad 2016, Di Paolo, Cuffari, and De Jaegher 2018, Brancazio 2019).

\textsuperscript{92} There is another sense in which I understand my approach to be critical, in that it has an “emancipatory interest” (Honneth 2017). That is, I take for granted that emancipatory praxis is a goal for our normative theories.

\textsuperscript{93} This section might be understood as offering an emancipatory analysis of the concept of well-being, alongside rational care, locative, positional, and suitability analyses (Campbell 2015).
I then turn to outline my immanent approach to well-being and how it might be able to lay the groundwork for any adequate answer to the Normative Problem in critical social research (§4).

1. The Normative Problem and Well-Being

Although not always explicit in the Critical Theory tradition, conceptions of well-being are thought to lie at the heart of it, which has as its “primary task [...] the diagnosis of processes of social development that must be understood as preventing the members of society from living a ‘good life’” (Honneth 2007, 4). Such a diagnosis is meant to allow agents to see through ideological distortions to reasons that would be “motivating so that [agents] are prepared to act in order to reconstruct [their] current social reality” (Haslanger 2020, 36).

In advancing an explicitly prudential approach to the Normative Problem here, I do not mean to thereby entirely supplant non-prudential approaches in the literature: such as epistemic or moral ones. Rather, I understand conceptions of well-being to be at least complementary to these. And, as I will now go on to suggest, many of these non-prudential attempts may be characterised as relying on conceptions of well-being—whether explicitly or implicitly. That is, with respect to the Normative Problem, well-being underlies the other normative goods or at least necessarily accompanies them.

Consider how answers to the Normative Problem with an epistemic focus would have to make background appeals to conceptions of well-being (cf. Elster 1980, Williams 2002). After all, that an agent’s attitudes are ignorant does not its own mean that they are genetically or functionally problematic. As I explained in the previous chapter (Ch. II), the mere fact that a belief is
held as a result of epistemically unresponsive mechanisms (and thus ignorant on the basis of a lack of safety or sensitivity) does not on its own explain why it is a matter of false consciousness. Learning that one’s beliefs that ‘Goldbach’s conjecture is true’ or ‘possible worlds exist’ are acquired and maintained only out of the wishful thinking of appearing like a traditional analytic philosopher is not reason in itself for them be understood as beliefs constituting false consciousness. As we saw with psychologism, the sort of wishful thinking—and belief-forming mechanisms more broadly—relevant for false consciousness are those relating to the agent’s disadvantaged circumstances. As Tommie Shelby elaborates it, our beliefs are sometimes formed for the sake of our psychological well-being: “we sometimes believe things because to do so would, say, bolster our self-esteem, give us consolation, lessen anxiety, reduce cognitive dissonance, increase our self-confidence, provide cathartic relief, give us hope, or silence a guilty conscience” (Shelby 2003, 171). And it is when an agent’s psychological well-being is systematically affected by their oppressed circumstances that ignorant beliefs so-produced would constitute false consciousness.

Stanley’s processualist account of false consciousness is another such instance of an epistemological answer to the Normative Problem that tacitly relies on some conception of well-being. As we saw earlier, the oppressed purportedly lose knowledge due to the pragmatic encroachment of their disadvantaged circumstances: one reason is that oppressed agents are frequently under high-stakes situation that shake their confidence in their own beliefs, and that “knowledge, as well as the norms of action, presupposes full belief” (Stanley 2015, 256). Notably, as with psychologism,  

---

94 I am taking an externalist stance on epistemic justification here. As Amia Srinivasan argues, epistemic internalism cannot on its own account for how an agent’s beliefs under false consciousness is problematic, since justification for the epistemic internalist “is a matter of fit with one’s evidence, or with one’s epistemic reasons, or more generally with how things look from one’s own perspective on the world—where it is presumed that such facts are facts about one’s (nonfactive) mental states” (Srinivasan 2020, 399–400).
this is partly a prudential concern: the oppressed is thought here to fail epistemically because their well-being, however conceived, is threatened (for example, one would be more confident in making an accusation against one’s employer for workplace negligence as a citizen compared to a refugee).

Moral and political answers—welfarist or otherwise—may be more straightforwardly characterised as assuming that notions of well-being are integral to the badness of false consciousness. The moral and political failings of false consciousness have been observed to include the ways in which it directly harms an agent in terms of their autonomy (Superson 1993, Hay 2013) or self-respect (Hay 2013, Silvermint 2013), as well as how it would indirectly compromise their personal security (Cudd 2006, Powers and Faden 2019), or health and bodily integrity (Nussbaum 2011, Haslanger 2020). That is, a common and primary way in which false consciousness is understood to be morally or politically bad is because it compromises prudential goods that are otherwise good for the agent, that are constitutive of their well-being (cf. Terlazzo 2015).

Whether or not these various non-prudential answers to the Normative Problem are successful on their own is not my concern here (though I am sympathetic to a great many of them). For, whether or not they are successful as epistemic, moral, or political answers, a general theory of a false consciousness should nevertheless consider how conceptions of well-being can provide an answer to the Normative Problem. And it would also be crucial that they do so, if some of the aforementioned non-prudential approaches are to rely on such conceptions. But before we can evaluate the extant prudential approaches to the Normative Problem (in §3), we need to first be clear about what is required of a solution to it. In what follows, then, I will analyse the Normative Problem and make explicit the requirements that need to be met for any successful answer to it. Given that the Normative
Problem takes place within the context of how oppressed agents may be enjoined to resistance, this means that I will be attending to the requirements for a successful critique of false consciousness.

2. The Requirements of a Solution

A successful critique, whether of ideology or false consciousness, would conduce some resistant praxis among the oppressed. The extent of such resistance depends on the proliferation and extent of false consciousness among this (likely) heterogenous group: from individual refusals to adopt pro-attitudes towards certain ignorant beliefs, to activism that targets structural mechanisms, and to revolutions against whole systems (Fanon 1961, Hay 2013). For example, the critiques by the suffragists were sometimes successful in at least suspending some of the anti-suffragists’ active support of the anti-suffragist campaign. And, ostensibly, the successful critiques by the suffragists contributed (to an arguable extent) to their eventual enfranchisement.

Critique (whether of ideology or false consciousness) would thus aim to disrupt the looping mechanisms—involving the classified, classifiers, institutions, experts, and knowledge—that had stabilised the relevant false consciousness (to recall from the previous chapter). A critique would disrupt these mechanisms by revealing the relevant set(s) of beliefs to be a result of these mechanisms, rather than ‘just the way things are,’ as well as how the beliefs and their impression of ‘naturalness’ systematically contribute to the harms the criticised suffer. Even accounting for non-ideological reasons (such as coordination problems) for why some more demanding forms of resistances might not obtain (cf. M. Rosen 1996, Cudd 2006), successful critique would, at the very least, bring about psychological resistance

95 I discuss this later in greater detail in Ch. IV.
against the relevant ideological beliefs: for example, non-binary individuals
who—albeit still interpellated under binarised social structures—would, in
Carol Hay’s words, “simply refuse to believe what oppressive social
messages were telling [them] about the character or worth of people like
[them]” (Hay 2013, 141).

Given the shared aim of both types of critique, philosophers might be
tempted to adopt a ‘top-down’ approach to the Normative Problem, deriving
from the wider project of ideology critique. I argued in general against such
an approach to the concept of false consciousness in Chapter I (in particular,
§2.2), but it is worth revisiting it here with specific reference to the Normative
Problem. One might think that all we need to do is explain “what makes
ideologies problematic and whether there is a single feature or set of
(systematically related) features that makes them problematic” (Celikates
2017, 55). The idea then would be that whatever the normative failures of an
ideology might be, false consciousness would simply inherit them in virtue of
the agent’s internalisation of that ideology. The critical norm presupposed in
ideology critique, then, can function as the critical norm to be employed in a
critique of false consciousness. One might, for example, take after Sally
Haslanger, who notes that “ideology sustains injustice by masking or
distorting what’s good, right, just,” and thus “fails us both morally and
epistemically” (Haslanger 2020, 36; cf. Railton 2000). Under such a ‘top-
down’ approach, no independent determination of the badness of false
consciousness would seem to be necessary: false consciousness would
similarly be said to be morally and epistemically problematic, inasmuch as it
masks or distorts what is ‘good, right, just.’

Yet, when the concept of false consciousness is used to criticise particular
agents, a top-down approach of ideological critique is likely to backfire. This

96 For a criticism of recent approaches to ideology critique by the New Ideology Critics, see
Sankaran 2020.
is because ideology critique addresses oppressive structures at the level of schemas. Because of this, and given the targeted focus on particular agents provided by the concept of false consciousness, there are two reasons why the Normative Problem of false consciousness should be distinguished from the normative problem of ideology. In my view, these reasons give rise to two corresponding requirements for any answer to the Normative Problem of false consciousness.

It might be useful to illustrate how these requirements arise with a reference to a tacit instance of ideology critique among the flurry of replies and rejoinders published in the wake of the June 1889 Appeal. In the July issue of the *Nineteenth Century* immediately following the Appeal, Dilke’s reply attributed the Appeal’s opposition to suffrage partly to the “strange” belief that the vote was “a masculine adjunct” (Dilke 1889, 98). Such a belief, she observed, arose from the “patriarchal” manners of the “old times” and was maintained in a way that was unresponsive to the increasing need of women to participate in parliamentary politics, so as to remedy the unjust denial of their “social and economic equality” (ibid., 100, 102). She described the women of the Appeal as follows:

People without votes who deliberately say they do not want them are like a crowd standing outside a concert-hall, eager to hear every note of the music, refusing to take the key and unlock the door so that they may enter, and yet triumphantly pointing out to those who advise the use of that simple implement that, the windows being partly open, faint echoes of the melody reach them now and again if they listen with sufficient attention. (ibid., 99)

---

97 This is distinct from the line of argument pursued by Robin Celikates against totality-oriented conceptions of ideology, and with whom I am much in agreement (Celikates 2017).

98 To note, Dilke’s understanding of the implications of this did not stretch as far as to challenge the gendered division of labour of the time or marital institutions—contra someone like Mona Caird (see Caird 1888).
In effect, Dilke was providing a structural explanation of anti-suffragism, that it was an expression of a patriarchal schema. Her reply is thus an instance of ideology critique, an implication being that the anti-suffragists were under false consciousness. An anti-suffragist’s false consciousness, then, would be problematic inasmuch as it expressed a problematic ideology that did not “protect the earnings” of working-class women or encouraged their “self-reliance” in a way that it did for men (ibid., 98).

This implication of a tacit accusation of false consciousness was not lost on Creighton (a co-instigator of the Appeal alongside Mary Ward), who took up Dilke’s metaphor in her rejoinder in the subsequent August issue of the Nineteenth Century. Criticising suffragists of “labor[ing] under […] delusions” when “[t]he present need [was] that women should do their own work better,” she fired back with the following (Creighton 1889, 349, 353):

Mrs. Dilke says that we who do not want the vote are like those who will not open the door to go into a concert hall to hear beautiful music, but content themselves with the faint echoes that reach them through the windows. We might retort by saying that those women who, not content with what they have, still demand the franchise, are like those who, deaf through misfortune or their own fault, stand within the concert hall but cannot hear the music. (ibid., 354)

Creighton affirmed that the vote was indeed a ‘masculine adjunct,’ being “a part of the machinery of government which has always belonged to one sex,” and that while “[w]e may try to change society, we cannot change nature; sex will remain.” Women, she further maintained, “are not a class but a sex, and that [the British] representative system knows nothing of classes as such” (ibid., 353). Suffrage, to return to the words of the Appeal, would

---

99 In her final appeal to the House of Lords before the Representation of the People Act 1918 was passed (which granted suffrage to women, given property qualification), Ward also employed a tacit notion of false consciousness, accusing suffragists of having been ignorantly beholden to the Labour Party’s “political calculation” to usher in a “Socialist promised land,” which would limit improvement of the “life and well-being of women” (Ward 1918, 52–53).
“[lead] to a total misconception of woman’s true dignity and special mission," one consonant with “the disabilities of [her] sex, or by strong formations of custom and habit resting ultimately upon physical difference, against which it is useless to contend” (“An Appeal,” 781, 785).

In this exchange between Dilke and Creighton, we see that the very norm presupposed in Dilke’s criticism, that ‘men and women ought to be treated as socially and economically equal,’ was in dispute. From Creighton’s perspective in this rejoinder, it was Dilke and the suffragists who were themselves acting under an ideology that was detrimental for women. So which norms should Dilke have appealed to, if she had hoped for her criticism to enjoin Creighton to resistance (however extensive) rather than backlash? Keeping in mind that the norm of ‘men and women ought to be treated as socially and economically equal’ was far from being widely accepted in British society at the time, I will now attend to Dilke’s tacit ideology critique, as opposed to a critique of false consciousness targeted at Creighton, as a negative heuristic for us to draw out two methodological requirements for a successful critique of false consciousness: Continuity (§2.1.) and Endorsement (§2.2.). I will then use these requirements to subsequently evaluate extant prudential approaches to the Normative Problem in the following section (§3).

2.1. Continuity

First, while ideology critique addresses oppressed agents as group members, a critique of false consciousness addresses oppressed agents as particular agents. As I have argued in the previous chapters, the specific utility of the concept of false consciousness lies in its agent-targeted
attention to the various mechanisms—both structural and individual—that conduce the acquisition and maintenance of the relevant set of beliefs by a particular oppressed agent. Ideology critique, unlike a critique of false consciousness, would not on its own articulate why a particular addressee’s beliefs and corollary behaviour are products of these mechanisms and instantiations of ideology. The trouble here, then, is that unless those criticised already see the normative standards of the critic as applying to themselves, they would have little to no reason to acknowledge the appropriateness of the criticism and be motivated to any form of resistance—especially given that they are conceptualised in ideology critique as having a diminished agency they do not (purportedly) experience.

Overlooking this difference often leads to a polarisation within the oppressed group: pitting the oppressed into those who are ‘cultural dupes’ against those others with ‘true consciousness’ (Khader 2011, Lukes 2011). As we see in the above exchange between Dilke and Creighton, the norm to which Dilke appealed was not only alien to Creighton’s own set of evaluative beliefs, but also thereby effectively backfired and kept her at bay from the suffragist cause at the time. Creighton’s dignified tone in the Appeal in June was abandoned to meet the polemical tone of Dilke’s reply. Although the Appeal had initially described “the pursuit” of suffrage as “not only vain but demoralising,” it refrained from passing judgments on the characters of the suffragists themselves (“An Appeal,” 785), it was only now that Creighton would publicly pronounce the suffragists deluded, malcontent, and ‘deaf through misfortune or their own fault.’ Dilke’s characterisation of the vote as a ‘masculine adjunct’ did not provide the critical force she had hoped it would—and was in fact co-opted affirmatively in Creighton’s rejoinder, which was accompanied by a twenty-fold increase in signatories compared to the initial hundred and four women of the Appeal (Creighton 1889, cf. Bush 2007). As we can see from this occasion, if successful collective resistance lies in the solidarity of at least a sufficient number of the oppressed to set in
motion the necessary social changes, the invocation of false consciousness, when the relevance of the standards of critique are not acknowledged, would be self-frustratingly counter-emancipatory.

In light of the above difference between ideology critique and a critique of false consciousness, I think we can articulate the first, *Continuity* requirement for answering the Normative Problem:

*Continuity:* An adequate answer to the Normative Problem must account for how the critical norm is *hermeneutically continuous with* the values of the particular agent that is being addressed.

I understand ‘continuity’ here to be more than a mere lack of conflict. The relevant norm for critique has to be inferable from—or even implied by—the very evaluative beliefs by which the particular agent uses to make sense of their social situation(s). Furthermore, continuity has to pertain to the actual pool of agent’s concepts. In order for a criticism to address a particular agent (*ad hominem*, as it were), the relevant values cannot be considered under idealised or hypothetical conditions: as what a representative member of such-and-such a social group would hold or what the agent would if certain situational variables were held fixed (cf. Enoch 2017). Without *Continuity* understood as such, it would not be possible for the critic to invoke standards to which the particular agent being criticised should, from their point of view, respond.¹⁰¹

¹⁰¹ Consider how this compares to moral criticism: if successful moral blame is to result in, say, expressions of regret and remedial action by the blamed, a success condition would be for addressee to share in the recognition of their wrongdoing on the basis of shared norms (Coates 2020). While we may sometimes engage in objective modes of blame, as expressions of our commitment to the relevant moral standards, such an act would not on its own address the particular blamed agents as fellow participants within the same moral community (Mason 2019). I say more about the relationship between moral blame and a critique of false consciousness in Chapters IV and V.
Consider, for example, how the concept of online data privacy would not be continuous with the evaluative concepts held by suffragists or anti-suffragists—or anyone in fin-de-siècle Britain for that matter, given that evaluative concepts then were formed and acquired more than a century before the “Age of Surveillance Capitalism” (Zuboff 2019). In contrast, say, Adrienne Rich’s concept of ‘matrophobia,’ which is “not the fear of one’s mother or of motherhood, but of becoming one’s mother” (Rich 1976, 235), would be arguably hermeneutically continuous for a suffragist like Mona Caird. Caird’s The Daughters of Danaus (1894) has been observed to “read like a fictional exposition of Rich’s theory from the perspective of a daughter,” providing an “analysis of mother-daughter relationships […] informed by the desire to devise a new script for female socialisation and bonding by laying bare its negative dynamics under patriarchy” (Heilmann 1996, 82, 86). Meanwhile, the concept of matrophobia would not be continuous for Creighton who, as we noted, had championed the maternal duties of women in both her anti-suffragist and suffragist campaigning, delivering a lecture even in the midst of the latter to University College, Liverpool about “the tremendous responsibility of the mother” and the “vital importance that all her powers should be fully developed, that she should be in every way the best that she is capable of becoming […] the guide, the source of inspiration to her children” (Creighton 1909, 113). Creighton, in fact, recalls having an extremely good impression of how her mother carried out her maternal duties, having “encouraged [Creighton’s] studies and never tried to get [her] to spend much time over sewing”—even though, despite Creighton mother’s “finer intelligence” than her father’s, “the strain of her large family must have made it impossible for her [mother] to maintain any definite habits of reading and study” (Creighton 1994, 3, 17, 28; cf. Bush 2007).
2.2. Endorsement

The second implication of the agent-targeted focus of a critique of false consciousness on a particular agent, as opposed to the schematic focus of ideology critique on a social structure, pertains to how such an agent is, as it were, a site where various schemas and ideologies may converge. As we have noted a number of times, and as most standpoint and critical theorists are keen to emphasise, the starting points of an answer to the normative problem of ideology critique are found in the lived experiences of the oppressed and their local contestations of the ruling ideology (Wylie 2003, Honneth 2017, Pohlhaus Jr. 2020, Celikates forthcoming). Critical theorists are keen to underscore that these contestations (or ‘oppositional consciousness’) are more widespread than in “the nightmare which haunts the Frankfurt School” of “a society of happy slaves, genuinely content with their chains” (Geuss 1981, 83–84; see also Ch. I, §1.4).102

Yet, a methodological worry arises when we consider how agents under false consciousness are not all impacted to the same degree. The arguments and campaigning of the leading suffragists very much reflected their own social positions as largely upper- and middle-class white women. Suffragist critiques of the prevailing gender ideology were largely based on a specific form of normative discourse that obfuscated and perpetuated the oppression of the working class as well as other racialised and gender groups. As mentioned, many suffragists campaigned for the vote on the basis of property qualification and were not always welcoming to working-class women. And similarly, Millicent Fawcett expressed to a crowded lecture hall a mocking disbelief that Maori women had been placed in “a superior position to that held by the women of England” by their enfranchisement (Fawcett 1909, 7). Why and how ideologies are bad for

---

102 This nightmare has been rightly criticised from various angles as being at least empirically very unlikely (see, e.g., Rancière 1983, Mills 1990, Scott 1990, Celikates 2017).
particular agents are not uniform across the relevant social structures as well as these structures’ overlaps and interections with others (cf. Cunningham 1983, Balkin 1998). As such, solving the normative problem of ideology critique cannot involve mere simple extensions of how it negatively affects some agents (such as middle- to upper-class white women) to characterise how the group is negatively affected as a whole: how heterogenous contestations are to scale up to the level of schematic critique has long been debated over the decades (see, for example, Fanon 1961, Maynard 1995, Celikates 2018, Haslanger 2021). This is further complicated by the need to take into account how oppositional consciousness sometimes arises out of, not a resistant schema, but yet another ideology (cf. Lugones 2003, Curry 2020). The suffragists, for example, relied heavily on disability tropes in order to distinguish themselves from groups of men whom they regarded as undeserving of the vote: Dilke’s reply, for instance, juxtaposes women with “ill or crippled, immoral or sentimental, illiterate or drunken” men (Dilke 1889, 98). This is not to say that they were necessarily consciously ableist, but that their use of such tropes nevertheless reproduced an ableist ideology.

For these reasons, I think that the Normative Problem of false consciousness, which attends to how oppressive structures are bad for a particular agent, should have methodological priority in the present context over the normative problem of ideology critique. In this way, we may take better precaution to avoid misrepresenting those agents under the ideology being considered. Otherwise, social philosophers would risk outright neglecting and even reinforcing their experiences of oppression by reproducing an ideology or other ideologies within the critique itself.

If resistance aims at ultimately succeeding to bring about structural change, while at the same time avoiding a reproduction of ideology, then an answer

---

103 For example, the longstanding discussion on the feminist notion of consciousness-raising (e.g. Rowe 1975, MacKinnon 1989, Stanley and Wise 1993, Haslanger 2020).
to the Normative Problem of false consciousness should aim to appeal to as broad a range as possible of oppressed agents, to enjoin them to contestations as part of broader resistant movements. This would also include those, who—while they may not make up an entire society as in the Frankfurt School’s nightmare of a society of happy slaves—do experience oppression as enablement under false consciousness, and especially since false consciousness is an important dimension of oppression that would be otherwise neglected. Because of this, I think we can articulate a second, *Endorsement* requirement for answering the Normative Problem:

*Endorsement:* An adequate answer to the Normative Problem must be able to elicit some form of actual *endorsement* (desire, wishes, etc.) of the critical norm by the particular agent being criticised for false consciousness, specifically endorsement that is necessarily tied to their own motivations (to resist or contest a system of oppression).

That is, the relevant norm should be included in the “*subjective motivational set*” of the agent (Williams 1979, 102), especially since the critic—as themselves embedded within social structures—cannot on their own guarantee that any other norm would harbour the possibility of belonging to another ideology. Like continuity, such endorsement has be *actual*, in order for critique to move the agent to resist. Idealised or hypothetical endorsement, while it might have some limited use in theorising resistance, has no utility in emancipatory praxis.

Creighton’s later conversion to the suffragist cause in 1906 provides some clue as to headway that Dilke’s 1889 reply could have made with respect to *Endorsement*. Creighton explained that her eventual conversion had not been a result of suffragist critique but, rather, had been based on her gradual realisation that the franchise and party politics was a necessary instrument
for women to carry out their gender roles, whether in the household or in
education and healthcare (Bush 2007). That is, epistemic friction effected
the erosion of her false consciousness: her economic beliefs, which changed
as a result of her increasing familiarity with working-class women (as herself
an upper middle-class woman) after becoming the first president of the
National Union of Women’s Workers (NUWW) in 1895, now ran up against
her political beliefs about the parliamentary exclusion of women.

In Dilke’s 1889 critique of antisuffragism as an expression of patriarchal
ideology, however, the salient norm was ‘women ought to be treated as
social and economic equals’ rather than ‘the realisation of gender roles
ought not to be impeded.’ Further, Dilke’s critique, as it stands and as
noted above, reproduced an ableist ideology (whether or not this was also
subscribed to by Creighton). Had Dilke’s criticism taken place after 1895 and
relied more heavily on the shared beliefs between her and Creighton that
men and women’s “interests and capacities are different,” and had Dilke
focused instead on how Creighton’s “use [of] the same arguments again and
again” reflected an unfamiliarity with the changing economic schemas of
working-class women, she could have had appealed to a norm that would
be endorsed by Creighton and taken up the frustration of realising gender
roles as the normative basis of her criticism (Dilke 1889, 97). In this way,
such a (counterfactual) criticism could have garnered some endorsement
from Creighton in 1889, if Dilke had simply emphasised inconsistencies
within Creighton’s set of evaluative beliefs (call it a ‘criticism from mere
inconsistency’). And Dilke would not have effected such a backlash, as
well as avoided introducing/reproducing another ideology in her criticism.
This, however, is not to say that this would suffice for a successful critique of
Creighton’s false consciousness that would enjoin her to resistance, but only
that the criticism would meet Endorsement: the mere frustration of the

---

104 I pay more attention to the details of Creighton’s conversion in Ch. IV.
105 The latter norm, of course, requires further and more comprehensive critique.
106 Jaeggi calls this “internal” critique (Jaeggi 2009).
realisation of gender roles, after all, does not on its own imply the need for parliamentary representation. Recall that epistemic friction on its own does not guarantee emancipatory consciousness, and may be responded to with, say, wilful ignorance (which is likely why Creighton ignored these aspects of Dilke's reply). **Continuity** is still required for Creighton to endorse a critical norm that relates the inconsistency in her beliefs to structural oppression, in order for her to realise that she was under false consciousness.

To note, the two requirements for a solution to the Normative Problem, **Continuity** and **Endorsement**, may sometimes come apart in critique. In such cases, emancipatory praxis is not brought about and critique is said to have failed. We see **Continuity** without **Endorsement** in cases where oppressed agents are able to recognise evaluative norms as addressing them and their beliefs as problematic, but owing to some form of defeatism (or nihilism or pessimism) are not at all motivated to resistance (West 1993, Tessman 2009, Wilderson 2020). This is congruent with the recent post-Marxist observations about the problem of “enlightened false consciousness” in late capitalist society: where “one knows the falsehood very well, one is well aware of a particular interest hidden behind an ideological universality, but still one does not renounce it” (Žižek, 26; cf. Bourdieu and Eagleton 1992).^{107} We can find a concrete instance of this phenomenon when we pay closer attention to Natalie Wynn’s retrospective on her false consciousness about her own sexuality (which we had briefly considered in Chapter II, §3.3) and, in particular, the failure of her self-criticism of it. She confesses, realising that she was under false consciousness about her sexuality as a trans woman, after being introduced to the concept of compulsory heterosexuality (‘comphet’):

^{107} A version of this may already be found in the *Dialectic of Enlightenment*, when Adorno and Horkheimer gesture to this at the end of “The Culture Industry,” noting that “the triumph of advertising in the culture industry is that consumers feel compelled to buy and use its products even though they see through them” (Adorno and Horkheimer 1947, 167).
So, to be honest, I still experience comploth thoughts and fantasies. Like, for a long time, I’ve had what seemed to be romantic fantasies about marrying a man. But after a lot of introspection, I’ve realized that this is not really a fantasy of private happiness and fulfillment. It’s a fantasy of social prestige and acceptance.

Like, can you imagine if I got married to a tall, handsome man, at an opulent wedding with lots of photographers posting to Instagram? Ugh, I’d be queen of the trannies! I’d be a super heckin valid uwu.

But I don’t actually wanna be married, I wanna get married. I wanna have a wedding. Basically, I wanna wear a wedding dress. I want attention, okay? Please pay attention to me.

So I can totally deconstruct the fantasy, but it’s still really hard for me to let it go. I cling to the reassuring conformity of heterosexuality. Because there’s a level of gender instability and deviancy that’s inherent to homosexuality. And until recently, I wasn’t secure enough in my gender to handle that.

[...]

God there’s so much pressure on trans women to be extremely gender conforming. And even more so if you’re a ‘representative of the community.’ You feel like you have to be a Stepford wife or else cis people are gonna throw us all in camps.

And, early in my transition, I kind of worshiped straight trans women. Like, I looked up to them and I wanted to be just like them.

Now back then there were a handful of trans lesbians on YouTube trying to make the case to the masses that being a trans lesbian is a legitimate thing. But in 2017, sweaty, the masses were not having it.

I watched those women get eaten alive. And as a result of my witnessing that at a super fucking vulnerable moment for me, I internalized the message that a trans lesbian is just not an okay thing to be if you wanna be tolerated or respected by anybody other than a handful of communists. So I began my transition with essentially no role model of a widely respected and admired trans lesbian.

And it doesn’t help that most forms of transphobia are harsher on gay trans women than they are on straight trans women. Like, take this trope that trans women are men who transition to creep on women in bathrooms. In response to that, it feels really good to be able to say, ‘I’m not even attracted to women. I’m just a petite heterosexual biogirl. I’m surely not some kind of six-foot monster who likes women.’

125 of 237
*sigh* 😞
Shame! 😢

It does make me feel like a monster sometimes, like a mutant that has no place in society. And this shame has actually made it more difficult for me to accept that I’m a gay woman, than it was for me to come out as trans in the first place. It’s like I made a kind of subconscious bargain where I traded my sexual orientation for my gender identity, and so I finally transitioned only to spend the next couple years living with a different kind of denial. And that denial got pretty deep and pretty dark. (Wynn 2020)

From this retrospective, we can see that Wynn was able to criticise her own instance of false consciousness, where the criticism presupposes the norm ‘the sexuality of trans individuals ought to be just as unencumbered as that of cis individuals.’ Yet, even despite realising that the beliefs about herself were a result of hermeneutical injustice and stereotype threat, the self-criticism was yet unsuccessful in motivating resistance against ‘the compket thoughts’ (and hence become critique proper). There was a defeatism in her ‘clinging to the reassuring conformity of heterosexuality.’ Her false consciousness was only eroded as a result of a “combination of […] being politically canceled by Twitter, being romantically canceled by Joanne, and also losing [Joanne] as [her] closest friend,” which sent Wynn “into a catatonic, rock-bottom kind of depression” (ibid.). In other words, her eventual resistance was brought about as a result of radical marginalisation under the relevant gender schema, rather than successful critique.

We also see actual cases of Endorsement without Continuity with oppressed agents who are able to accept the presence of structural disadvantages and corresponding ideological distortions—and may perhaps even participate in collective resistance—but yet do not see themselves as structurally oppressed. That is, there is an endorsement of a norm which enjoins them to resistant practices, but it does not address them ad hominem. For examples of this, Tamar Jugov and Lea Ypi cite “women in academia who agree that in
general their social group (women in academia) suffers from some form of structural injustice but affirms that it is immune from it” or “hard-working immigrants, who acknowledge the existence of xenophobic norms in their new society but think that they are lucky not to be affected by those because they are blessed with nonxenophobic neighbors, colleagues, or employers” (Jugov and Ypi 2019, 18; cf. Godfrey and Wolf 2016).

For successful critical application of the concept of false consciousness (i.e. a successful critique of false consciousness), therefore, both requirements have to be satisfied in providing a solution to the Normative Problem. In what follows, I consider how extant prudential approaches to the Normative Problem fail to satisfy these two requirements.

3. Critical Failures of Extant Prudential Approaches

In this section, I will present the case of the false consciousness of Simkins (§3.1), before going on to argue that extant prudential approaches to the Normative Problem fail to satisfy Continuity and Endorsement when trying to account for what exactly is prudentially bad about Simkins’ false consciousness, and therefore fail to provide a general account of what is prudentially bad about false consciousness as such (§§3.2–3.3).

As we noted at the start of the chapter, prudential norms seem particularly apt as providing some standards for critiquing false consciousness, since as subject-relative standards, they would have less risk of being alien or alienating for the target of critique. Corresponding to the distinction between realist and anti-realist theories of well-being, ‘subject-relativity’ can be disambiguated in two respective ways: either well-being is relative to a subject inasmuch as it pertains to facts about them that are attitude-
independent; or well-being is relative to a subject inasmuch as it pertains to facts about them that are attitude-dependent.

Realist prudential approaches that have been taken to the Normative Problem have tended come in the forms of objective list theory (Silvermint 2013) and perfectionism (Nussbaum 2011, Enoch 2020, cf. Khader 2011). In the former, the determining particulars of an agent’s well-being are a list of goods that either are meant to have intrinsic value independent of the agent’s individual features, or that specifically manifests the capacities characteristic of human beings (for example, virtues). In the latter, the determining particulars concern the agent’s specific development of the “potentialities, capacities, and faculties, that (under favorable conditions) they naturally have at an early stage of their existence”—a development which may then pertain to the “peculiar circumstances and idiosyncrasies of that particular individual” (Kraut 2007, 131, 4). The appeal in such approaches is that an agent, irrespective of their ignorance about their own well-being (which may be problematic under false consciousness), can be straightforwardly understood as being in a prudentially bad state relative to the facts of them being the kind of agents they are. This seems, on the surface, to also be consonant with the way Marx describes “man” to be “a species-being, and has to confirm and manifest himself as such both in his being and in his knowing,” such that critique is aimed at overcoming one’s alienation from this ‘species-being’ (MECW 3: 337).

Anti-realist approaches that have been taken to the Normative Problem, on the other hand, have tended to come in the forms of desire-satisfaction or preference-satisfaction theories (see, for examples, Geuss 1981, Superson 2005, Harvey 2010, Terlazzo 2016). In either case, what is intrinsically good for an agent is simply what they desire or prefer—however such desire or preference might be qualified. The appeal here is that such anti-realist theories maintain that what is good for an agent has a necessary connection
to motivation. Anti-realists, on the surface, also have some claim on their approach being consonant with the canonical understanding of the task of critical theory as the “self-clarification [...] to be gained by the present time of its struggles and desires [emphasis mine]” (MECW 3: 145). However, as I will now go on to argue, both extant realist and anti-realist prudential approaches to the Normative Problem fail to satisfy either Continuity, Endorsement, or both.

3.1. The Case of Simkins

So far, I have been assuming a critique of false consciousness to pertain to a critique of an entirely consistent set of beliefs. It is, of course, entirely plausible that there are norms that are inferable or implied by one’s evaluative beliefs that come into conflict with each other—or even evaluative beliefs outside of the set of beliefs belonging to false consciousness. The conflict of norms when these sometimes occur is an instance of the epistemic friction discussed in the previous chapter (Ch. II, §3.4), which occasions meta-lucidity that afford the possibility of erosion. We saw this earlier in the case of Creighton’s conversion to suffragism (in §2.2). Creighton’s erosion of false consciousness occurred as a result of the epistemic friction she experienced, having been able to acquire sociological beliefs about the macroeconomic situation of working-class women during to her presidency in the NUWW—a position that would not have been available to her had she not been a member of upper middle-class society. Put in Medina’s words, she was thus able to gain “insights in to the functioning of perspectives that [made] it possible to redraw [her] cognitive maps, to redescribe [her] experiences, and to reconceptualize [her] ways of relating to others” (Medina 2013, 47).

108 Recall also Lukács’ concern for class consciousness as involving “thoughts and feelings” (Lukács 1923, 31).
Because of this, I wish instead to now turn to the case of the false consciousness of Simkins, in assessing the adequacy of realist and anti-realist prudential approaches to the Normative Problem—despite having considered the false consciousness of Creighton in the earlier part of this chapter. As a case of false consciousness among working-class women anti-suffragists, Simkins has no similar sociological access to the macroeconomic beliefs that Creighton acquired after 1889. And, from Simkins’ response to Ward’s anti-suffragist campaigns in 1909 up to 1949, Simkins remained steadfastly committed to the idea that suffrage had not been prudent for women as a whole. This thus presents a particularly poignant challenge for formulating a critique of false consciousness, which seeks (in principle) to enjoin resistance from any oppressed individual. Any criticism from mere inconsistency (such as the counterfactual criticism that Dilke could have made against Creighton), would not enjoin resistance for Simkins or even give her pause, since it would only serve to render consistent her anti-suffragist beliefs. Attending to how an answer to the Normative Problem may be provided in the case of Simkins would thus allow us to see how critique might be possible for oppressed individuals who lie within the overlaps of systems of oppression that do not afford erosions of false consciousness through epistemic friction—that is, how we might have a foothold for critique, even if the nightmare of the Frankfurt School were to come true.

Notably, Simkins’ false consciousness was much more representative of the British population then. As Simkins herself is keen to stress, her position cannot easily be dismissed as an anomaly—as many theorists, then and

---

109 At least according to her genre-bending Polonius (1949), which was circulated among family and friends. The publication was made twelve years before her passing in 1961.
110 Criticisms based on mere inconsistencies, notably, leave untouched whether or not a given belief set might be problematic in itself (Jaeggi 2009).
now, might be so wont to dismiss false consciousness (for example, Terlazzo 2016, Dorsey 2017):

Finally M. E. S. [Simkins’ initials] would like to add that she is very sure from the personal contact with the working classes enjoyed by one who is a worker, not an aristocratic lady residing in a fashionable quarter and theorising about workers from a safe distance, that she voices an enormous public of working women in protesting against the imposition on them of the suffrage, and does not only represent her own personal friends […] (Simkins 1908, 149)

As historians have corroborated, prior to the First World War, both the referenda in the US and the polling of women municipal voters in Britain were consistently in favour of anti-suffragism—and British suffragists themselves lacked confidence in carrying out an official referendum of women on the matter (Bush 2007, 5–6). There were, in fact, close to half a million anti-suffrage petition signatures by women before the war.

With that said, let us now turn to the case of Simkins:

Simkins: Maud Ellen Simkins writes impassioned pamphlets that bolstered the anti-suffragist campaigns [functional ‘falsity’], proudly on the basis that ‘women are domestic creatures’ [epistemic ‘falsity’]. She holds this belief in virtue of her position in Victorian class society [genetic ‘falsity’]. Simkins argues that suffrage would disproportionally affect women workers like herself, who would be saddled with “the triple burden of wage-earning, housekeeping, and political responsibility,” maintaining that the initial double burden of wage-earning and housekeeping was “because the wealthier women have been neglecting the first elementary duties of women” (Simkins 1909b, 790–793). She argues that the fulfilment of the ‘elementary duties’ were desirable in themselves and definitely preferred over the “exile” of wage labour—much more over the responsibility of the vote.
Simkins was, of course, quite mistaken about the reason she had to bear the double burden of both wage-earning and housekeeping in the first place: in her writings, she did not seem to consider that the economic oppression she astutely perceived and experienced in fact systematically overlapped with the gender oppression that came with a necessarily domestic conception of womanhood.

As an aside, it is noteworthy that her writings also included imperialist and social-Darwinist strands. Simkins mentions, for example that “women’s methods of competing with men have so far landed them in being the Chinamen of the Western labour market” (Simkins 1908, 81). Such a belief about (white) women’s position in the labour market was not uncommon at the time: many others, such as the sociologist Beatrice Webb, also shared the view that working (white) women were on the same rung of the labour ladder as Chinese men (Benton and Gomez 2008). Yet it is crucial to underscore that this ladder was seen as not only economic but also evolutionary (and thus the ladder would also have in principle included the aforementioned Maori women that Fawcett decried): Chinese people in Britain were often treated as objects of disgust by both suffragists and anti-suffragists, being generally seen as members of a less (or at best more slowly) evolved race (Lamont 1894, Tchen and Yeats 2014, Witchard 2014).

Specifically, Chinese people were seen as “the mule among the nations—capable of the hardest task under the most trying conditions; […]

---

111 Webb, for example, writes in her diary during her 1873 visit to San Francisco that “it was impossible for [her] to stay in the [Chinese theatre] for more than five or six minutes, the noise was so deafening, and then being in such close quarters with John Chinaman [i.e. a stock caricature of Chinese labourers] was not exactly pleasant” (Webb 1926, 67).
unlovable and useful in the highest degree”; and who has “no more sympathy than the lower animals” (Warnford-Lock 1907, 39–40). As such, Simkins’ comparison had not only economic implications but imperial and racial ones too: by comparing (white) women to ‘the Chinamen,’ she was tacitly implying that (white) women not only were now operating in environments unsuited to their nature, but also were faced the threat of devolution—especially given that it was understood under Victorian social Darwinism that “men and women of civilized races had evolved pronounced sexual differences,” while the absence of such difference was a mark of savagery (Bederman 1995, 25).

Nevertheless, given her circumstances, it might seem genuinely bad for her as an individual to be saddled with the additional burden of suffrage. What is in dispute, however, is the conception of well-being that would be able to acknowledge the apparent negative value of suffrage for her and yet eventually enjoin her to regard it as positive. With regard to the Normative Problem, what we are concerned with in Simkins is what kind of approach to well-being must be true in order for a critic to justifiably claim that Simkins’ false consciousness was indeed bad for her, but yet provide her with motivating reasons for resistance. That is, what kind of approach to well-being is required for the critical norm, implicitly referenced in criticising her for false consciousness, to be hermeneutically continuous with the reasons Simkins gives for why suffrage would be prudentially bad for her (Continuity), yet also one that Simkins would actually endorse (Endorsement).

---

112 The proliferation of social Darwinism extended as far as the crown colony of Singapore outside of the imperial metropole: many of the Straits Chinese there saw (and sometimes still see) themselves as less-than-competitive compared to the Chinese in China precisely because of their part-Malay heritage—inherting the British conception of the Malay race as indolent savages and seeing their British or even Chinese education as a form of racial uplift (Chew, Young, and Tan 2019, Ong 2020, Wilson and Lai forthcoming, cf. Alatas 1977).

113 As mentioned, suffragists were not actually campaigning for the working class to get the vote, given the property qualification. Working-class women like Simkins would only get the vote with the Representation of the People (Equal Franchise) Act 1928, a whole decade after women who satisfied the property qualification got the vote.
3.2. Realism

A realist approach seems to provide a clear critical stance with respect to Simkins: Simkins’ ignorance, constitutive of her false consciousness, is irrelevant for an evaluation of her well-being. What would matter, under such an approach, is whether she attains a pre-determined list of goods, according to objective list theorists, or develops and exercises the relevant capacities, according to perfectionist theorists. False consciousness is bad, under the realist approach, simply because it prevents her from achieving these goods or exercising these capacities. I think, however, that such realist theories in general fail to satisfy both requirements of the Normative Problem.

Readers might have noted that the Endorsement requirement parallels the anti-alienation requirement often imposed on theories of well-being. Connie Rosati’s formulation of this is as follows:

[S]omething X can be good for a person A only if two conditions are met:

1. Were A under conditions C and contemplating the circumstances of her actual self as someone about to assume her actual self’s position, A would care about X for her actual self;

2. conditions C are such that the facts about what A would care about for her actual self while under C are something A would care about when under ordinary optimal conditions. (Rosati 1996, 307; cf. Railton 2003)

That is, for those who subscribe to the anti-alienation requirement, any theory of well-being should involve an account of prudential goods that have a necessary connection to an agent’s motivations. The debate surrounding
this requirement is lengthy and extensive and I will not attempt to reproduce it here (see, for example, Sarch 2011, Fletcher 2021). What is crucial for us here is that, whether or not anti-alienation is a general requirement for theories of well-being, *Endorsement* is a specific requirement for theories of well-being that aspire to answer the Normative Problem. To be clear, unlike the anti-alienation requirement, the *Endorsement* requirement is not a condition on the nature of prudential goods as such. Rather, it is a more specific condition on the nature of those prudential goods that would be relevant to motivating resistance. This means that even if we arrived at some prudential goods—however they are related—that might be true for all welfare subjects, unless at least one of the items on the list would motivate resistance, realist theories would fail the *Endorsement* requirement.

In response to the anti-alienation requirement, realist may opt for what Guy Fletcher calls a ‘constitutive strategy,’ in which the prudential goods are necessarily constituted by motivational attitudes: “a person who is happy has the affective and/or attitudinal states that are constitutive of happiness, and a person who has friendship has the attitudes of concern and enjoyment that are constitutive of friendship” (Fletcher 2015, 157) The idea here is that realists would be able to provide a motivational connection while also allowing for the possibility that agents might be wrong about their own well-being. However, it is unclear that adopting a parallel strategy in the case of the *Endorsement* requirement would always be sufficient. There are a great many such goods that are already satisfied even under SIMKINS: Simkins herself lists pleasure, happiness, and friendship as among the goods that she takes to contribute to her well-being and that she is able to achieve through a life of household management. And these are in fact compromised in the suffragist resistance against parliamentary exclusion. Suffragist campaigns were not exactly known for being pleasant endeavours for campaigners: Dilke’s critique of the Appeal, for example, notes the “storm of ridicule and contempt” that suffragists received (Dilke 1889, 102); and
Simkins reports being disappointed by “[t]he bitterness, the dissensions, [and] the family feuds” among women that was engendered by the campaigns (Simkins 1908, 77). That is, appealing to these goods (pleasure, happiness, and friendship) are likely to provide Simkins with little to no reason for, or even reasons against, joining the suffragist resistance.

But there are, of course, other goods that are attained in and through resistance: for example, “[c]ontrol over one’s [political] environment” (Nussbaum 2011, 34) and autonomy. These goods in particular have tended to be appealed to by realists attempting to answer the Normative Problem. Indeed, if such goods are taken together with the constitutive strategy, these would constitute the relevant goods that would meet the Endorsement requirement on two accounts: false consciousness would be bad because it compromises the attainment of these motivationally-connected goods, firstly, by perpetuating the lack of these goods in oppression, and, secondly, in masking the reasons for resistance against such oppression. Yet, an approach involving such goods would still fail to satisfy the Continuity requirement in all cases. For, at least in SIMKINS, these goods are not hermeneutically continuous with Simkins’ extant set of valuations.

Consider the case of autonomy, which is the good or capacity par excellence for realist approaches to the Normative Problem (see Silvermint 2013, Enoch 2020). This, importantly, has to be understood in the content-neutral (or procedural) sense (Friedman 2003), which does not determine the content of the acts or ends according to which autonomy would be valuable, rather than a substantive sense, where autonomy is only valuable in relation to predetermined acts or ends, in order not to beg the question against the

114 Autonomy here, notably, would not be treated as a condition on value-determining attitudes but as itself an intrinsic prudential good.
Normative Problem (cf. Stoljar 2000, Oshana 2006, Charles 2010).\textsuperscript{115} Moreover, it is crucial to distinguish between the relevant kinds of content-neutral autonomy here that might be said to be hermeneutically continuous or discontinuous with Simkins’ set of valuations. *Global* autonomy concerns one’s ability to determine one’s own ends across the span of one’s life, while *local* autonomy concerns one’s capacity to respond appropriately to reasons for the execution of individual acts (Oshana 2005). It would indeed be accurate to characterise Simkins’ valuations as continuous with local autonomy, involving the capacity to decide on questions about whether one should have, say, potatoes and molasses for supper—as part of the duties of household management.

However, false consciousness, and here the question of suffrage, pertains to global autonomy, which is in conflict (and hence discontinuous) with Simkins’ belief in the biologically determined ends of womanhood. She particularly emphasises this conflict in her interpretation of suffragism as “an endeavour to impose on all women liabilities” that would interfere with their ‘elementary duties’ as women, which she embraces with pride (Simkins 1909a, 77). In fact, as Suzy Killmister observes, under cases of oppression like those of Simkins’, global and local autonomy are often at odds with each other, such that there is a effective “trade-off” between an agent either having their individual acts frustrated or ensuring their ends align with those determined by oppressive norms (Killmister 2014, 173; cf. Oshana 2006, Killmister 2018). As such, the move cannot also be made for global autonomy to be simply inferred from, or implied by, local autonomy under oppression—*contra* the impression in non-oppression cases where,

\textsuperscript{115} These feminist autonomy theorists (and more) often appeal to a substantive conception of autonomy in order to address the problem of ‘internalised oppression,’ appealing to moral or political norms to inform the substance of autonomy. Whether or not these appeals are a result of a Münchausen over-confidence in moral and political theorising to generate norms in face of the threat of ideology (cf. Curry 2020), our inquiry here is circumscribed solely with the prudential domain.
“presumably, more locally autonomous preferences and choices contribute to a more autonomous life” (Enoch 2020, 180n45).

The realist approach thus on its own does not account for how goods such as (global) autonomy—much less control over one’s political environment—can be drawn from the content of Simkins’ domestic values. Moreover, even if it can for Simkins, the realist is still left with the task of accounting in general for how Continuity might be satisfied for every unique set of beliefs in every case of false consciousness.

3.3. Anti-Realism

The requirements of the Normative Problem might appear like straightforward requirements for an anti-realist account of well-being. But, importantly, what is specifically needed for the Normative Problem is not merely a theory of well-being that can account for motivation to resistance but also account for how norms for resistance are continuous with the criticised’s extant valuations—this, as we will see, is the primary challenge faced by anti-realist theories of well-being. However, for the sake of theoretical thoroughness and at risk of being overly meticulous, I will examine the anti-realist approach here, focusing on the preference-satisfaction theory (how the failure of this approach may extend to those of desire-satisfaction theory should be clear enough).

Naïve preference-satisfaction approaches, where what is good for an agent is just having those things for which they have an actual preference, are straightforwardly inadequate for the Normative Problem. Simkins’ actual preference simply not to have the franchise, as well as her preference for devoting her time and energies into household management over both economic and political participation, would simply mean that neither
Continuity nor Endorsement are met because there is no room for any critical foothold for suffragist critique to appeal to. In response to this, preference-satisfaction theorists often tend to add informational and/or (content-neutral) autonomy requirements (Sumner 1996, Rosati 2006). That is, for a preference to count as the kind that would determine prudential value, respectively, the agent has to have sufficient information about the circumstances relevant to their attitudes, and/or to have had formed the attitudes, as well as endorsed them among others, under some process of critical reflection. According to such non-naive accounts, since the agent’s ‘proper’ preferences remain frustrated under adaptation, false consciousness would be bad for the agent insofar as it contributes to the frustration of these preferences.

These qualifications, however, are insufficient to resolve the Normative Problem in the case of SIMKINS. Simkins actually had sufficient information about the circumstances relevant to her attitudes about suffrage and demonstrated evidence of critical reflection upon them. She was educated at Notting Hill High School (alongside several prominent suffragists such as Barbara Ayrton-Gould; see The Notting Hill High School Magazine 1909), engaged in various forms of literary and historical work, and was publicly involved in the suffrage debates, responding with nuance to suffragist arguments about the nature and role of women in her writings. Simkins’ preferences would thus not be construed as adaptive and would qualify as legitimate preferences (cf. Terlazzo 2016). As such, neither Continuity nor Endorsement would be satisfied, since even with these qualifications there remains no room for a critical foothold, providing no advance on the Normative Problem.

116 Similarly, Raymond Geuss notes that “[t]o speak of an agent’s ‘interests’ [in the concept of false consciousness] is to speak of the way that agent’s particular desires could be rationally integrated into a coherent ‘good life’ [emphasis mine]” (Geuss 1981, 47–48).

117 Simkins was nineteen years old when the Appeal was published, having graduated from high school only a couple of years before.
A way to get around this might be to cast doubt on the legitimacy of Simkins’ actual critical reflection as satisfying the (content-neutral) autonomy condition. After all, false consciousness extends beyond mere preferences to also include an agent’s wider epistemic situation. Had Simkins’ reflection been performed under ideal, non-oppressive circumstances, she would not have arrived at such an endorsement of the preference to domesticity. Again, the debate surrounding the role of idealisation in anti-realism is lengthy and extensive and I will not attempt to reproduce it here (see, for example, Enoch 2005, Sobel 2009, Dorsey 2017). However, what is crucial is that any form of idealisation would still fail to resolve the Normative Problem in SIMKINS—whether understood as a matter of coherence of preferences or full consideration of the ways the preferred good might be (Dorsey 2017). First, Simkins’ preferences would satisfy coherence: she explicitly emphasises the consistency in her preferences against suffrage from when she was twenty years old to when she was forty, and even when she was seventy-nine (Simkins 1908, 1949). Second, her explicit considerations of suffragist arguments about the value of the vote in her publications *Mixed Herbs* (1908) and *An Englishwoman’s Home* (1909), which includes speculative fiction, demonstrated a range of considerations of what the domestic arts might possibly mean. Thus, in both cases, unless we fix in advance what counts as coherence or full consideration by some realist means to include content specific to resistance or political participation (that is, turn to a substantive conception of autonomy; cf. Enoch 2005), it seems that Simkins’ critical reflection, as demonstrated in her writings, would be considered to have been performed under ideal circumstances.

---

118 Dorsey understands idealisation to pertain only to an agent’s values and not to their attitudes—but this make any difference in the case of false consciousness, where values are themselves ideological.

119 This is as far as the records I could gather suggest.
Yet, even if we were to grant under idealisation that she would have arrived at a positive evaluation of political participation, it would still fail to satisfy the Continuity and Endorsement requirements. Idealisation does not alone account for how a pro-attitude towards political independence would be hermeneutically continuous with her actual set of pro-attitudes, which are circumscribed by the domestic context. To stipulate that full information or full consideration must include, say, a revaluation of her concepts of womanhood and domesticity would be to require from Simkins an imaginative capacity towards her preferences that, as Simkins herself and historians have emphasised, were not typical of the British women population at the time (despite the public prominence of suffragist efforts). More worryingly, such a stipulation would be to (substantively) set in advance normative standards that are in dispute to begin with.

Further, since the relevant preferences would be idealised, Endorsement, which requires actual endorsement, is not met either. One might be tempted argue that the demand for actual endorsement is too restrictive, such that resisting oppression has to be part of an agent’s conceived well-being, and that some dialectical path from their actual preferences to resistant preferences must be open to them if the Normative Problem is to have any answer. But, as I will soon go on to suggest in the next section, this demand only seems too restrictive because of a methodologically individualistic framing of motivation. Further, it is difficult to see how such a dialectical path would be open, given that, as we have seen, the Continuity requirement is not satisfied.

In the foregoing parts of this section, I have argued from a consideration of the Normative Problem in Simkins that neither realist nor anti-realist approaches can (at least thus far) provide an answer to the Normative Problem that satisfies Continuity and Endorsement. Realism either fails Endorsement because motivation for resistance is not constitutive of the
prudential goods they enumerate, or because it does not in principle provide an account of Continuity for all instances of false consciousness. On the other hand, anti-realism, even if non-naïve, fails either because the informational and autonomy conditions can already be met under false consciousness, thus providing no critical stance for the Normative Problem, or because it involves idealisations that is inconsistent with the actuality conditions in Continuity and Endorsement.

Admittedly, the above worries for the realist and anti-realist approaches to the Normative Problem, as I have considered above, should not come as a surprise. They have been expressed (though with much less detail and still less sympathy) by those philosophers and theorists that have taken the Normative Problem to be the problem for the concept of false consciousness (cf. Cunningham 1983, MacKinnon 1989, Balkin 1998). In what follows, however, I take the failures of these realist and anti-realist approaches to suggest that turning to substantive theories of well-being has been the wrong strategy to begin with. In their place, I will go on to outline a formal, non-substantive alternative. Of course, I welcome the possibility that a substantive theory (yet to be considered in the literature as it stands) that satisfies the two requirements of the Normative Problem might be found. It would, after all, be consonant with my overall goal to recover the concept of false consciousness for its use in the toolkit of social philosophy. But I hope to also suggest below that such a theory would likely be subsumed under my approach.

4. A Radical Approach to Well-Being

One may accept that false consciousness is not always bad for those under it (a position adopted by Rosa Terlazzo (2016) with regard to adaptive preferences). Perhaps all a theory of well-being can provide—according to
this line of thought—are reasons for those not under false consciousness to work towards the dismantling of the mechanisms that gave rise to the ideology and oppressive circumstances to begin with (Terlazzo 2016). The thought here might be to follow after Marx’s adage that “[t]he weapon of criticism cannot, of course, replace criticism of the weapon, material force must be overthrown by material force” (MECW 3: 182).

However, situating the above Marxian adage within its own context of the Introduction to A Contribution to the Critique of Hegel’s Philosophy of Right (1844) suggests an important place for a critique of false consciousness—one that targets the agent under false consciousness ad hominem—to such a project:

> The weapon of criticism cannot, of course, replace criticism of the weapon, material force must be overthrown by material force; but theory also becomes a material force as soon as it has gripped the masses. Theory is capable of gripping the masses as soon as it demonstrates ad hominem, and it demonstrates ad hominem as soon as it becomes radical. To be radical is to grasp the root of the matter. But, for man, the root is man himself. (MECW 3: 182) [emphasis mine]

As I understand it, there is a Marxian suggestion here that critiques of false consciousness can mobilise mass resistance (which would be especially important for him, if emancipatory hopes are to be found in the proletariat). More than that, reading this in conjunction with Marx’s Comments on James Mill (written also in 1844), we learn that, for him, this possibility is due to radical politics being rooted in a social and dynamic ontology of what it means to be human. That is, Marx specifically understands the good of a human being to be grounded in social interaction:

> Let us suppose that we had carried out production as human beings. Each of us would have in two ways affirmed himself and the other person. 1) In my production I would have objectified my

---

120 Here, Marx is specifically referring to a critique of religion (cf. Ypi 2017).
individuality, its specific character, and therefore enjoyed not only an individual manifestation of my life during the activity, but also when looking at the object I would have the individual pleasure of knowing my personality to be objective, visible to the senses and hence a power beyond all doubt. 2) In your enjoyment or use of my product I would have the direct enjoyment both of being conscious of having satisfied a human need by my work, that is, of having objectified man’s essential nature, and of having thus created an object corresponding to the need of another man’s essential nature. 3) I would have been for you the mediator between you and the species, and therefore would become recognised and felt by you yourself as a completion of your own essential nature and as a necessary part of yourself, and consequently would know myself to be confirmed both in your thought and your love. 4) In the individual expression of my life I would have directly created your expression of your life, and therefore in my individual activity I would have directly confirmed and realised my true nature, my human nature, my communal nature. (MECW 3: 227–228)

As we can see from this passage, “Marx assumes that the good for individuals is complementary: in advancing my good I advance yours and in advancing your good you advance mine” (Rosen 2000, 32). This, for him, contrasts an insular, individualist understanding of what it is like to be a good specimen of one’s kind in bourgeois society (cf. Rosen 2000, Wood 2014). In what follows, I wish to loosely take after Marx’s orientation here and explore how we might be able to overcome the impasse of the above substantive approaches to the Normative Problem through a biodynamical reconsideration of the ontology of welfare subjectivity.121

The aforementioned accounts of well-being grounded welfare subjectivity in an individualist ontology of a human being.122 That is, prudential standards are subject-relative inasmuch as they are grounded in properties instantiated

121 In Adorno and Horkheimer’s words, I seek here to “[remember] nature in the subject” (Adorno and Horkheimer 1947, 40).
122 Michael Bishop’s ‘network theory of well-being’ perhaps comes closest to the biodynamic picture I will outline here. However, attempts to answer the Normative Problem have not availed themselves to his framework and, at the same time, he does not provide an account of the formal normativity of such a network (Bishop 2015).
an individuated biological unit. Consider the following characterisations of well-being or ‘the prudential value of a life’:

A theory of well-being must explain which facts constitute my being better off. So they must be facts about *me*. (Kagan 1992, 171) [emphasis in original]

Welfare assessments concern what we may call the prudential value of a life, namely how well it is going for the individual whose life it is. This relativization of prudential evaluation to the proprietor of the life in question is one of the deepest features of the language of welfare: however valuable something may be in itself, it can promote my well-being only if it is also good for beneficial for *me*. (Sumner 1996, 20) [emphasis in original]

Substantive theories of well-being purport to tell us what ultimately makes something good or bad for an individual and, more broadly, what makes a life go well or poorly for the one who is living it. (Campbell 2015, 402) [emphasis mine]

The relevant kind of evaluation I am making is in terms of well-being, about how well lives go for the person who lives them. (Fletcher 2016, 1) [emphasis in original]

This talk of being a ‘proprietor of the life in question’ is often taken to mean that an approach to well-being assumes that an individual's well-being cannot be at bottom constituted by normative properties instantiated in another individual (see also Haybron 2008, Hall and Tiberius 2015, Fletcher 2021). Admittedly, this allows space for, say, objective list theories to include autonomy in a relational sense, or preference-satisfaction theories to have other-regarding preferences (such as the preference for one's child to succeed). Consider the prudential value of friendship, where the state of a friend is often thought to impact our own well-being. It should be underscored that even realist attempts to justify why friendship is constitutive of well-being tend to rely on particular ways in which friendship is itself indexed to some normative property instantiated in the welfare subject. We have seen, for example, that Fletcher analyses friendship as
being necessarily constituted by “the [individual’s] attitudes of concern and enjoyment that are constitutive of friendship” (Fletcher 2015, 157). What is crucial, however, is that these approaches preclude normative properties instantiated in another biological unit from also constituting one’s welfare-subjectivity, or, conversely, another’s well-being from being constituted by normative properties instantiated in oneself (although these properties may, as in the above examples, nevertheless be directed towards or cause them).

Much of this assumes an underlying model of cognition and human subjectivity wherein an agent’s affects, cognitions, and valuations are ontologically independent of the (social and material) environment in which they are otherwise embedded, or wherein the kind of being that they are an instance of are static and ahistorical. And after all, at first glance, it seems as though false consciousness is precisely the kind of problematic situation in which the properties of another (preferences or exercise of autonomy) is mistakenly believed to constitute one’s well-being. To recall the Millian description of the false consciousness of Victorian women: “women are brought up […] in the belief that their ideal of character is […] submission, and yielding to the control of others,” that “it is their nature” to “live for others […] and to have no life but in their affections […] to the men with whom they are connected, or to the children who constitute an additional and indefeasible tie between them and a man.” (Mill 1869, 132–133). However, appropriating a phrase from Isaiah Berlin, I suggest that instead of a “retreat into the inner citadel” of well-being (Berlin 1958, 182), the Normative Problem can be in fact resolved if we attend more closely to the social dimensions of the biodynamical organisation of the human being within a form of life.

Many philosophers of well-being aim to espouse theories about human well-being that are “intended to be thoroughly naturalistic, representing a view of philosophy as more or less continuous with the sciences” (Haybron 2008,
Yet the individualistic model of cognition and human kinds undergirding extant prudential approaches—especially those pertaining to the Normative Problem—has been extensively challenged in the biological and cognitive sciences over the past decades. This is particularly seen in the work of philosophers and theorists associated with autopoietic theory and enactivism: Humberto Maturana and Francisco Varela—whose ideas have been further extended by Evan Thompson and Wynter. Wynter, in particular, draws directly on the work of Maturana and Varela to provide an ontology (or, following Frantz Fanon, ‘sociogeny’) of the human that goes beyond what she calls “the ‘ethno-class’ or Western bourgeois conception of what it is to be good man and woman of one’s kind” (Wynter 2006, 15). From such a conception, we derive an understanding of prudential evaluation that is relativised only to the ‘proprietor of the life in question.’ Under a Confucian conception of ‘what it is to be a good person of one’s

123 As Edward Baggs and Anthony Chemero observe, enactivist theorists may be characterised as adopting an “epistemic strategy” in explaining “behaviour in terms of animal-environmental relation,” “characterizing the exploratory, self-regulating behavior of the individual organism, and [using] this to understand how that organism brings forth its animal-specific umwelt” (Baggs and Chemero 2018, 1). This is in contrast to the more “ontological strategy” adopted by ecological psychologists, which “[describes] the habitat of the species, and [uses] this to explain how action possibilities are constrained for individual actors” (ibid.). Here, I follow Baggs and Chemero in regarding these two strategies as two sides of the same explanatory coin for animal-environmental relations.
In what follows, I will outline my immanent approach to well-being that follows from autopoietic theory and enactivism. I take after Wynter’s social-theoretic extension of autopoietic theory’s biosocial dynamics for my proposed approach to the Normative Problem, but, in addition, also the critical theorists’ formal turn away from substantive approaches to the normative problem of ideology critique. That is, in light of the failures of analytic substantive theories of well-being, I suggest a shift in framing the prudential approach to the Normative Problem away from these theories by—and this is the critical-theoretic move—reflexively locating the very act of conceptualising well-being within its “matrix in the total activity of society” (Horkheimer 1986, 199).

In the final section of this chapter, then, I will briefly sketch an enactivist (social) ontology of the human (§4.1), before drawing on Jaeggi’s conception of critique as facilitating the problem-solving nature of forms of life to

124 As Richard Kim observes:

On the Confucian view, humans are social creatures, susceptible to a variety of social influences; our individual identities are significantly constituted by an interlocking set of relationships formed within the family and society, and therefore, what is beneficial and harmful to us is importantly constituted by the interests of those that come to inhabit our social domain. [...] In contrast, contemporary philosophers working on well-being have tended to ignore how the nature of human relationships and our social environments are related to human flourishing, because, at best, they are thought only to provide us with knowledge of the necessary empirical conditions for achieving well-being. But if the human self, as the early Confucians believed, is substantially constituted by those relationships which we come to establish during the course of our lives, then it may turn out that any satisfying account of well-being must explain the connections between well-being, family, and community. (ibid., 86) [emphasis mine]

125 What follows may also be seen as a narrower (and perhaps belaboured) articulation of autopoiesis in relation to the human good than the wide-ranging and comprehensive discussions of the autopoiesis of the human in Wynter 1984, Wynter and McKittrick 2015 (cf. Erasmus 2020).
complete the groundwork for any future answers to the Normative Problem (§3.2). I will then return to the case of SIMKINS and consider how the immanent approach to well-being might present us with an advance over the previous approaches (§3.3).

4.1. Enactivism

Instead of approaching cognition and the ontological grounds of well-being by taking human experience and subjectivity as pre-given starting points, enactivism provides a complex, ‘ground-up’ approach to cognition and subjectivity for which I can only provide a sketch here.\(^{126}\) Crucially for our purposes, enactivism involves an understanding that “a cognitive being’s world is not a prespecified, external realm, represented internally by its brain, but a relational domain enacted or brought forth by that being’s [autopoietic] agency and mode of coupling with the environment” (Thompson 2007, 13).\(^{127}\) This requires some unpacking. First, a cognitive being, along with its coupling with the environment, is a *dynamical system*: it

---

\(^{126}\) For more extensive accounts see, e.g., Varela, Thompson, and Rosch 1991, Thompson 2007.

\(^{127}\) To note, there are longstanding arguments within biological and cognitive sciences circles about the precise way to specify the biodynamics or organisational properties that give rise to cognition, and therefore whether ‘autopoiesis’ is the appropriate term to characterise sensorimotor and intersubjective levels of operation. Other terms, for example, include ‘operational closure’ or ‘autonomy’ (specifically understood by enactivists to characterise the biodynamics of all organic life beyond the molecular level), and ‘adaptivity’ (see, for examples, Varela, Thompson, Rosch 1991, Di Paolo, Burhmann, Barandiaran 2017). But the underlying idea here is simply that cognition should be understood as arising out of the dynamics of life and that living systems are autopoietic systems. The precise standard and context of use for the term ‘autopoiesis’ is used inconsistently in various literatures, from the biological and cognitive sciences, to cybernetics, law, sociology, and contemporary Daoist philosophy (see, e.g., Maturana and Varela 1980, von Foerster 1984, Luhmann 1986, Tuebner 1993, Moeller 2006, Thompson 2007). Here, then, I will simply use ‘autopoiesis’ in an inclusive sense to capture whatever biodynamics or organisational properties of living systems turns out to be the basis of cognition. This loose use of the term is fairly common in the aforementioned literatures and is consonant with Wynter’s own social-theoretic appropriation of the term ‘autopoiesis’ in her work. Additionally, this would also help us avoid confusing enactive autonomy with the personal autonomy discussed in the previous sections.
is a “state-dependent [system] whose states are numerical (in the abstract case, these will be numbers, vectors, etc.; in the concrete case, numerically measurable quantities) and whose rule of evolution specifies sequences of such numerical states” (Van Gelder 1995, 368; cf. Thompson 2007). In the case of the human being, these ‘numerically measurable quantities’ would range from biochemical to sociological data (see Lee, Kubzansky, and VanderWeele 2021). Second, foundational to the ‘enacting’ of the relational domain is the notion of an ‘autopoietic’ system which is canonically defined in Maturana and Varela’s 1980 collaboration *Autopoiesis and Cognition* as follows:

An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces components which:

(i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produce them; and

(ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as a network. (Maturana and Varela 1980, 78–79).

Here, ‘machine’ is understood as a dynamical system “defined in terms of its organization” and “hence can be explained in terms of the relations constituting that [particular] organisation”—unlike, say, a computer program that is explained “in terms of the structural components realizing that organization in a particular concrete system” (Thompson 2007, 100).\(^{128}\)

\(^{128}\) With respect to the use of the term ‘machine,’ there is a disagreement between Maturana and Varela’s early collaboration here and Varela’s later works. The former does not see organisms as different in kind from machines, or being distinguished by having any sort of intrinsic teleology. As such, Maturana is often distinguished as an autopoietic theorist but not an enactivist (Villalobos and Ward 2016).

150 of 237
Autopoiesis is not specific to cognitive beings, but is understood to be characteristic of organic life as such and provides the basis of organic cognition (cf. Godfrey-Smith 2016). It would be helpful to illustrate first how it is characteristic of organic life as such by turning to how enactivists understand bacteria to qualify as autopoietic systems and hence organic life, as opposed to an RNA strand or a virus. With respect to Maturana and Varela’s second condition (ii), unlike an RNA strand whose components (guanine, uracil, adenine, and cytosine) do not specify its own system (topological) domain (and thus cannot be said to have its own production network), a bacterium has a molecular membrane that specifies its own domain and within which the network of its processes of production is located. With respect to the first condition (i), unlike a virus whose regeneration takes place in a host cell instead of within its domain (a protein coat), the regeneration and realisation of the network of processes of a bacterium take place within its domain. In other words, unlike a virus, a bacterium has its own metabolic process which is located within, and actively maintains, the system domain over time. Autopoiesis thus characterises the self-regulating and self-constituting organisation of an individual biological system over time.

Since the autopoietic components are not bound to cellular entities or processes, such an organisation may be multiply realisable at higher levels of organic life. In higher primates, autopoietic organisation occurs at various layers of operation, which are inextricably linked to what it means for such organisms to flourish. The metabolic layer, as we saw with bacteria, involves a biochemical network of processes in the self-regulation of the organism. The sensorimotor layer involves a perceptual, affective, and practical network of processes in the dynamic coupling between the organism and its material environment (Umwelt). And, finally, the

---

129 As I mentioned in Note 41, the extension of the term ‘autopoiesis’ beyond molecular levels of operation is controversial but not uncommon.
The *intersubjective* layer involves an interactive network of processes in the dynamic coupling(s) between the organism and its social environment (Thompson and Varela 2001). No layer “is reducible to [or is fully determined by] phenomena in the others,” but each level may be mutually constrained and enabled by the others (Di Paolo, Buhrmann, and Barandiaran 2017, 5; cf. Di Paolo, Cuffari, and De Jaegher 2018).

As enactivists observe, organisms “cast a web of significance on their world” in virtue of autopoietic organisation (De Jaegher and Di Paolo 2007, 488). That is, there are aspects of the world that are significant (and insignificant) for the organism given the organisation of its system components, and the network of these aspects forms an environment in which *sense is made* for the organism. It is in this way that organisms are said to cognise. Returning to the example of bacteria, Thompson provides the following illustration:

> Consider motile bacteria swimming uphill in a food gradient of sugar. These cells tumble about until they hit on an orientation that increases their exposure to sugar, at which point they swim forward, up-gradient, toward the zone of greatest sugar concentration. This behavior happens because the bacteria are able to sense chemically the concentration of sugar in their local environment through molecular receptors in their membranes, and they are able to move forward by rotating their flagella like a propeller. These bacteria are autopoietic and embody a dynamic sensorimotor loop: the way they move (tumbling or swimming forward) depends on what they sense, and what they sense depends on how they move. This sensorimotor loop both expresses and is subordinated to the cell’s [...] maintenance of its autopoiesis. As a result, every sensorimotor interaction and every discriminable feature of the environment embodies or reflects the bacterial perspective. Thus although sucrose is a real and present condition of the physicochemical environment, its status as food is not. That sucrose is a nutrient is not intrinsic to the sucrose molecule, but is a relational feature, linked to the bacterium’s metabolism. Sucrose has significance or value as food, but only in the milieu that the organism itself enacts. Thus, thanks to the organism’s [autopoiesis], its niche has a “surplus of

152 of 237
The layers of autopoietic organisation not only enable but also constrains the significances of an organism’s environment and corresponding behaviour. For example, no sense is made with frequencies above 80 kHz by the sensorimotor system of higher primates, given the kinds of organisation they have that prioritise the visual system, whereas such frequencies make sense for the sensorimotor system of echolocating bats. As such, higher primates do not have the same audio-visual cognitions as bats do and their web of significance is thus specified differently.\footnote{Notably, Wynter herself explicitly relates Fanon’s question of ‘what is it like to be Black’ in sociogeny to Nagel’s question of ‘what is it like to be a bat?’ (Wynter 2001).}

It is at the third, intersubjective layer of operation that a system of sense-making distinctive of what we call ‘the human’ is found.\footnote{While this third, intersubjective level maps onto Wynter’s understanding of the Third Emergence (1997), the First and Second Emergences—the advents of the universe and biological life, respectively—do not map onto the metabolic and sensorimotor cycles. These cycles pertain only to the Second Emergence.} That is, an ontology of ‘the human’ is fully realised in social interaction, where

\begin{quote}
Social interaction is the regulated coupling between at least two [autopoietic] agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent [autopoietic] organization in the domain of relational dynamics, without destroying in the process the [autopoietic] of the agents involved (though the latter’s scope can be augmented or reduced). (De Jaegher and Di Paolo 2007, 493)
\end{quote}

To note, this is not to determine \textit{a priori} that ‘the human’ consists in social interactions—these interactions, after all, may also be found between other cognising organisms or in symbiotic relationships (see Sims 2020). Rather, ‘the human’ is one of the senses that is ‘cast’ by the emergent social interaction, subject to dynamical negotiation among its constituent agents, which manifests in various ways or \textit{genres} that are “relatively well-delimited
starting points that precoordinate the expectations of [agents]" (Di Paolo, Cuffari, and De Jaegher 2018, 178; Wynter 2015 cf. Curry 2017a). When we understand these dynamic couplings to be scaled up to the level of social practices and institutions, we find that ‘the human being’ is thus itself a historical constitution of the various structures of participatory sense-making in which the organism is spatio-temporally embedded, as well as of how these constrain and enable the metabolic and sensorimotor layers of its autopoietic operations within their systems of meaning (see L. Barrett 2017, Heyes 2018).

4.2. Enactive False Consciousness

We can now see how the enactivist picture might be continuous with the social constructionist account of false consciousness that I set out in Chapter II (albeit with a more radical ontological, biological commitment here, extending beyond the social construction of kinds of agents to the human kind itself). As I argued earlier (in Ch II, §3), the interpellation, recognition, and guarantee of an individual under an oppressive schema are what give rise to self-regarding cognitions of false consciousness. Putting both the Althusserian and enactive accounts together: an individual human being has their biodynamic organisation not as a pre-given entity, but in virtue of the interpellations of their agency within the various social schemas that emerge out of the interactions between themselves and other human beings. These various social schemas provide a range of constraints and enablements on the human organism, in its various layers of operation, as well as the ways in which it relates to the larger social system as a constituent agent. Where these schemas are ideological, so would these

This insight has been anticipated in (at least) the Zhuangzi, particularly in the Inner Chapters, which not only deny any a priori sense for ‘the human [人 ren]’ but also understands any a posteriori sense to be a result of dynamic negotiation between organisms within concrete contexts (cf. Perkins 2010, Fraser 2012).
constraints and enablements be oppressive, since the web of significance (the public meaning, scripts, etc.) constituting such schemas would guide one to engage in problematic social practices. It is in this sense that Wynter designates such “culture-specific subjective experiences of what it is like to be human, and therefore of what it is like to be self-evidently conscious, in the terms of each of their culture’s adaptive order of consciousness” as “the phenomenon identified by Marx as that of ideology or ‘false consciousness’” (Wynter 2001, 58–59; cf. Di Paolo, Cuffari, and De Jaegher 2018, Heyes 2018).

To illustrate this, consider the following account of the sociologist Irving Zola’s experiences of airports:

In Zola’s case, his experiences of his body while traveling were shaped by the designs of airports and parking lots, as well as by the social expectation that people should accomplish physical tasks, such as navigating an airport, independently and in as ‘normal’ a fashion as possible. This expectation led him to walk through the airports under his own power, even though walking was tiring for him, rather than using a wheelchair, with (worse) or without (better) help. The feel of his body as sore, tired, and cramped when arriving at his destinations was also constructed by these social structures. Social structures also limited the ways he felt—and so could explain—his personal bodily experiences. Zola suggests that he did not really even notice the original feel of his body while traveling until after a different social structure—the independent living movement for disabled people—made it possible for him to use a wheelchair and experience his body in another way. (Maybee 2019, 74)

The configuration of Zola’s metabolic and sensorimotor layers of autopoietic organisation were constrained by the social structures of airports, yet in a way that was thought of as ‘normal’ for a human being—to the extent that he ‘did not notice the original feel of his body.’ The Independent Living Movement’s renegotiation of the previously ableist genre (on the basis of which the airport environment was organised up until this point), to one that
was less ableist led to a corresponding shift in the configuration of the metabolic and sensorimotor layers and thus also a shift in the operative conception of ‘normal’ human bodily experience for Zola. As Julie Maybee observes, “such ‘structural limitations’ blocked him—a sociologist who was trained to spot socially constructed experiences in other contexts—from noticing or articulating his own bodily experiences as a disabled person for twenty years” (ibid.). Zola’s case thus presents a clear example of, first, how properties instantiated in other individuals (namely, their able-bodied sensorimotor configurations) can constitute the genre within which one makes sense of oneself as a human being, which in turn affects one’s autopoietic organisation. That is, the autopoietic organisation that is taken to define the genre of the human need not have been due to a symmetrical negotiation among the autopoietic agents dynamically coupled together (see also Gill-Peterson 2018, Pearce 2018).\(^{133}\) Second, Zola’s case presents an example of how the agent does not have to be cognisant of this asymmetry in making sense of themselves accordingly—that is, they may be under false consciousness about their own well-being.

But all this only says that individuals under false consciousness cognise themselves within schemas that constrains and enables the various levels of their autopoietic organisation ‘in a problematic way.’ This is not yet to answer the Normative Problem, since ‘in a problematic way’ is still unspecified. It is crucial to note that the constraints on their own are not problematic for one’s well-being and do not thereby solve the Normative Problem for us: for example, having one’s eating patterns (metabolic organisation) constrained as part of being categorised as a patient being treated for bulimia nervosa is not on its own problematic. The constraints and enablements accorded by social schemas simply constitute the ‘social categories we live by,’ including the operative genre of the human (Ásta

\(^{133}\) Moreover, self-regarding cognitions can extend to whether or not one is even cognised as a human being (Curry 2017a).
In the above consideration of Zola, it is only retrospectively determined by him that ‘the original feel of his body’ was problematic after the change in genre brought about by the Independent Living Movement. What concerns us particularly here with regard to the Normative Problem, however, is the task of justifying such a presupposition prior to the changes in genre (such that someone under false consciousness has prudential reasons to change and participate in such a change): to determine how genre-constraints can be problematic for an individual’s well-being, especially since welfare subjectivity is meant to be grounded what it means to be a human being in the first place.

4.3. The Unity Norm

I suggest that the necessary first step towards articulating the formal grounds for any prudential answer to the Normative Problem is to observe that autopoiesis does in fact provide a very general, enactive form of normativity. As Ezequiel Di Paolo notes, autopoietic “self production is a process that defines a unity and a norm: to keep the unity going and distinct [emphasis added]” (Di Paolo 2005, 434). Any approach to well-being, under the enactive framework, would have to be formally grounded in—but not substantively determined by—the continued unification and individuation of a human being’s self-constitution at metabolic and sensorimotor layers, as well as their intersubjective constitution within the social systems they are embedded in. Just as in Thompson’s example of the bacteria, particular components that bring about the continuation and individuation of an autopoietic system—according to the various layers—could therefore gain substantive significance for the individual human being (within the particular genre in which they are understood) as their prudential goods. Further, as the very acts of conceptualising well-being are themselves acts of participatory sense-making, certain negotiated conceptions of prudential
goods can acquire a stabilising function for a given genre: in that “what really regulates human societies is the subject of each order wanting to be a good man or woman of your kind,” however construed (Wynter 2006, 15). I call this the **Unity Norm**.

Since the Unity Norm only specifies a very general condition under which something may be said to contribute to a welfare subject’s biodynamic organisation qua living organism, the enactive approach to well-being remains non-substantive. It could, for example, be compatible with a variety of genres (such as the Western bourgeois or Confucian genres), as well as the variety of prudential goods these might be thought to entail (such as preference-satisfaction or family). But this compatibility with a wide range of genres, as long as the organism’s unity is maintained, is precisely why the Unity Norm alone still does not address the Normative Problem.134 The multiple ways in which these autopoietic systems of significance may be realised, may enable and constrain, and thus may manifest ‘the human,’ threaten a prudential relativism that unhelpfully returns us to the failures of the naïve anti-realist approach. That is, it may well be the case that prudential value is relative to whichever genre of the human is salient, but with regard to whichever prudential value is operative, neither **Continuity** nor **Endorsement** can be met because there would no room for any critical foothold for critique to appeal to. With only the Unity Norm, what would be good or bad for Zola or Simkins was simply determined by the ableist genre or Victorian genre within which they understood themselves. This is not necessarily problematic for enactivism as a descriptive theory in the

---

134 As Nathaniel Barrett has observed, attending solely to the level of biodynamic articulations, the various systems of meaning, enactive normativity on its own involves “nothing more than prescriptive norms” as a corollary of the Unity Norm (N. Barrett 2017, 438). That is, as Varela puts it, “any action undertaken by the system is permitted as long as it does not violate the constraint of having to maintain the integrity of the system and/or its lineage” (Varela, Thompson, and Rosch 1991, 205). For an example of how the Unity Norm might be in operation even in cases of anorexia nervosa—which tends towards the self-destruction of the biological unit—from an active inference perspective, see Barca and Pezzulo 2020.
biological and cognitive sciences, but if enactivism is to be critical, and avoid an unhelpful relativistic approach to the Normative Problem, we need a further norm apart from the Unity Norm.

As Nathaniel Barrett argues, any further enactivist norm has to meet two criteria. First, such a norm has to be “a plausibly emergent feature of a self-organized system that can serve as the basis for non-representational self-regulation” (N. Barrett 2017, 441). That is, the norm has to remain a general feature of autopoietic organisation and not a product of a particular sense-making process (a fortiori, any genre), so as to characterise any organism. Second, the norm “needs to be open-ended enough to allow for the formal identity of the system (the ‘self’) to change in adaptive ways” (ibid.). That is, the norm cannot specify in advance what the particular (range of) configuration(s) of the system should be, i.e., it should remain non-substantive.

4.4. The Adaptation Norm

To briefly recap, in the previous subsections (§§4.1–4.3), I provided an articulation of a descriptive enactivist approach to well-being. But in order for such an approach to become critical and address the Normative Problem, we require a further enactivist norm that is not yet found in the enactive literature. In what follows, I suggest that a further norm meeting these two conditions can indeed be found. I do so by turning to the critical theorists’ insight that critique targets second-order pathologies—specifically drawing on Jaeggi to provide a formal solution to the Normative Problem.

Recall that the task of critical theory as being the ‘self-clarification to be gained by the present time of its struggles and desires.’ Jaeggi places emphasis on the self-clarification of struggles, understanding this to mean
analysing forms of life to involve “systemic blockages or disruptions with regard to the perception and solution of problems and correspondingly are the result of failed or deficient transformation processes” (Jaeggi 2019, 216). Importantly, she understands critique to involve not the mere articulation of how a system fails to resolve a problem, but how it fails to do so as a result of its prior adaptations to prior problems. A central example she gives of this is Marx’s critique of capitalism. Briefly, bourgeois civil society was critically analysed by Marx as an adaptation of the European system of production from the mode of feudalism, with the former constituted by the norms of civil liberty and equality that were adaptations to the systemic crises that were actualised in the conflicts between peasants and the aristocracy in the latter. However, these norms of civil liberty and equality, that are constitutive of the capitalist mode of production, are seen to be in contradiction with their very expressions in the system itself:

The capitalistic contract of employment is concluded between (formally, i.e., legally) free and equal parties. After all, these are independent contract parties confronting each other, and, in a certain sense, they interact as free and equal: the employee is not a bondslave and there are no feudal law status restrictions. On the other hand, however, the reality of capitalist work relations seems to contradict these norms. Most employees, are in fact, forced to enter into such relations (on pain of starvation); furthermore, the existing material inequality developing between the parties of the contract is, according to Marx’s analysis, not accidental but systematically induced. (Jaeggi 2009, 67)

It is in this sense that capitalist ideology can be critically analysed by attending to how, at moments of crises, it expresses the maladaptations of the capitalist mode of production to previous crises. While Jaeggi’s approach pertains explicitly to ideology critique, it can also be adapted to the critique of false consciousness insofar as the underlying concern is an articulation of how certain failures to perceive and respond to problems are due to prior adaptations of a dynamic system to prior problems.
Put in enactivist terms, a second, general norm can be derived from the Unity Norm. I call it the Adaptation Norm: *adaptations of the system should not block or disrupt future adaptations*. That is, when an adaption is actually required at some future time, a prior adaption is *maladaptive* if it blocks or disrupts this. Such a blockage or disruption thus occasions a systemic conflict (that is, a crisis), which can only be determined as a *systemic* problem in *post facto* analyses for a dynamic system. For example, the dodo’s increase of fat stores and defenceless breeding behaviour, given the climate of Mauritius and relative absence of predators, only proved fatally maladaptive when it became a prime target for Dutch sailors to hunt and its eggs vulnerable to the invasive species the sailors introduced. The extent of the systemic conflict and the degree of further adaptation required to overcome it (even maladaptively) depends both on the complexity of the dynamic system and the extent of the change(s) to which adaptation is required.

To note, the notion of maladaptation that I employ here is distinct from Alvaro Moreno and Matteo Mossio’s characterisation of a *malfunction* within an organisational system. The latter is “an unresponsive trait does not modulate [the system’s] activity as required by the intervention of regulatory functions and therefore prevents adaptive regulation to shift to a different first-order organisational regime, so that the whole system can only remain in a specific organisational regime in which the trait match the functional presuppositions” (Moreno and Mossio 2015, 84). An example of malfunction that Moreno and Mossio give is a case of angina, where the regulatory, autonomous nervous system, in a fight-or-flight situation, attempts to shift the cardiovascular regime from ‘at rest’ to ‘under stress,’ but the coronary artery fails to dilate proportionally to an increase in blood flow pumped by the heart. In such cases of malfunction, the “degree of malfunction of a trait could be assessed in terms of the set of first-order organisations of which it
prevents the realisation,” such that “[t]he degree of malfunction is, therefore, inversely proportional to the degree of adaptivity of the organism” (ibid., 85). Maladaptive systems, however, can be extremely adaptive and yet remain problematic, as we see in the case of capitalism. Or, put another way, it is the regulatory function itself that is malfunctioning.

The Adaptation Norm thus meets the two criteria for enactive normativity set out by Barrett, inasmuch as it is a plausibly self-emergent feature that can serve as the basis for non-representative self-regulation, as well as precisely insuring open-ended adaptation of the system. Given that successful adaptations already presuppose an absence of blockage, this norm can only be retrospectively salient at a moment where its violation is actualised as a conflict between a system’s components: at which point the Adaptation Norm would guide theoretical attention into analysing the system’s earlier adaptations as self-frustrating—to diagnose a pathology of the system, as it were.

4.5. False Consciousness as Maladaptation

With the Adaptation Norm, we can now articulate the autopoietic grounds for any adequate prudential answer to the Normative Problem. Contra the ‘top-down’ approach from ideology critique, false consciousness is not problematic because an agent’s set of beliefs simply inherits the normative failures of the social system(s) in which the agent is embedded. Rather, false consciousness is problematic because it violates the Adaptation Norm.

To be more specific, false consciousness is problematic for an oppressed individual inasmuch as it is maladaptive for them: the individual thereby participates within genres and social categories—and hence acquire corollary constraints and enablements—that threaten the unity of their
particular autopoietic organisations. As result of the maladaptation of their beliefs, components of the welfare subject’s metabolic, sensorimotor, and intersubjective processes may at some point come into conflict within themselves or with each other and threaten the autopoietic organisation’s unity and are blocked from straightforward adaptation. It is in this sense that individuals may be negatively affected by false consciousness in different ways while within the same genres, depending on the further social categories according to which they are constructed within it.

We can see, for example, how, while a Western bourgeois genre may generally satisfy the Adaptation Norm in the autopoietic organisation of (at least) white middle-class men in the US, it would more clearly be differentially maladaptive for the autopoietic organisations of Black women in the US or bakla in the Philippines. With respect to the former, as Wynter puts it, the interpellation of Black women by the same genre is expressed as a class-based social ladder: “white women as the ‘absolute’ thing; the ‘J-Lo,’ the sort of ‘exotic concubine’ type; and then Black women at the bottom of this thing.” (Wynter 2006, 25). With respect to the latter, the bakla’s interpellation by the gender binarism corollary to the Western bourgeois genre is expressed as a ‘homosexualisation’ that predominates the effeminacy and gender-transitivity that used to be their primacy characteristic. As J. Neil Garcia puts it: “[a]s if coping with his swishy ways in a helplessly macho culture were not enough,” which was brought on by Spanish colonisation, “the bakla must now also contend with private demons of self-loathing on account of his intrinsically ‘pathological’ or ‘sick’ desire” (Garcia 2008, 168). These maladaptations then, would be made salient in occasions of crises for such individuals.

Put generally, under false consciousness, an agent’s autopoietic organisation is compromised because pernicious belief-forming mechanisms had maladapted their beliefs. These mechanisms, as we saw, could be
found at the structural level (such as the culture industry) or the individual level (such as testimonial injustice). Further, required changes to the system-regulating attitudes (including valuations) of the agent at an occasion of crisis are, as a result of the maladaptive beliefs, blocked: by being themselves obscured from the agent as requiring adaptation. And such attitudes may even extend to the very beliefs about what it is to be a human being and hence what well-being consists in.135

To return to Simkins, consider how the Adaptation Norm allow us to determine the problematic nature of Simkins' false consciousness, particularly with respect to how she recounted her experience of unemployment. Simkins interprets the disagreeableness she experienced from a person who turned her away at a potential workplace as rooted in “the ancient, the eternal rivalry between two women” (Simkins 1908, 49). While an understanding of rivalry between women was neither uncommon in fin-de-siècle Britain nor seen as generally problematic (see Marcus 2007), this rivalry, for Simkins, was problematic in the context of work allocations because, according to her, “[w]omen in authority are so apt to grow dictatorial, autocratic, narrow,” especially with their “freshness, individuality, and delicacy lost under that deadly transforming ‘grooviness’ of masculine careers” when they were forced to leave their domestic ambit (Simkins 1908, 49).

135 Consider, by way of comparison, Margaret Urban Walker's account of the ideological distortions of the oppressors through “necessary identities,” in her critique of Bernard Williams’ account of how the ancient Greeks saw the difference between slavery and womanhood:

I don’t disagree with Williams's claim (taken as a very rough generalization) that the Greeks saw being a woman as a necessary identity, and being a slave as not always or so clearly one. Nor am I suggesting that the conditions of slaves and women, or the Greeks’ perceptions of them, were the same. My point is that the evidence for the involuntary, vulnerable, coercive nature of women’s situation seems quite comparable, in the forms both of known treatment and of testimony, to that for slaves. The difference would not seem to be in the recognizability of an unchosen and unenviable situation of which any rational person could complain, but rather in its being recognized, in particular by those in a position to enforce that situation for others while remaining protected from it themselves. (Walker 2007, 174)
Simkins 1909a, 135). As her own writings suggest, Simkins is herself able to observe how her unemployment and disagreeable experience was expressive of conflicts within her autopoietic organisation as a woman in the British labour market (albeit in her own terms): she makes a passing remark, for instance, that “[w]e have yet much to learn about the best manner of adjusting work to women and women to their work”; and she argues even for the setting up of national job registries (taking after the Germans) as a means to remedy the situation (Simkins 1909a, 129). Further, she boasts that her arguments for such “a national system of labour exchanges,” were found among those employed by The Times as a recommendation to address “the serious crisis of unemployment from which the right to vote and to legislate has not been able to save the British man to-day” (Simkins 1908, 157–158).

Despite these insights, Simkins is unable to see how the problem with her continued unemployment was not so much due to the industrial corruption of femininity, as it were, but an inefficient labour market. Nor could she see how women such as herself could have contributed to parliamentary politics, if given the vote, and thus contribute to addressing the inefficiency. The maintenance of her false consciousness in face of her continued unemployment, as we (as social philosophers) can see, were due to her maladaptive beliefs in, and valuations of, the domestic boundaries of women. And while Simkins is able to continue to satisfy the Unity Norm by attributing her continued unemployment to women being forced into the workforce at the expense of their (and her) appropriate domestic activity, this is yet another maladaptive response to the conflicts in her autopoietic organisation: a conflict between the sensorimotor configurations (her seeking to earn a wage and yet inability to find work) brought about by the

136 Simkins’ commitment to the idea that “the most womanly woman is just the housekeeping woman” lasted even after the two world wars, but her later writings nevertheless suggest that some conceptual concessions were made, speaking, for example, of such a genre as “an old order very nearly extinct” (Simkins 1908, 135; Simkins 1949, 83).
intersubjective layer which involves the social category of ‘(domestic) woman’ and that of ‘worker.’—the very conflict Dilke points out in her reply to the Appeal. What could satisfy the Adaptation Norm for Simkins, then, would be for her to revise her category of ‘woman’ in accordance with the suffragists’.

That said, the task of a critique of false consciousness (enacting the critique as opposed to merely theorising about it) would only begin when such autopoietic conflicts occur. Under the immanent approach to the Normative Problem, what would be required from the critic is to articulate, in accordance with the agent’s system of meanings (including their valuations), to them how these conflicts of autopoietic organisation are pathological to their system of meaning: that these conflicts are due to the self-frustrating significances the agent acquired through their interpellation as such-and-such a subject by given social schemas. In this way, the critic would be able to facilitate at least psychological resistance by bringing to the fore the relevant maladaptive interpellations to the attention of the agent. Notice, however, that the relevant prudential norm being appealed to in such a critique is not any enactive norm (whether the Unity or Adaptation Norm) but the very prudential norm (or norms) that is being frustrated in the agent’s attempt to realise it within their given autopoietic organisation. In Simkins’ case, for example, this is the norms of femininity within the context of the social categories salient for her as a working woman. A crucial point to note in this immanent approach then is that, in order for critique to concretely satisfy the Continuity and Endorsement requirements of the Normative Problem (and thus bring about resistance in the criticised), the critic themselves should be understood as attempting to form a dynamic coupling with the agent being criticised in order to present to the agent ‘the self-clarification of their struggles and wishes.’ As Simkins wryly remarks:

Forty years ago the philosopher, John Stuart Mill, published his work, the *Subjection of Women*, but women in general are said
not to have paid it much attention. In ordinary life, indeed, we are not accustomed to see women regulating their conduct by the advice of the philosophers. [...] Whether his doctrines had any practical connection with life and conduct she would never trouble to argue about. (Simkins 1909a, 47)

That is, given that the task consists in an articulation of relations between social significances and a particular agent’s autopoietic organisation, a solution to the Normative Problem cannot be determined \textit{a priori}, as it were, by the social philosopher or theorist of well-being, but must await empirical research and be understood as itself a mode of participatory sense-making \textit{between critic and criticised} (cf. Celikates 2018). The approach is thus \textit{immanent}. Theorising about the Normative Problem on its own does not engage in such concrete sense-making, and neither \textit{Continuity} nor \textit{Endorsement} requirements can be satisfied by simply laying out this present account of the immanent approach to well-being to someone under false consciousness.

From the above, we can thus see how, under the immanent approach, \textit{Continuity} might be satisfied: while the Adaptation Norm is formally presupposed in the background, critique acquires substantive content \textit{a posteriori} by drawing attention to the need to reconceptualise the agent’s extant valuations and their corollary genres and social categories, whatever they may be, to address the present crisis (with an eye to prevent similar future ones). With respect to the \textit{Endorsement} requirement, the immanent approach satisfies it since the substantive content of critique pertains to valuations which the agent already holds, that constrain and enable their affective (and hence motivational) system at the sensorimotor layer.

Consider how, given the enactivist picture, this approach compares with the realist and anti-realist approaches to \textit{Continuity} and \textit{Endorsement}. While the anti-realist approach would also draw on the agent’s extant valuations, its attempt to secure a critical stance via informational and autonomy
conditions reaches at most at the same impasse as the first enactive norm: how a system may best maintain itself given its existing arrangement internally and in relation to its environment, without taking into account how these attitudes themselves may be problematic for an agent. And, keeping in mind how Mill’s arguments were deemed by Simkins to lack ‘practical connection with life,’ the realist approach may at most sometimes pick out prudential goods that that are continuous with an agent’s extant evaluative concepts (as it did for the suffragists). Thus, they fail Contiguity. Alongside idealised anti-realism, realist approaches lack an account of how these values are to be introduced, at the intersubjective level, to the actual system of an agent under false consciousness in a way that would lead to sufficient uptake at the sensorimotor level to effect at least psychological resistance—thus failing Endorsement.

5. Summary

I have argued in the foregoing sections that the Normative Problem of false consciousness, given its Continuity and Endorsement requirements, cannot be resolved by substantive approaches to theories of well-being in spite of their extant usages (Geuss 1981, Enoch 2020). An implication of the discussion in this chapter is thus a dilemma for social philosophers who have thus far used the concept of false consciousness: either the substantive theories of well-being they invoke have no fundamental role in social criticism or they have to reject the possibility of a prudential critique of false consciousness. In the case of the latter, they would have to spurn much of the Critical Theory tradition and resort to moral and political critiques that cannot be ultimately grounded in claims about well-being.

To overcome this dilemma, I have suggested in the previous section (§4) a formal account of how the problem might be resolved in the context of
social research, by combining enactivism with insights from Critical Theory. Specifically, I drew upon Wynter’s socio-theoretic extension of autopoiesis and Jaeggi’s understanding of critique as targeting maladaptations of dynamic systems. In this way, I sought to explore a possible avenue for empirically-informed, prudential critiques of false consciousness. It may well be the case that a substantive theory of well-being emerges in the dynamic coupling between critic and criticised, but it should not mistakenly seen as grounding the normativity of critique so much as a particular expression of critique within a genre.
IV.  
Does False Consciousness Necessarily Preclude Moral Blameworthiness?  
The Refusal of the Women Anti-Suffragists

CHAPTER ABSTRACT

There is a worry that appealing to false consciousness in questions of responsibility for the harm suffered by members of oppressed groups is victim-blaming. Individuals under false consciousness allegedly systematically fail the relevant rationality and epistemic conditions due to structural distortions of reasoning or knowledge practices, thus undermining their status as responsible moral agents. But attending to the constitutive mechanisms and heterogeneity of false consciousness allows us to see how having it does not *eo ipso* render someone an inappropriate target of blame. In this chapter, I focus on the Appeal, arguing that its signatories, despite false consciousness, satisfy both conditions for ordinary blameworthiness. I consider the writings and activities of three prominent signatories, Louise Creighton, Beatrice Webb, and Mary Ward, observing that the irrationality characterisation is unsustainable beyond group-level diagnoses, and that their capacity to respond appropriately to reasons was not in fact compromised. Following recent work on epistemic injustice, I also argue that culpable mechanisms constituted their false consciousness, rendering them blameworthy for the Appeal.

In the previous chapter, I considered the Normative Problem, which was understood in terms of how social criticism ran up against extant theoretical assumptions about well-being in ethics. In this chapter, I turn to the Agential Problem, which concerns how social criticism runs up against our extant
assumptions about moral responsibility. This amounts to the general intuitive worry that appealing to the notion of false consciousness in accounting for the harm suffered by members of oppressed groups amounts to victim-blaming by conceptual fiat. I focus here on moral responsibility—rather than social, political, prudential, or epistemic—because this has been the sole focus of a large part of detractors who raise this worry and perhaps the most extensively theorised form of responsibility thus far within the literature. Some of my considerations may not extend to other kinds of responsibility, but I hope at least the worries about the concept of false consciousness in the moral context will be allayed.

In Chapter II, we saw how the concept of false consciousness might be best understood as a regulative concept that designates a wide variety of possible looping mechanisms that sustains, and is sustained by, epistemically defective beliefs. Here, I will argue that attending to the constitutive mechanisms and heterogeneity of false consciousness allows us to better see how suffering from it does not eo ipso render an agent an inappropriate target of blame—that is, how one’s status as a morally responsible agent is not necessarily undermined by a characterisation of false consciousness.

It is important to observe that, in addition to the victim-blaming intuition, one may also have the conflicting intuition that it is precisely because of false consciousness, and its attendant epistemic sophistication and lack of coercion, that one can legitimately lay part of the blame on such agents. There is a sense that the concept brings into focus crucial but otherwise overlooked agents that are partly responsible for contributing to anti-suffragism. And, as recent historians have noted, narratives that the women

---

137 For some considerations of these other kinds of responsibility, see, for examples, Medina 2013 for epistemic responsibility, Matthews 2015 for prudential, and Jugov and Ypi 2019 for political.
of the Appeal were simply uncritical handmaidens in the anti-suffrage cause of men crucially overlooks the unique and thoughtful contribution of the signatories in strategising and succeeding in “inflecting anti-suffragism with new, more positive emphases designed especially to appeal to female public opinion” in both parliamentary and public debates (Bush 2007, 157; see also Nelson 2004, Delap 2005).

This intuition should not come as a surprise given the arguments of the foregoing chapters. The concept of false consciousness, after all, provides a constitutive story about how they became a particular type of agent such that we would justifiably attribute the relevant act to them as theirs. False consciousness would thus allow us to see the women anti-suffragists as having the same status of agency as their male counterparts (who were themselves also largely under the grips of patriarchal ideology), as equally full-fledged members of the moral community. Exempting them as moral agents in limine would not only be unjustly presumptive (the anti-suffragists themselves only expressed opposition to political equality), but also be counter-productive to the emancipatory cause. Ordinarily, our blaming practices are thought to have a crucial communicative function, addressing agents and appealing to shared reasons for modifying their attitudes or behaviour (see, for example, Calhoun 1989, Houston 1992, Moody-Adams 1994, Mason 2019). In contrast, by theorising such agents outside the moral community, we preclude ourselves from enlisting them in collective resistance against a shared oppression and to bring about structural change (cf. Jugov and Ypi 2019). Furthermore, if individuals can be blamed for contributing to their own oppression despite being under false consciousness, then they can a fortiori be blamed for contributing to the oppression of others despite being under false consciousness (for example, the women anti-suffragists’ imperialism or ableism).

---

138 These worries have also plagued the debate about the usefulness of the concept of adaptive preferences (cf. Terlazzo 2016).
So while the intuition of exemption from blameworthiness has been often presumed, the task of this chapter is to cautiously vindicate this second intuition and allay the worry of victim-blaming. To do this, I take up the women of the Appeal as a historical case study to show how they, even under false consciousness, can satisfy both the rationality and epistemic conditions—and thereby qualify for blameworthiness in the ordinary sense. I adopt a broadly Strawsonian approach here, following moral philosophers such as Gideon Rosen and Elinor Mason in using ‘blameworthiness in the ordinary sense’ or ‘ordinary blameworthiness’ to refer to agents being liable to certain blaming responses within a given moral community, having failed to act according to its standards (see Mason 2019). My main contention here is that, too often, the broad group-level strokes used to paint those under false consciousness as irrational, along with the attention on whether a whole group of people can be relevantly aware under false consciousness, pay insufficient attention to the variety of ways in which the relevant ignorance at the individual level might be nevertheless actively acquired and maintained by the individual agent as part of their false consciousness.

In what follows, I will first suggest how there might be conceptual space for the culpability intuition (§1). I will then attend specifically to the case of the Appeal, observing how its signatories, despite false consciousness, satisfy both rationality and epistemic conditions for ordinary blameworthiness (§2). I focus on the details of three prominent women of the Appeal—Louise Creighton, Beatrice Webb (then Potter), and Mary Ward—arguing that the irrationality characterisation is unsustainable beyond the group-level diagnosis, and that the women of the Appeal's capacity to respond appropriately to reasons was demonstrably not compromised despite them being correctly understood as having false consciousness. Then, drawing on

139 I exclude the quality of will condition here (see, e.g., Talbert 2013, Arpaly 2015) as the women of the Appeal, as we will see, were clearly well-motivated, despite being morally mistaken.
recent work on epistemic injustice and attending to the women’s own writings, I argue that there were indeed culpable mechanisms at play in the case of their false consciousnesses—for example, testimonial injustice and wilful hermeneutic ignorance—rendering them blameworthy for the Appeal. Lastly, I return to what the case study means for the relationship between the concept of false consciousness and blameworthiness in general, particularly in terms of how such careful historical attention helps refine and delimit our use of the concept as social philosophers (§3).

1. Conceptual Space for Culpability

As I have argued (in Ch. II), false consciousness involves stabilising mechanisms that occur systematically at both individual and structural levels, resulting in a looping effect in an individual’s self-understanding given their position in the relevant structure. This looping effect involves more than the classification and the classified people, but also feedback loops among institutions, knowledge, and experts. To recall, looping mechanisms may occur at both individual and structural levels:

**Individual**: Nash equilibria, stereotype threat/boost, implicit bias, wishful thinking, confidence-shaking, speech act accommodation, epistemic laziness, testimonial injustice, meaning vertigo etc.

**Structural**: market equilibration, culture industry, consumer production, hermeneutical injustice, epistemic bubbles, echo chambers, recommender systems etc.

Given this analytical framework, we are able to account for how the diverging intuitions of exculpation and culpability arise by considering the diverse mechanisms and different aspects of the heterogeneity of false consciousness: an oppressed individual’s agency and reasons is
conceptualised as both *constrained* and *constituted* within a structure of oppression. When the individual-level mechanisms salient for the constitution are culpable, we would be able to regard the agent as blameworthy for the issuing action. This does not mean, however, that false consciousness is, or should be, reduced to such constituent mechanisms. After all, the mechanisms alone do not guarantee that the beliefs would be ignorant to the extent that they systematically motivate acts of oppression. Explanations appealing to false consciousness are, to underscore, ‘structural explanations’—that is, they help us understand “the behavior of the individual *given* their place in a structure” and also understand “the individual as the instance of a type—a type defined by the conditions for existing at that node” (Haslanger 2016, 129).

So the *ethical* task here of determining whether a specific agent is nonetheless morally responsible would have to involve individual-level analysis of the form of false consciousness, attending to the constitutive stabilising mechanisms that are salient for the agent, given their position within the relevant structure(s)—and attend the structural-level only insofar as it might restrict or mask rationality, or render ignorance non-culpable.

2. **The Appeal against Female Suffrage**

Before I get to my main argument in this chapter, it is worth summarising the argument of the Appeal. The Appeal’s rejection of suffragism was grounded in the belief that “the limits fixed by the physical constitution of women, and by the fundamental difference which must always exist between their main occupations and those of men” made it such that “the necessary and normal experience of women [… ] does not and can never provide them with such materials for sound judgment as are open to men,” with respect to

---

140 To be clear, by ‘individual-level analysis,’ I do not mean a methodologically individualist analysis, wherein the structural context is occluded (see Haslanger 2016 and Ch. II).
parliamentary concerns (“An Appeal,” 782). Adhering to these limits, as they repeatedly underscored, was also crucial for the security and success of “colonial affairs” and “colonial policy” (ibid.). Effectively, the Appeal argued that women already influence politics in an indirect way that would ensure that their different interests and capacities are accorded due weightage in parliamentary decision-making: suffrage was thus unnecessary and counter-productive. What undergirded this was the self-understanding of women of the Appeal as being naturally different in a way that excluded them from the political sphere. According to the Appeal, women’s suffrage was both unnecessary and antithetical to their own proper domain of activity, in that direct participation would lead to epistemic and ethical weightlifting (cf. Williams 1995). That is, while men were political animals, women were domestic ones.

From this, it is clear how it would be appropriate for us to apply the concept of false consciousness to the women of the Appeal: their belief that women were naturally deficient in political judgment is patently false (epistemic ‘falsity’), arising out of the social conservatism of the time (genetic ‘falsity’), and it motivated them to undersign a manifesto that buttressed the anti-suffragist movement (functional ‘falsity’). But despite satisfying the analytic conditions of the concept of false consciousness, I will now go on to show that historical attention to the women of the Appeal’s position in fin-de-siècle Britain reveals that the stabilising mechanisms involved do not undermine their status as morally responsible agents. That is, they would still be blameworthy insofar as the mechanisms do not undermine the rationality and epistemic conditions.

2.1. The Rationality Condition

As I mentioned at the start of this chapter, the concept of false consciousness has often been taken to imply an exempting irrationality on
the part of the affected agent. Carol Hay, for example, argues that oppression may damage one’s rational capacities permanently and thus disqualify them as a blameworthy agent, citing false consciousness (she calls it ‘self-deception’) as a “classic” case of such resultant irrationality, wherein “oppressive social systems create incentives for oppressed people to believe certain falsehoods about themselves, contrary to their own evidence” (Hay 2013, 123–124; cf. Superson 1993). It would thus be tempting to think that the women of the Appeal, under conditions of patriarchy, suffered from some sort of collective motivated irrationality, inasmuch as this suggested an inability to respond to reasons to the extent that they would compromise their own interests. After all, the Appeal’s success in maintaining the status quo of the parliamentary exclusion of women was not only in its novel, women-friendly packaging of anti-suffragism, but, as suffragist Ray Strachey solemnly recorded in 1928, it “presented [a new argument] to the [anti-suffragist] side. ‘Women themselves don’t want the vote,’ they could now say; unfortunately it was partly true” (Strachey 1928, 285). And as some suffragists wryly noted, the anti-suffragists’ efforts ran counter to not only to the growing need to represent women’s interests in legislation but also the very anti-participatory sentiments that the anti-suffragists professed to uphold (see, for example, Fawcett 1889).

But this ‘top-down’ approach is mereologically suspect: it considers the social-level properties of the women of the Appeal as being instantiated at the individual-level simply in virtue of them being a part of the group. While the claim of systematic irrationality might be maintained at the level of the signatories as a collective, it is harder to maintain at the individual level, when we attend to the various mechanisms that formed and maintained the belief in the natural incapacity of women for political activity.
To reemphasise, despite the relative dearth of historical attention on them compared to suffragists, anti-suffragism was neither a fringe nor homogeneous stance among women at the time. While suffragists such as Millicent Fawcett and Margaret Dilke swiftly and acerbically responded in the July issues of the periodicals *Nineteenth Century* and *Fortnightly Review*, with hundreds undersigned, thousands more signatures came in the August issue of the *Nineteenth Century* in support of the Appeal (Sutherland 1990, Bush 2007). The political equality of women was far from clear as a settled moral fact for the zeitgeist and there were internal disagreements even among those who later organised themselves as the Women’s National Anti-Suffrage League (1908–1910). Further, much anti-suffragists’ sentiments was left unexpressed in the public sphere, not only partly as a “logical extension of their reluctance to take to the stage of parliamentary politics” but also because of “their positive commitment to a paradigm of womanhood characterized by altruistic femininity, devotion to family duties, and inconspicuous public service in the extended domestic setting of local communities” (Bush 2007, 3; see also Heilmann and Sanders 2006, Crozier-De Rosa 2018).

As with all forms of false consciousness, the content of the ignorant beliefs of the women of the Appeal were noticeably bound to the domain of its domestic social structures (where what ‘domestic’ consisted in was only slowly being extended by the influx of women into the labor market). These women were perfectly capable of responding to relevant reasons in other aspects of social life and were well-respected in doing so: Creighton was the co-founder and first president of the bipartisan National Union of Women Workers (NUWW), which coordinated women volunteers throughout Britain; Webb was a leading sociologist of the cooperative movement and a co-founder of the London School of Economics; and Ward was an accomplished novelist who was a prominent advocate for the education of the poor as a means for ‘equalisation’ in society.
In trying to avoid this ‘top-down’ mereological problem, one might perhaps opt to follow Harriet Taylor and John Stuart Mill in adopting a ‘bottom-up’ mechanical approach to the question of irrationality, guided by a notion of false consciousness (see, for example, Cudd 2006). In “Enfranchisement of Women” (1851), the central question for Taylor and Mill was “why the existence of one-half the species should be merely ancillary to that of the other—why each woman should be a mere appendage to a man, allowed to have no interests of her own, that there may be nothing to compete in her mind with his interests and his pleasure” (Taylor and Mill 1851, 12). They observed that women have been made to “consider as their appropriate virtues” their (at minimum) political dependence on men (ibid., 13).

But, returning to the distinction between global and local autonomy I made in the previous chapter (Ch. III, §3.2), while it might well have been the case that global autonomy was compromised in the declaration of the Appeal (i.e. the women’s ability to determine their own ends across the span of their lives), it is less clear that local autonomy was at all compromised for the women of the Appeal (i.e. their capacity to respond appropriately to reasons for the execution of individual acts) since questions about whether one should have, again, potatoes and molasses for supper would not fall under the purview of the relevant form of false consciousness here (McKenna 2005, Oshana 2005). And although some might argue that one’s capacity to respond appropriately to reasons may be suitably compromised under oppression, these arguments have largely been concerned with cases where preference-adaptation occurs (see, for example, Hay 2013, Cudd 2015). While preference-adaptation might have been a salient mechanism for some anti-suffragists, it is unclear how the mechanism of preference adaptation would be very salient for the women of the Appeal, given that they were upper- and middle-class women (none of them were working-class). Webb, for example, herself admits that “had [she] been a man, self-respect, family
pressure and the public opinion of [her] class would have pushed [her] into a money-making profession; as a mere woman, [she] could carve out a career of disinterested research” (Webb 1929, 355). While these women might have been politically excluded, they were by no means frustrated to the extent that their responsiveness to reasons would be compromised. Moreover, they nevertheless did have overlapping social concerns with suffragists as seen in their involvements with the NUWW (and also in their friendships across partisan lines).141

Further, it is hard to maintain that the women’s local autonomy were so compromised to warrant exculpation, given what we know about the degree of sophistication and strategising that went into publishing the manifesto. The Appeal was primarily a result of the efforts of co-instigators Ward and Creighton. Creighton herself had expressed, and was compelled by, fears that the increased frequency of franchise attempts in the House of Commons and a sympathetic Prime Minister (William Gladstone) “[made] it seem likely that Parliament might suddenly pass a bill granting female suffrage” (Louise Creighton, letter to Ida Koch, May 23, 1889). It would not come as a surprise that the editor of the Nineteenth Century, James Knowles, was a known anti-suffragist and saw Ward and Creighton’s enthusiasm as a welcome means for his periodical to intervene in the ongoing suffragism debates at such a critical juncture (Bush 2007). But the production and publication of the manifesto was an affair of joint decision-making rather than one of Knowles’ sole orchestration. Amongst themselves, moreover, Ward and Creighton also self-consciously attempted to distance themselves from the “semi-religious beliefs on the natural and necessary position of women” held by their male counterparts, which they did not “altogether share” (Mary Ward to Louise Creighton, April 18, 1889).

141 Further, given the Millian account of patriarchal Victorian upbringing, it seems odd to speak of adjusting preferences if they were something one was brought up with.
From these considerations, then, we can safely conclude that the capacity to respond to relevant reasons were neither restricted nor masked in any absolute or exemptive way. After all, false consciousness cannot be exemptive *simpliciter*, if we are not willing to also exempt oppressors who suffer it by the same token—there is surely an absurdity in excluding a good portion of the population *in limine* as morally responsible agents. So while irrationality might actually be the right diagnosis for other forms of false consciousness elsewhere (for example, those involving preference-adaptation), the false consciousness of the women of the Appeal would be better characterised as a matter of having bad premises. That is, the more fruitful approach is not whether they had the capacity to adequately weigh up what was right or wrong in their act of undersigning, but whether they had the awareness of what was right and wrong to begin with.

2.2. The Epistemic Condition

It is not uncommon for agents affected by false consciousness to be regarded as exempt from blameworthiness on account of failing to satisfy the epistemic condition. Ann Cudd, for example, thinks that false consciousness renders an agent non-culpably ignorant, noting that “the oppressed may well not understand the oppression they suffer, for it is often a part of their oppression that it is hidden from them under the guises of tradition or divine command or the natural order of things” (Cudd 2006, 198). It is thus also tempting to compare the women of the Appeal with the usual examples found in discussions on culpable ignorance in the moral responsibility literature, such as ancient slaveholders or sexist US fathers in the 1950s (see, for example, Benson 2001, G. Rosen 2003, Guerrero 2007, Sher 2017). Much of the debate surrounding these examples concern whether the circumstantial epistemic bad luck that these agents find themselves in is sufficient to exculpate, given that, for example, “it would
have taken a moral genius to see through to the wrongness of chattel slavery” (G. Rosen 2003, 66; cf. Fricker 2007, 98–107). But there are at least two disanalogies here, at individual and structural levels of analyses.

At the individual level, the usual comparisons are almost always cases wherein the agents under analysis are clearly members of the oppressor group—for example, the “run-of-the-mill” First World, cishet, white, middle-class man in Gideon Rosen’s classic case (Rosen 2003, 66)—and not the oppressed: they would not be victims but paradigmatic beneficiaries of epistemic injustice. Perhaps supporting weaker intuitions to exculpate in these cases, the relevant forms of false consciousnesses and corollary stabilising mechanisms in these would be substantially different from those cases pertaining to members of oppressed groups (for example, stereotype boost vs stereotype threat).

At the structural level, the applicability of the concept of false consciousness to the women anti-suffragists itself distinguishes this case from these comparisons: where a lack of epistemic friction is often thought to characterise cases such as ancient slaveholders, evidence-resistance in the case of the anti-suffragists presupposes encounter with dissenting views. In this case, it was clear that there was ample counter-evidence for the women of the Appeal: longstanding suffragist arguments (the centenary of Wollstonecraft’s *A Vindication of the Rights of Women* was just around the corner) and recent shifts in the labor market.\(^\text{142}\) Importantly, the latter meant being faced with the conflict between the social consciousnesses arising out

\(^{142}\) Many suffragists, however, did not necessarily share the so-called ‘liberal feminism’ of Wollstonecraft, which argued for the cause of women from individual rights rather than social role (‘social feminism’). Social feminists thus could also hold that ‘women are domestic creatures,’ but that they still needed the vote in order to perform their roles—which often included expanded participation in the civilising or educating missions outside of the imperial metropole. To note, the passing references to ‘feminism’ here is meant only in the broad sense of the promotion of equal consideration of women’s interests (for more on the relation between feminism and suffragism, see the methodological note to this thesis in Ch. I, §4.2).
of membership in Victorian society as a politically dependent person and
that out of membership in the labor market as an economically independent
person—both of which schematised over the same resources. As Dilke
replied to the Appeal,

The supporters of woman suffrage do not believe in indirect
representation under any circumstances, but least of all when
the influx of women into the labour market brings them,
whether they will it or no, into competition with those whose
interests and capacities are different; it is not the Woman
Suffrage Societies that have brought about this great social
change. A man is no longer expected, even in well-to-do
middle-class society, to support his adult sisters and daughters
as well as his wife and infant children. The societies, accepting
the new state of things, wish to protect the earning of these
women, to teach them self-reliance, to help them in the only
way human beings can be efficiently helped—shown how to
help themselves.

[...] The rapidly increasing wealth of the middle classes
has deprived thousands of women of the necessity for
household toil; but education and increased opportunities for
intellectual and public work draw these same women, if not in
the first, then in the second generation, into busy useful lives,
giving satisfaction to themselves and benefit to the community
at large. (Dilke 1889, 98–100)

We see concrete demonstrations of such epistemic friction giving rise to
emancipatory consciousness in the public defections of Webb and
Creighton to the suffragist cause, as well as in their vocal support for it,
between the Appeal’s 1889 publication and the formation of the Women’s
National Anti-Suffrage League in 1908. In 1906, Creighton announced to the
NUWW that she realised that, given the increasing role women were playing
in party politics, they needed to have “the responsibility of the vote”
(Creighton 1994, 89, 146). A few days after, Webb wrote a public letter to

143 Creighton’s own memoirs seems to suggest that this refers to a conference in 1912, but
Fawcett’s letter to the editor of The Times in 1906 mentions Creighton’s change in her views
in October 1906, which also would be consistent with her Women in Council obituary
(Millicent Fawcett, letter to the editor of The Times, 5 November 1906; “Louise Creighton
Obituary,” Women in Council Newsletter, May 1936; cf. Creighton 1994, 172n30; Covert
2000, 368n32; Bush 2007, 162n61).
Fawcett explaining that she now also realised that there was a need “to permit this growing consciousness among women—that their particular social obligations compel them to claim a share in the conduct of political affairs” by “finding a constitutional channel” (Beatrice Webb, letter to Millicent Fawcett, November 2, 1906). On the contrary, the suffragist responses only served to reinforce Ward’s anti-suffragist position, and she intensified her efforts against suffragism in the following years.

The question concerning us, therefore, is whether there were stabilising mechanisms specific to the women of the Appeal—which were disrupted for Creighton and Webb, or reinforced for Ward—that would have made them culpable for their ignorance, such that judgments that they were blameworthy for undersigning the Appeal are justified. And from the suffragists’ indignant replies, we can gather an affirmative answer: testimonial injustice and wilful hermeneutical ignorance. That is, caused by the prejudices surrounding women and lower classes, the women anti-suffragists not only accorded a deflated level of credibility to the suffragists, but also refused to adopt the alternate epistemic resources offered by them. Dilke, for example, laments precisely the neglect of suffragist claims in her reply to the Appeal:

> Those who have spoken and written repeatedly on this subject for the last dozen years have a feeling of hesitation and shyness at being obliged to use the same arguments again and again, and to bring but little fresh fuel to feed the furnace of public opinion; but it is only necessary to read through the appeal with care to find that the opponents of further progress have simply burnished up the old weapons and sharpened the time-worn steel.144

> […] These ladies take upon themselves to say the time has come to arrest all further progress; ignoring the fact that as the old bonds and fetters fall away from women’s limbs new requirements arise, new possibilities open out before them,

---

144 To be fair, while Dilke was right to note that the underlying arguments had not changed despite the new rhetoric, she missed out on Strachey’s aforementioned observation of the new (though problematic) argument that women themselves did not desire the vote.
and careers that but a short quarter of a century ago would have seemed far out of their reach now open before them and seem to call able and well-educated women to fill posts for which their training has fitted them.

[...] Ladies of intellect and social standing can always make their voices heard, can always write to the papers and magazines, can command the sympathy and attention of public men whenever they feel they receive less than justice. But the supporters of woman suffrage aspire to help those other women whose lives are spent in humble toil, whose work is ill paid, whose education has been defective or entirely neglected. (Dilke 1889, 97–98)

Fawcett was also keen to point out the exclusion, alongside the consequences of the continual lack of enfranchisement of women:

A further consideration of the Nineteenth Century list of names shows that it contains a very large preponderance of ladies to whom the lines of life have fallen in pleasant places. There are very few among them of the women who have had to face the battle of life alone, to earn their living by daily hard work. Women of this class generally feel the injustice of their want of representation. The weight of taxation falls upon them just as if they were men, and they do not see why representation should not go with taxation in their case, simply because their physical strength is less than that of men. (Fawcett 1889)

The ladies in The Nineteenth Century support their case by stating that ‘all the principal injustices of the law towards women have been amended by means of the existing constitutional machinery.’ They may not know that the law still recognises in a mother no legal rights over her children during the lifetime of her husband. [...] The inequality of the divorce law is another well-known instance of the cases in which the existing constitutional machinery has remained placidly content with a state of things unjust to women. The inequalities of the law of intestacy, as regards men and women, are so flagrant as to be almost ludicrous. Existing constitutional machinery has arranged that in almost every case of intestacy the male relatives got the lion’s share.

[...] A similar kind of inequality is maintained as regards probate. [...] It is unnecessary to point out that to the whole professional class this necessity of paying probate on what in
many cases is the widow’s own property, comes at a time when she is impoverished by the death of the chief breadwinner of the family. No such harassing and exacting demands are made upon a man who loses his wife. The assumption of the law is that all their joint property is his only, and he pays no probate on plate, furniture, &c., which they may have worked for and bought together. (“Women’s Suffrage: A Reply”)¹⁴⁵

Fawcett and Dilke’s replies expressed feelings of resentment and indignation at how the women of the Appeal seemed to have not only unjustly misrecognised the political status of women but also contributed to normalising the socio-economic injustices that were being passed over by the systematic underrepresentation of women’s concerns in legislation.¹⁴⁶ But in enacting testimonial injustice, the women of the Appeal not only harmed women by prolonging parliamentary exclusion of their interests, but also their capacity as knowers (Fricker 2007).¹⁴⁷ That is, women were excluded in limine as full participants of the epistemic community, where default inclusion is meant to be constitutive of being an epistemic agent. Further, as often happens in cases of testimonial injustice, the epistemic harm is compounded in diminished self-confidence in one’s beliefs or own epistemic capacities: Strachey observed a lamentable “phase of discouragement” among suffragists in the aftermath of the Appeal, leading eventually to the militancy of the Suffragettes (Strachey 1928, 285).

¹⁴⁵ Although the publication’s author was unspecified, this is commonly attributed to Fawcett by historians given her writing style (Nelson 2004, Bush 2007).
¹⁴⁶ There might be a concern that the salient brand of oppression here would actually be class rather than gender, such that our concern here with the women of the Appeal should not longer be a case of the false consciousness of the oppressed but rather that of members of the oppressing group. But as we observed earlier, given that social structures overlap, while gender oppression is foregrounded in the general notion of suffragism, there is a need to appreciate the overlapping nature of systems of oppression and the heterogeneity of the stabilising mechanisms of false consciousness. Moreover, the undersigned of the Appeal were, as Fawcett observed above, at least partly composed of women who would be directly affected by the success or failure of the suffragism.
¹⁴⁷ That is, whether the credibility deficit is understood non-distributively or otherwise (see Medina 2013, Coady 2017).
There might be a natural concern that the testimonial injustice was itself a product of other exculpatory stabilising mechanisms constituting false consciousness (a further maladaptive response, as it were). Anti-suffragist sentiments, after all, were widespread at the time and there was definitely some measure of hermeneutical injustice. That is, there was some lacunae in the collective conceptual resource: many women were unable to make sense of their own social experience (for example, the concept of an independent woman). But, at best, this period of overall moral uncertainty involved a situation of a “midway” rather than “maximal” hermeneutical injustice (Fricker 2016b, 167). That is, rather than a complete absence of the relevant conceptual resource, suffragist groups involved “localised or in-group hermeneutical practices that [were] nonetheless not shared across further social space” (ibid., 166). But why these practices did not extend beyond suffragist social spaces into those of the women of the Appeal despite their best efforts, then, might be said to be a case of contributory injustice—wherein “an epistemic agent’s willful hermeneutical ignorance [i.e. a willful refusal to acknowledge epistemic tools developed by marginalised groups] in maintaining and utilizing structurally prejudiced hermeneutical resources thwarts a knower’s ability to contribute to shared epistemic resources within a given epistemic community by compromising her epistemic agency [emphasis mine]” (Dotson 2012, 32.; cf. Pohlhaus, Jr. 2012).

A vice-based approach is often adopted in approaching cases of epistemic oppression in social philosophy (Fricker, 2007, Medina 2013). So just as the instantiation of the (moral-epistemic) vices of arrogance and laziness may render someone’s cultural ignorance culpable, and on that basis they would be blameworthy for an action motivated by culturally ignorant belief, the fact that the women of the Appeal instantiated the vices of testimonial injustice and willful hermeneutical ignorance should suffice to render their ignorance culpable (cf. Mason and Wilson 2017). But, whether approaching this case in
vice-epistemological terms or otherwise, we can conclude from the above considerations that the women of the Appeal were culpably ignorant on account of testimonial injustice and wilful hermeneutical ignorance, and accordingly blameworthy for the publication of the Appeal that hindered the suffrage movement.

Nevertheless, there might still be lurking worries that the testimonial injustice of the women of the Appeal was a result of a prior state of false consciousness that was non-culpably constituted. Simply having a midway rather than maximal case of hermeneutical injustice, for one, does not guarantee that wilful hermeneutical ignorance was the salient difference-making mechanism for any of the women of the Appeal. Still, if all we seem to require is better historical evidence to determine which were the salient mechanisms for the women of the Appeal, this would not be problematic for the general position set out here with regard to the Agential Problem, since my claim is simply that the concept of false consciousness does not necessarily rule out blameworthiness. And, in fact, attending to Creighton and Webb’s explicit reflections about their aforementioned defection, we find that a number of culpable stabilising mechanisms sustained their beliefs—none of which they regarded as justificatory and all of which they felt remorse for. Ward, unfortunately, was sadly unmoved even in spite of losing the anti-suffrage cause—but with the relevant culpable stabilising mechanisms no less discernible.

Creighton’s 1906 public announcement of her change in stance was no doubt due at least to her involvement with the NUWW from her presidency in 1895, as a result of which she traveled broadly and “gained a wide knowledge of all kinds of women’s questions,” coming to eventually feel that “women made great progress, both in speaking and in business capacity” (Creighton 1994, 116).148 Admitting in her later memoirs that instigating the

148 This was briefly observed in the previous chapter (Ch. II, §2.2).
Appeal “was a mistake on [her] part,” and that she had “always hated everything that was concerned with political parties,” Creighton noted that “[w]hat was most decisive in leading [her] to change [her] opinion some years afterwards” was her own observation recognition that women needed the vote for their growing participation in “party politics” (Creighton 1994, 89). It was also claimed that “her sympathy with the industrial women in the north had led her to see the need of the vote to protect their interests as wage earners” (Covert 2000, 302).

Yet, Fawcett’s July 1889 reply (to which Creighton’s public rejoinder in August also addressed) had already pointed out the need for suffrage to protect working-class women, in addition to pointing out the contradiction of the Appeal in holding that women could not form ‘sound judgments’ on matters of ‘constitutional change’ and yet nonetheless organised and worked to elect male party candidates (for example, the Primrose League). Creighton did not obviously seem to pay much attention to these points at the time of the Appeal and recorded that she only realised these for herself after. In this, then, we find that Creighton’s testimonial injustice was a result of a wilful hermeneutical ignorance on her part. Rather than heeding Fawcett’s reply at the time and during her early tenure as the president of the NUWW, Creighton had refused to acknowledge the legitimacy of the suffragist hermeneutical resources as they were set forth at the time. In Creighton’s rejoinder, she had insistently maintained, against women’s participation in parliamentary decision-making, that

If the vote was the privilege of the wise and the educated, many women might justly claim it. But it is the propelling power of a part of the machinery of government which has always belonged to one sex. (Creighton 1889, 351)

149 She also notes that she thought the Suffragettes “needed to be steadied by responsibility” given their “wild performances” (Creighton 1994, 146).
Admittedly, there is some suggestion of the non-culpable mechanism of stereotype threat in her memoirs. Creighton noted that she was afraid that her recantation would take place too soon after her husband’s passing in 1901, such that she would seem to have been against suffrage because of her husband’s influence—even though no one had said this publicly and even though Mandell Creighton himself was not opposed to suffrage and had “left [her] absolutely free to form and express [her] own opinions” (Creighton 1994, 89). But the mechanism of stereotype threat here was more relevant for the *announcement* of her opinions rather than the maintenance of her beliefs about suffrage *per se*.

Unlike Creighton, Webb notes in her semi-autobiographical volumes that, although she “delayed [her] public recantation for nearly twenty years,” she “immediately and resolutely withdrew from that particular controversy” upon Fawcett’s “indignant retort” to the Appeal (Webb 1929, 354; cf. 1975, 361). An entry in her manuscript diary, dated only a few days after Fawcett’s reply, records her refusal of anti-suffragists’ requests of her to respond to Fawcett’s ‘retort’ in the *Nineteenth Century* (given Webb’s social and intellectual standing):

> At present I am anxious to keep out of the controversy. I have as yet accomplished no work which gives me a right to speak as a representative of the class Mrs. Fawcett would enfranchise: celibate women. And to be frank, I am not sure of my ground; I am not certain whether the strong prejudice I have against political life and political methods does not influence my judgement on the question of enfranchising women. (Webb 1929, 354)

Referring to this ‘strong prejudice,’ Webb later confessed that her “false step” arising out of her being “anti-feminist in feeling [was] easy to explain, though impossible to justify”:

> Conservative by temperament, and anti-democratic through social environment, I had reacted against the narrow outlook
and exasperated tone of some of the pioneers of women’s suffrage, with their continuous clamour for the Rights of Women. Also, my dislike of the current parliamentary politics of the Tory and Whig ‘ins’ and ‘outs’ seemed a sort of argument against the immersion of women in this atmosphere. But the root of my anti-feminism lay in the fact that I had never myself suffered the disabilities assumed to arise from my sex. (Webb 1948, 361) [emphasis mine]

This appears almost verbatim in Webb’s earlier volume My Apprenticeship (1929), with a removed note describing her anti-feminist feeling as a part-reaction to “[her] father’s overvaluation of women relatively to men,” and another note about “a luncheon given by an American lady to American suffragists (who had not given [Webb] a cigarette to sooth [Webb’s] distaste for the perpetual reiteration of the rights of women) venting this irritation by declaring provocatively—‘I have never met a man, however inferior, whom I do not consider to be my superior!’” (Webb 1929, 355).

Thus, unlike Creighton, Fawcett’s ‘indignant retort’ was what made Webb hesitate about the claim that the ‘principal injustices of the law’ towards working-class women were indeed ‘amended by means of the existing constitutional machinery,’ as she had undersigned in the Appeal. Although it took some time for her “old prejudice” to eventually “evaporate” altogether, it is noteworthy that it was precisely the suffragists’ reactive attitude of indignation (often thought to be constitutive of moral blame) against the signatories that conduced Webb’s eventual epistemic erosion (Webb 1948, 361).150 Webb’s testimonial injustice against the women of ‘exasperated tone’ at the time of the Appeal, then, was admittedly not a matter of wilful hermeneutical ignorance. However, she herself confessed later that it had been failure of her own procedural epistemic duties as an “impartial investigator of women’s questions,” due to the intellectual arrogance she mentioned (ibid., 360).

150 For those who think resentment or indignation are essential to blame, see, e.g., Strawson 1962, G. Rosen 2015.
Unfortunately, unlike Creighton and Webb, Ward passed away less than a year after she planned to write about her own reflections on her participation in the anti-suffrage campaigns (Ward 1919). But when we attend to her response to suffragist replies over the years after the Appeal, we see not only a repetition of the ‘same’ arguments she had set forth with Creighton in the Appeal ‘again and again,’ but an intensifying thirty-year effort to withhold the extension of suffrage to women—leading eventually to her deflation and defeatism when it was finally won. In a 1910 polemic for The Times, Ward went as far as to describe women as “necessarily ignorant” (Ward 1910).

And in her final plea to the House of Lords in January 1918, before the vote for the Representation of the People Act 1918, Ward clearly heeded none of the arguments that had won over her other contemporaries: Ward continued to insist that, “because of the conditions of their sex,” women were “the less educated, and more excitable, the less skilled, and less responsible elements in [Britain’s] population” and indeed that, with the War, “[t]he physical force argument—that physical force is the ultimate sanction of the Parliamentary vote—stands stronger to-day than ever” (Ward 1918, 48–49, 55). She even reached towards bold, conspiratorial reasoning towards the end of her plea: “I do not doubt its sincerity on the lips of many good men for a moment. But the real motive power behind the clauses, so far as the House of Commons, and political parties are concerned, has been simply political calculation” (Ward 1918, 52). Presuming that she spoke on behalf of more women than the suffragists, Ward continued to deny not only that women were capable of political participation, but that they desired suffrage—in spite of organisations such as the Primrose League or the Suffragettes. After the vote, she wrote defiantly to Creighton:

Well now, thank goodness it is over [...] Now the question is what the women will do with their vote. I can only hope that
you and Mrs Fawcett are right and I am wrong. (Mary Ward, letter to Louise Creighton, 14 March 1918)

What is crucial in interpreting these efforts by Ward is that in her acclaimed novels, she nevertheless very explicitly and penetratingly explores various suffragist positions through her characters: Marcella (1894), which she began writing three years after the Appeal, ends with a balance between the ‘liberty’ and ‘equality’ of woman and her social duties; in Delia Blanchflower (1914), the eponymous heroine oscillates convincingly between militant and non-militant forms of suffragism, even prompting an American reviewer unfamiliar with Ward’s anti-suffragism to mistake her for a suffragist (Sutton-Ramspeck 2000). Given that, in spite of these aesthetic sympathies, Ward was still politically unsympathetic to the suffrage cause, it would thus not be unsafe to conclude that both epistemic arrogance and wilful hermeneutical ignorance were the salient mechanisms lying behind the testimonial injustice that maintained her false consciousness—if not before, then at least definitely in the years after the Appeal.

3. False Consciousness and Blameworthiness

We have seen from the case study in the preceding section that false consciousness does not necessarily exempt one from moral blameworthiness: where culpable mechanisms (for example, testimonial injustice) are salient, it may even render one blameworthy. Given the heterogeneity of false consciousness’ mechanisms, neither the rationality nor epistemic conditions are necessarily compromised, as the initial intuitive

151 Ward’s initially close friendship with Creighton was now strained at this point. In November 1912, she was removed from the NUWW and in 1917, writes to Creighton, “I thought of telling you something of what your old friends have gone through in the last four years and how changed our life has been. But on the whole it seemed better not—Silence is best” (Mary Ward, letter to Louise Creighton, 12 September 1917). To note, she publicly maintained that her “view about the Suffrage has divided [her] in opinion, though not in feeling and affection, from many friends with whom I have work in social or educational questions” (Ward 1918, 59).
worry set it out to be. An agent’s responsiveness to reasons under false consciousness is largely limited to the particular ideology’s domain, and may be uncompromised if it does not involve mechanisms such as preference-adaptation. An agent may also be culpably ignorant under false consciousness, such as if the relevant evidence-resistant attitudes are sustained by (moral-epistemic) vices where non-culpable mechanisms are not salient in the acquisition of those vices. The use of the concept of false consciousness, therefore, is not as morally problematic as some detractors have suspected—although it should nevertheless be invoked judiciously.

Further, as in the case of Webb, we have also seen that ordinary moral blame, if understood at least in terms of appropriate reactive attitudes (for example, resentment, indignation) or in terms of addressing the agent and appealing to shared reasons for modifying their attitudes or behaviour, may be itself instrumental to the epistemic erosion of a given false consciousness. In cases where false consciousness does not thematise over the reasons to which blame would appeal, the act presents to the relevant agent an instance of epistemic friction. This, as Medina pointed out and we have seen in Webb’s awareness of her intellectual prejudices, provides an opportunity for developing meta-lucidity. But, to emphasise, this is only a possibility. For, in other cases, perhaps closer to Creighton’s (whose own rejoinder brought her national fame as an anti-suffragist), ordinary blame might seem to only have a backfire effect.

It might be objected that, in the examples given in the above case study, Creighton and Webb did not suffer false consciousness—at most Ward did, as evidenced from her persistent beliefs and behaviour. However, such an objection involves unnecessarily strict requirements for what qualifies as false consciousness that is not justified by the role of the concept in the critical project, as set out in this thesis. The regulatory concept is only meant to guide research into the contributory elements of structures of oppression.
and point towards potential sites of amelioration. Both Creighton and Webb fall squarely under the definition of false consciousness: having evidence-resistant, ignorant beliefs (epistemic ‘falsity’), which motivated them to reinforce a particular structure of oppression (functional ‘falsity’), and in virtue of which they formed and maintained those very beliefs (genetic ‘falsity’). To deny that false consciousness was present, if it were eroded, seems to be a matter of conceptual confusion and being overly pessimistic about the possibility of emancipatory consciousness forming among the oppressed. Moreover, the erosion of false consciousness for both Creighton and Webb, involving a fairly comprehensive reformation of their social self-understanding, took years before the hold that the ideological mechanisms had on them was loosened and also took great emotional tolls, such as in straining close friendships (as in the case between Creighton and Ward).

Still, the above objection might be coming from a more fundamental worry that, if social philosophers and theorists simply obtain sufficient evidence for the mechanisms behind an agent’s psychological makeup and development, the concept of false consciousness itself would thereby be made explanatorily redundant for our blaming practices (apart from functioning as a heuristic for our evidence-gathering). Further, we seem to be moving away from a structural approach into a merely methodologically individualist analysis. Under the individualist analysis, it would seem that the concept of false consciousness loses explanatory traction since, as a structural explanation, it characterised an individual as an instance of a type existing at a structural node. What we would need, according to this line of thought, is not structural analyses—although they may be useful heuristically—but simply better epistemic warrant concerning the individual. Another related worry might be that wilful hermeneutical ignorance cannot be singled out from the nexus of the myriad constitutive individual-level and structural-level mechanisms of false consciousness as the mechanism which specifically resulted in the testimonial injustice of the women of the Appeal. So even if
there were culpable mechanisms, the general worry is that they work systematically alongside non-culpable ones in a way that renders them inextricable, insufficient conditions on their own.

But these worries are misplaced and conflates a number of concerns. First, mechanisms such as testimonial injustice, wilful hermeneutical ignorance, or stereotype threat, are structural notions: structural marginalisation is constitutive of them. Second, we need to distinguish between two intersecting, but very different projects here: one that concerns regulating the social research of groups and another that concerns how we, as agents within a moral community, are to respond to fellow agents within it. In the former, we find concerns for the systematicity of mechanisms, culpable and non-culpable, working together as jointly sufficient conditions for ideology. And, as we pointed out earlier, this systematicity is only found when we analyse structures and their ideologies from the perspective of the group. Not all of these mechanisms are relevant in the latter project of our blaming practices: we are only concerned with those mechanisms that were salient for the individual agent in question, such that they would be blamed or exempted accordingly. That preference-adaptation might have been a salient mechanism for one individual under a given case of false consciousness, does not mean that it was so for another. Webb, for example, in addition to noting that she did not suffer the ‘disabilities assumed to arise from her sex,’ pointed out that she was in a better socio-economic position than men as a “writer on economic questions” (Webb 1929, 355). To maintain otherwise would be to return to the aforementioned mereological problem, confusing characterisations of entire groups with individuals.

My approach to the question of whether the women anti-suffragists were blameworthy differs from a number of other approaches that have been put forward recently to the question of moral blame in similar cases of systematic social ignorance: forward-oriented responsibility and extended
blame. These, it would seem, are promising ways to approach the question of responsibility in cases of false consciousness.

References to Iris Marion Young’s “social connection model” of responsibility has been gaining a fair bit of traction recently among some social philosophers as a way to avoid the trappings of attributing fault to an individual for structural effects (Young 2011). That is, such responsibility is “politically assigned rather than discovered” and “potentially grounded in factors other than moral desert or causing harm” (Zheng 2018, 4; see also Fricker 2007, Medina 2013, Jugov and Ypi 2019). But while this might be a fruitful way forward in terms of the ameliorative task of the critical project, this also notably avoids the notion of blameworthiness entirely, and its attendant moral sentiments in the ordinary cases.

Another notion of blame also gaining traction is grounded in expectations for agents to take responsibility for their actions, independent of whether they committed a fault (see, for example, Fricker 2016a, Bagnoli 2018, Aragon and Jaggar 2018, Calhoun 2019, Knowles 2021). Elinor Mason calls this ‘extended blame,’ inasmuch as we “voluntarily extend our responsibility zone in order to secure the respect and trust of others, and as a way of showing commitment and investment in our relationships” (Mason 2019, 185). It would no doubt seem reasonable to expect the anti-suffragists to take responsibility for having undersigned the Appeal. For some, there might be a question as to how much this would be distinct from the former, forward-oriented notion of responsibility (for example, Fricker 2016a), but in either case it does not involve the full range of moral sentiments relevant to ordinary blameworthiness.
4. Summary

I have argued from a careful consideration of the women of the Appeal that an agent may, in fact, be ordinarily blameworthy under false consciousness. Victim-blaming is thus not intrinsic to the concept of false consciousness. However, to conclude, it should be underscored that despite this, it would be not only intellectually presumptuous but morally pernicious to move straightforwardly from this to the claim that we should blame them or that any act of blaming in similar cases is fully justified. An act of blame does not merely involve the blameworthiness of its recipient but also other crucial aspects such as the aim of blame or one’s standing to enact it. The ethics of blame requires that we take into account the justificatory conditions of these other aspects: failing to do so might risk redoubling any harm that might have already satisfied the deterrent aim of blame or even reinforcing oppressive institutional norms, if blame is performed publicly (cf. Snow 1994, Matthews 2014, Engen 2020). Until we take all of these into account, we cannot fully assess the justifiability of blaming even in cases like the blameworthy women of the Appeal.

In the following, concluding chapter (Ch. V), I will briefly examine one such way in which the aim of blame is frustrated when occurring at the same time as a critique of false consciousness. This perhaps may be seen as an attempt to make some headway into answering these broader questions.

---

152 See Coates 2020 for a survey of these other aspects of the ethics of blame in general.
Conclusion: A Methodological Limitation
Moral Criticism and Social Criticism

Chapter Abstract

This brief, concluding chapter summarises the arguments and advances of the previous chapters, before raising a methodological concern about a tension between a critique of false consciousness—wherein individuals are criticised in terms of social categories—and moral blame—wherein agents are criticised as individuals. This concern is extrinsic to the concept of false consciousness, but nevertheless presents an issue for how we are to understand cases such as Millicent Fawcett’s criticism of Beatrice Webb, which involves successful moral blame of someone under false consciousness. These seem incompatible as modes of cognising others, but, as we see with Fawcett’s criticism, there are nevertheless successful instances of their overlap. I thus conclude the thesis by gesturing to future work that may be engaged in order to address this concern.

1. Thesis Summary

In the foregoing chapters of this thesis, I have attempted to defend the concept of false consciousness against its detractors. I maintain that these philosophers, who often cite the Explanatory Problem, the Normative Problem, and the Agential Problem (or some combination of them), have been too quick to detract the concept at the expense of its critical utility.
In Chapter I, I first clarified the concept of false consciousness and its critical utility in analytic social philosophy and how it supplements extant work on ideology critique. Tracing through the genealogy of the concept in Marxian and Critical Theory traditions, I argued that the concept of false consciousness should be understood as a set of beliefs held by an individual that is epistemically, functionally, and genetically ‘false.’ Against alternative definitions that are motivated by a narrow view of the space of theoretical possibilities of the concept, I set forth the Explanatory Problem, the Normative Problem, and the Agential Problem as the key problems of false consciousness that are prima facie in tension with our understanding of social ontology and epistemology, normativity, and moral blameworthiness.

In Chapter II, taking up the case of Mary Ward’s false consciousness, I argued that the Explanatory Problem persists only because of a lack of clarity about the explanatory desiderata for any account of false consciousness. Clarifying them in my analysis of the Explanatory Problem (as Acquisition, Content, Scope, and Systematicity), I put forth a social constructionist approach to the concept of false consciousness that is able to satisfy all the desiderata as well as subsumed and supplemented previous approaches—namely, functionalism, psychologism, and processualism.

In Chapter III, I then argued that the Normative Problem for a critique of false consciousness was distinct from the normative problem for ideology critique due to the specific requirements of Continuity and Endorsement for the former. Focusing on prudential norms and the case of Maud Ellen Simkins, I considered how substantive realist and anti-realist approaches to well-being fail to meet Continuity and Endorsement, suggesting that this is due to an assumption about the individualist nature of (human) well-being challenged in the biological and cognitive sciences. And, in their place, I advanced a non-substantive, immanent approach to well-being to meet the two requirements. Such an approach is grounded in an enactivist ontology and
second-order normativity inspired by Wynter and the critical social ontology of Jaeggi.

In Chapter IV, in relation to the Agential Problem, I argued that the presumption that an agent under false consciousness is \textit{eo ipso} an inappropriate target of blame on the basis of irrationality and ignorance is false. Keeping in mind that false consciousness not only constrains but constitutes agency, I attended to explicit historical evidence of how particular women anti-suffragists satisfied both the \textit{rationality} and \textit{epistemic} conditions of moral blameworthiness. For the rationality condition, I considered how the irrationality charge is based on a mereological fallacy and that at least Louise Creighton and Ward’s local autonomy were not compromised in any responsibility-undermining way. For the epistemic condition, I considered how Creighton, Beatrice Webb, and Ward were culpably ignorant on the basis of the culpable mechanisms that maintained their ignorant beliefs. In this way, I provided at least three cases where false consciousness did not exculpate by conceptual fiat.

I hope to have shown that these various problems (Explanatory, Normative, and Agential), often thought to be intrinsic to the concept, are not in fact so and only \textit{appeared} to be problematic due to a lack of clarity with respect to the desiderata of each problematic, as well as particular ontological assumptions in social ontology and epistemology, well-being, and moral responsibility. I wish to conclude, however, by briefly conceding that there \textit{is} an occasional methodological limitation for the use of the concept of false consciousness which, while not rendering the concept incoherent, is in tension with our ordinary blaming practices. It is \textit{this} problem, which I will now go on to outline, rather than the other problems considered above, that I believe underlies—or should underlie—the unease we might have with using the concept of false consciousness.
2. A Methodological Limitation?

Consider the structural parallels between ordinary moral blame (for example, criticising someone who reneged on a promise) and critique (for example, criticising Creighton for believing that, and acting as if, men are politically superior to women). While these different types of criticisms most obviously diverge in terms of their targets of criticism (an individual agent vs an agent qua structural node), both types of criticisms are characterised as having diagnostic and ameliorative aims, bringing to attention problems of the criticised targets and orienting them towards correction.

<table>
<thead>
<tr>
<th>Concerns about Appropriateness</th>
<th>Moral Blame</th>
<th>Critique</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the appropriate norms to which the criticism appeals?</td>
<td>(a) What are the norms on the basis of which an individual agent is blamed?</td>
<td>(i) What are the norms on the basis of which a case of false consciousness is critiqued?</td>
</tr>
<tr>
<td>What the appropriate characteristics that the critic must have?</td>
<td>(b) What conditions must the blamer fulfil to blame an individual agent?</td>
<td>(ii) What conditions must the social critic fulfil to critique a case of false consciousness?</td>
</tr>
<tr>
<td>What are the appropriate characteristics that the criticised must have?</td>
<td>(c) What are the conditions under which an individual agent is blameworthy?</td>
<td>(iii) What are the conditions under which an agent is under false consciousness?</td>
</tr>
</tbody>
</table>
Concerns about (a) the norms on the basis of which an individual agent is blamed is the business of normative ethics as such; whereas the other three concerns about the appropriateness of moral blame have been the subject of extensive discussions in the moral responsibility literature. Justin Coates enumerates these as considerations of: (b) jurisdiction, pertaining to the conditions under which agents may legitimately blame another (e.g., business, standing to blame); (c) justice, pertaining to the conditions under which agents may be themselves blameworthy (e.g., rationality, awareness); and (d) transaction, pertaining to the conditions under which acts of blame are appropriate (e.g., proportionality, aim) (Coates 2020).

The concerns of this thesis, on the other hand, have been with the first three concerns about the appropriateness of a critique of false consciousness. The Normative Problem was concerned with (i) the norms of critique and (ii) the status of the critic, whereas the Explanatory and Agential Problems were concerned with (iii) the conditions under which agents could said to be under false consciousness. My concern in this concluding chapter, however, is with (iv) the conditions that the act of critique itself must fulfil. Specifically, I merely wish to observe how, in cases of overlaps, moral blame interacts with critique in a way that limits the latter. That is, it seems as though even if someone is blameworthy and under false consciousness with regard to the same action, there are nevertheless ‘transactional’ considerations for why they should not be blamed alongside critique, or vice versa.
The diagnostic and ameliorative aims of moral blame are clear: cognitive and affective recognition, by the blamed agent, concerning the moral significance of their action (for example, remorse) and their consequent modification of attitudes and behaviour in reconciliation with the moral community. It is evident that the participant, reactive attitudes constitutive of blame itself is oriented towards such reconciliation, as the blamer is related as a fellow member and representative of the moral community. In critique, the aims are similarly clear: cognitive and affective recognition, by members of a given social structure, of the ideological significance of their action and their consequent modification of attitudes and behaviour in reconciliation among themselves—that is, resistance or solidarity. Where moral criticism addresses the agent as an individual, social criticism addresses them ‘as the instance of a type.’ Recall that Lukács attends to class consciousness inasmuch as it instantiates “the appropriate and rational reactions ‘imputed’ to a particular typical position in the process of production” (Lukács 1923, 51). Or, as Marx puts it in his preface to the first German edition of Capital, Volume 1:

I paint the capitalist and the landlord in no sense couleur de rose. But here individuals are dealt with only in so far as they are the personifications of economic categories, embodiments of particular class relations and class interests. My standpoint, from which the evolution of the economic formation of society is viewed as a process of natural history, can less than any other make the individual responsible for relations whose creature he socially remains, however much he may subjectively raise himself above them. (MECW 35: 10) [emphasis mine]

There is thus an overlap in two different types of cognition where social and moral criticisms overlap. Put in Strawsonian terms, the incompatibility lies in having to adopt both the participant attitude and the objective attitude to the target in the selfsame criticism even though “they are, profoundly, opposed to each other” (Strawson 1962, 9). In the participant attitude, one addresses another human being based on “the attitudes and intentions towards us of
those who stand in [various interpersonal] relationships to us, and the kinds of reactive attitudes and feelings to which we ourselves are prone”—for example, resentment, indignation, compassion, and gratitude (ibid., 6). In the objective attitude, one addresses another human being “as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided” (ibid., 9).

What is particularly problematic for us is also that the application of the concept of false consciousness by the social critic threatens to redouble oppression in theoretically fixating the individual under an oppressed social category. Anita Superson unwittingly hints at this problem when she argues that

it is not the restriction of choices in itself that makes it wrong to blame right-wing women, but the way in which their choices get restricted, namely, by a system that is patriarchal, through and through. Right-wing women are raised in a system that is entrenched with patriarchal ideology, whether in its religious or secular form. Acting out of that system, they choose sexist lifestyles, and adopt the sexist values and beliefs associated with them. Since their choice of lifestyle, and of the values and beliefs associated with it, are severely restricted as they are made within the context of patriarchy, it is wrong to blame right-wing women. Blaming them limits their choices even further as it means they ought or ought not to act in certain ways, and as such, makes them victims twice. (Superson 1993, 55–56)

Although Superson’s argument is ultimately that the problem lies in the restriction of choice for those who are autonomous, I believe that there is a more plausible way to understand the problem with blaming ideologically compromised agent: one that locates the issue in the criticism itself. I suggest that the problem is that such criticism fixates them within an asymmetrical economy of representation (since blame requires that the
agent cognitively and affectively recognise the action as theirs) and yet enjoins that they overcome this fixation in the same act (cf. Dotson 2014, Laruelle 2015). That is, even if the critic is enactively engaged in a dynamic coupling of participatory sense-making with the criticised, this interpersonal dynamic remains circumscribed within the wider dynamic of the criticised being such-and-such an oppressed agent at the particular structural node (cf. Ritchie 2021, forthcoming).

The coinciding of social criticism and moral criticism thus seems to be self-frustrating because the agent is being blamed in the participant attitude for an act that is only apprehended in the objective attitude and the critic’s categorisation itself reproduces the very oppression which it aims to overcome. Whereas I argued in Chapter IV that the concept of false consciousness does not *eo ipso* undermine moral blameworthiness, and hence is not always *inappropriate* (such as when agents under false consciousness meet the rationality and epistemic conditions of blameworthiness); here, we find that the overlapping of social and moral criticisms seem to frustrate each others’ aims, such that a critique of false consciousness is not always *appropriate*.

One seemingly straightforward response is simply to keep these criticisms separate. Even if Creighton, Webb, and Ward were blameworthy, we should refrain from moral criticism of them for contributing to oppression. Social criticism of them on the basis of crises in well-being (more specifically: their autopoietic systems), on the other hand, would not be problematic—as was the case with Simkins. Or, if we were to blame them as individual moral agents, we should not make reference to the structural significance of that action or their situation, attending only to their violation of moral obligations.

But this seems deeply unsatisfactory: any stipulated distinction in the moral and social domains is greatly complicated by the fact that conceptions of
well-being tend to inform moral norms (as I noted in Chapter III, §2) and that our understanding of an agent's position within a social structure often informs our justice, jurisdictional, and transactional considerations in moral blame (the very concepts of testimonial and hermeneutical injustices, for example, assume social marginalisation). Further, as I have observed at the outset of the previous chapter (Ch. IV), our blaming practices are thought to have a crucial communicative function and by maintaining moral and social criticisms as distinct projects, we preclude ourselves from appreciating how moral criticism can itself enjoin collective resistance.

Perhaps a better solution might be found in Robin Zheng's recent proposal to distinguish between “summative” and “formative” forms of moral criticism: that is, to distinguish between criticism that involves “a negative sanction against agents who wrongly exercise their agency in ways that violate moral standards,” and criticism that involves “feedback that identifies behaviour falling short of moral ideals, in order to motivate, inform, and reinforce our efforts to improve,” respectively (Zheng 2021, 11–12). The latter would include responses that rely on the notions of forward-oriented responsibility and responsibility-taking that I observed towards the end of the previous chapter (Ch. IV, §3). The thought here then is that overlaps between social criticism and moral criticism can occur if the latter is understood in the *formative* sense. It seems entirely consistent, for example, that one can point out to ways in which, by refusing the hermeneutical resources of the suffragists, Creighton fell short of both emancipatory consciousness and moral-epistemic ideals of humility and open-mindedness.

---

153 That is, at least under Fricker's account (2007). For an account of hermeneutical injustice that does not require marginalisation, see Simion 2019.
Towards a Theory of Moral Criticism as Social Criticism

And yet, there are actual instances of successful moral blame for an act of false consciousness as overlaps between social criticism and moral criticism in the *summative* sense. We saw this in the previous chapter, with Webb’s attributing her defection to suffragism to Fawcett’s blaming her for supporting the Appeal and calling her out on her classism. If we understand resentment to be constitutive of blame, the exchange between Fawcett and Webb appears to be a case where both a critique of false consciousness and an act of moral blame (as summative criticism) coincided in a single act. Yet it represents an instance in which both criticisms were *successful*. Not only did Webb demonstrate meta-lucidity but also remorse (as I observed in Chapter IV, §2.2).

What the conditions are for ordinary blame to contribute to successfully eroding false consciousness, then, is an important topic for further investigation. It is perhaps that such erosion occurs not only over time and within a context of alternative schemas, but with a dynamic coupling (and corollary participatory sense-making) formed independently of pernicious looping mechanisms. For example, such an investigation might begin with the observation that, after Webb had privately withdrawn from the anti-suffragist campaign, her partner Sidney Webb likely also played an instrumental role in encouraging her to eventually make public her renunciation of the anti-suffrage cause and to actively support the suffrage cause (cf. Caine 1983):

> You asked me to ‘form’ you. That means criticism: very difficult for me! But it has struck me that you underestimate the power of ‘party’ in social reform; the ‘great instrument of progress’ which party organisation is; and the great need there is for people to stimulate such parties by belonging to them. [...] And

---

154 Although she had learnt of his existence in 1888, she only met him two years later in January 1890, six months after Fawcett’s retort (Webb 1929, 402–406).
I think you will one day feel the need of more obviously and actively ‘repenting’ about Woman’s suffrage. Are you acting quite ‘honest’ about that? (Sidney Webb, letter to Beatrice Potter, n.d., 1890)

My wife bids me add—I have converted her! (Sidney Webb, letter to Ruth Cavendish Bentinck, Jun 12, 1912)

Attending to the instances and conditions of successful overlaps of social criticism and summative moral criticism, of course, should not understate the importance of formative moral criticism—just as attending to the instances and conditions of successful critiques of false consciousness should not to understate the importance of ideology critique. I suggest, then, that future work in social and moral philosophy should also pay attention to cases such as this, to provide us with a better understanding of the methodological tension between critiques of false consciousness and interpersonal blame and how it can be resolved. Such attention would allow us to better refine the function of moral blame within critical projects, even in cases of false consciousness: this would explicate an additional approach to solidarity-building in emancipatory praxis, one that is grounded in our quotidian practices of holding each other morally accountable. That is, we should also attend to moral criticism as social criticism.
Bibliography

Note on Abbreviations


Primary Texts of Anti-Suffragists and Suffragists


Other References

YouTube’s recommender system.” Synthese. https://doi.org/10.1007/s11229-020-02724-x.


213 of 237


Crozier-De Rosa, Sharon. 2010. “The New Woman at Home and Abroad: Fiction, Female Identity and the British Empire.” In Europe’s Expansions


Fanon, Frantz. 1961 [2004]. *The Wretched of the Earth*. Translated by
Farris, Sara R. 2017. *In the Name of Women’s Rights: The Rise of
   the Psychological Turn.” *Signs* 43(4): 775–795.
Fisher, Mark. 2009. *Capitalist Realism: Is There No Alternative?* Winchester:
   Zero Books.
Fletcher, Guy. 2013. “A Fresh Start for the Objective List Theory of Well-
   Being.” *Utilitas* 25(2): 206–220.
   Routledge.
   ——. 2019. “Against Contextualism about Prudential Discourse.” *The
   Justification*. Translated by Ciaran Cronin. Oxford: Oxford University
   Press.
van Fraassen, Bas. 1996. “Science, Materialism, and False Consciousness.”
   In *Warrant in Contemporary Epistemology: Essays in Honor of Alvin
   Plantinga’s Theory of Knowledge*, edited by Jonathan L. Kvanvig, 149–
Free, Melissa. 2016. “British Women Wanted: Gender, Genre, and South
   African Settlement.” *The Oxford Handbook of Victorian Literary Culture*,


