COPYRIGHT AND USE OF THIS THESIS

This thesis must be used in accordance with the provisions of the Copyright Act 1968.

Reproduction of material protected by copyright may be an infringement of copyright and copyright owners may be entitled to take legal action against persons who infringe their copyright.

Section 51 (2) of the Copyright Act permits an authorized officer of a university library or archives to provide a copy (by communication or otherwise) of an unpublished thesis kept in the library or archives, to a person who satisfies the authorized officer that he or she requires the reproduction for the purposes of research or study.

The Copyright Act grants the creator of a work a number of moral rights, specifically the right of attribution, the right against false attribution and the right of integrity.

You may infringe the author’s moral rights if you:

- fail to acknowledge the author of this thesis if you quote sections from the work
- attribute this thesis to another author
- subject this thesis to derogatory treatment which may prejudice the author’s reputation

For further information contact the University’s Director of Copyright Services

sydney.edu.au/copyright
Identity, Morality and Communicativeness: The Principles of Conscience

Andrew Donovan Donnelly

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Department of Philosophy
Faculty of Arts
University of Sydney
February 5 2014
I declare that the research presented here is my own original work and has not been submitted to any other institution for the award of a degree.

Signed: ....................................

Date: ........5/2/14
Abstract

What is the conscience? In this thesis I argue that, contrary to received wisdom, the conscience is not simply the moral sense, or the capacity to make moral judgements. I argue that conscience cannot be analysed in terms of any individually necessary conditions. Conscience (or as it is sometimes called ‘conscientiousness’), is a cluster concept. It can be analysed in terms of three principles which do not necessitate some thought or act being conscientious but, rather, count in favour, of that thought or act being conscientious. The three principles of conscience are identity (‘conscientious identity’), morality and communicativeness. Over the course of the thesis I articulate an argument in favour of my account of conscience through reference to both historical accounts of conscience (such as the accounts of Bishop Butler, Kant and Heidegger) and to recent discussions in moral philosophy (including the work of Kimberley Brownlee, Michael Bratman and Marina Oshana). Ultimately, I argue, these three principles are the principles of conscience because they capture, in different ways, what is valuable about conscience; they capture what a right of conscience purports to protect.
Acknowledgements

First and foremost, many thanks to my Principal Supervisor, Caroline West. Caroline has always provided insightful commentary on my work and kept me enthusiastic about my project. Thanks also to my Associate Supervisor, Tom Dougherty. Tom’s careful reading of drafts and attention to detail has been invaluable.

Finally, thank you to the many people I have had conversations with over the years about the content of my thesis; especially, Macintosh Stewart, Nick Malpas, Ben Cross, Robyn Kath, Luke Russell, Matthew Hammerton, Emma Wright, Tom Hill and Walter Sinnott-Armstrong
# Contents

**Part I: Theories of Conscience**

| Introduction | 2 |
| 0.1. The Assumption of Conscience as Moral Agency | 2 |
| 0.2. Conscience in the Mid-Twentieth Century | 5 |
| 0.3. Conscience in Recent Moral and Political philosophy | 8 |
| 0.4. The Conceptual Methodology of Conscience | 11 |
| 0.5. The Structure of the Thesis | 14 |

**Chapter One: Conscience as Moral Agency**

| 1.1. The Moral Agency/Identity Distinction | 20 |
| 1.2. Butler, Smith and ‘Empirical’ Conscience | 26 |
| 1.3. Kant and the ‘Inner Court’ of Conscience | 31 |

**Chapter Two: Conscience as Identity**

| 2.1. Heidegger and the Project of *Being and Time* | 41 |
| 2.2. Dasein, Existential Conscience and Existential Guilt | 45 |
| 2.3. Existential Conscience and Morality | 50 |
| 2.4. Arendt and Conscience as 'Living with Oneself' | 55 |

**Part II: Communicative Conscience**

| Chapter Three: Brownlee and the Communicative Principle | 64 |
| 3.1. Varieties of Disobedience | 65 |
| 3.2. The Consistency Requirement of the Communicative Principle | 72 |
| 3.3. The Universal Normative Scope of Conscientious Conviction | 75 |
| 3.4. The Non-Evasion Requirement | 78 |
Chapter Four: The Communicative Principle and the Normative Heads of Conscience

4.1. The Humanistic Head of Conscience

4.2. The Psychological Benefit Head of Conscience

4.3. The Societal Value Head of Conscience

4.4. The Normative Standing of the Communicative Principle

4.5. The Dialogue Requirement

4.6. The Communicative Principle and Personal Disobedience

4.7. Ceteris Paribus Clauses and the Principles of Conscience

4.8. The Cluster Concept of Conscience

Part III: Conscientious Identity

Chapter Five: Les Misérables and Conscientious Identity

5.1. Paradigm Cases of Conscience

5.2. The Transfiguration Case

5.3. The Deliberation Case

5.4. The Distinctness Requirement and Practical Identity in Korsgaard and Taylor

5.5. The First-Personal, Practicality and Endurance Requirements

5.6. The Meaning of Conscience in Les Misérables

Chapter Six: Conscientious Identity and Identification

6.1. Frankfurt and Identification

6.2. Bratman and the Planning Theory of Agency

6.3. Bratman and Policy-Based Identification

6.4. Bratman and Self-Governing Policy-Based Identification
6.5. Identification and the Four Requirements of Conscientious Identity ..................169

Chapter Seven: Conscientious Identity as Narrative Self-Conception ...............174

7.1. Neo-Aristotelian Character ............................................................................174
7.2. Bernard Williams and 'Deep' Character .......................................................178
7.3. The First-Personal Perspective and Self-Conception ................................185
7.4. The Narrative Constraint on Self-Conception ............................................193
7.5. Narrative Self-Conception and the Practicality Requirement ....................197
7.6. Valjean and Narrative Self-Conception: ......................................................199
7.7. Narrative Self-Conception and the Normative Heads of Conscience ..........201
Part I: Theories of Conscience
Introduction

0.1. The Assumption of Conscience as Moral Agency

What is the conscience? It has become common recently, in both philosophy and related fields, to assume without acknowledgement that the conscience is the individual’s capacity for moral agency (henceforth, this view is referred to as ‘conscience as moral agency’). More specifically, conscience is often thought to be the capacity of the individual to make moral judgements.¹ In this thesis, I argue that morality is only one aspect of conscience. To focus on the moral aspect of conscience is to neglect the communicative and identity-implicating aspects of the concept. It is the identity-implicating aspect of conscience, perhaps, which is the most controversial aspect of my proposal. In broad terms, this is the idea that thoughts and actions of conscience are cases where the agent thinks or acts true to self. This alternate view does have historical precedent (particularly in the work of Heidegger and Arendt), and some version of this view is quite often appealed to in contemporary bioethics. However, those historical accounts are woefully obscure in formulation. Furthermore, contemporary accounts in bioethics make reference to identity only in passing; there is no attempt to demonstrate how the notion of identity implicated in conscience relates to other notions of identity commonly defended in contemporary ethics.

A large portion of this thesis is devoted to trying to capture this connection between conscience and identity. I eventually conclude that a thought or action of conscience that implicates identity, that involves being ‘true to self’, is a thought or action that implicates the

¹ I will consider some other formulations of ‘conscience as moral agency’ at the beginning of chapter one. These other formulations include the view that the conscience is the capacity to know the moral truth, or the capacity to have certain moral feelings (such as guilt). What unites these different accounts as formulations of conscience as moral agency is that each account seeks to explain some aspect of our moral thought.

In a recent survey of accounts of conscience, John Skorupski (2010) argues that there are two main conceptions of conscience: Conception A holds that conscience is a matter of self-judgement; the individual’s judgement as to whether her actions are in accordance with her moral beliefs. Conception B holds that conscience is the source of moral knowledge ((2010, p552).
individual’s narrative self-conception. I define narrative self-conception in the following way: An individual thinks or acts in accordance with her conscience (in the identity-involving sense) when that thought or action implicates the individual’s beliefs that (a), involve ideals which make the individual’s life meaningful and (b), ‘cohere’ with her thought and action generally. Of course, there is a lot more that is going to need to be said about such a formulation before it can be assessed for its plausibility.

I claimed that it has become common to assume that conscience just is moral agency; that conscience is the capacity to make moral judgements or some other related moral capacity. What do I mean when I say they have simply assumed it? I mean that when the term ‘conscience’ is used by these authors, it seems to be used as a synonym for moral agency or some aspect of moral agency. The substantive discussion is then devoted to talk of moral agency and it is just taken as a given that such talk counts as talk of conscience.

For example, in a recent paper ‘Conscience: what is moral intuition?’ Tracey Finn and Paul Thagard claim that:

Conscience is the internal sense of moral goodness or badness of one’s own actual or imagined conduct. The products of conscience are moral intuitions, which are the feelings that some acts are right and others are wrong. This paper offers a new theory of conscience and moral intuition as a particular kind of emotional consciousness, produced by brain processes that combine cognitive appraisal with perception of bodily states.

In that paper, Finn & Thagard are actually offering an analysis of moral agency as emotional consciousness. They argue that moral judgement incorporates both cognitive and emotional elements. Moral judgement is cognitive as it concerns the individual’s beliefs about whether some act is appropriate. Moral judgement has an affective or emotional component, as our emotional reactions to situations (such as anger and disgust) are often of deep moral relevance. Finn & Thagard contrast their account with various other attempts to explain the human capacity for moral judgement. But they make that comparison irrespective of whether

\[\text{\textsuperscript{2}}\text{ I offer my own specific formulation of narrative self-conception, but this is not the only possible such formulation. In chapter seven I describe Schechtman’s own distinct account of narrative self-conception.}\]

\[\text{\textsuperscript{3}}(2011), p150.\]
the accounts in question even *purported* to be accounts of conscience. The assumption has been made that an analysis of moral agency *just is* an analysis of conscience. A crucial motivation in this thesis is that that connection cannot be assumed; it has to be demonstrated by substantial argument.

To a certain degree, it is trivial whether or not any particular capacity or ability of humans gets given the name ‘conscience’. And it may well turn out that, as a substantive matter, the conscience and the capacity for moral judgement are one and the same capacity. But making that assumption from the very beginning is a mistake, I suggest, for two reasons. First, it risks overlooking important aspects of our moral/practical psychology (such as those aspects involving identity) which only come to the fore when we are more open-minded about what conscience entails. Second, it overlooks some of the reasons we have to think moral rights of conscience are important.

A similar case of this pernicious assumption, this time from popular neuroscience rather than philosophy, is Robert Hare’s *Without Conscience: The Disturbing World of the Psychopaths among Us*. There, Hare offers an extended discussion of the moral deficit of psychopaths, something that he refers to as a lack of ‘conscience’. Hare uses the term ‘conscience’ to refer to “the pesky inner voice that helps us to resist temptation and to feel guilty when we don’t”. Again, the assumption is that the conscience is a specifically moral capacity (albeit in this case a capacity tied up with moral *feeling* rather than moral *judgement*). There is no consideration as to whether the conscience might involve something besides specifically moral capacities. I do not intend this point to be a criticism of Hare adopting that particular

---

4 For example, Finn & Thagard compare their account with accounts of ethical intuitionism; the view that there is a moral truth we can access via our moral intuitions. One intuitionist, G E Moore, did refer to his account of the moral sense as an account of conscience ((1903), p152). But some of the intuitionists referred to (such as W D Ross) did not offer their accounts as accounts of ‘conscience’.

5 A moral right of conscience, in this thesis, is the entitlement of an individual to act on the dictates of her conscience rather than whatever the law or some quasi-legal directive requires her to do.

6 (1993).

7 (1993), p275. A recent account of the moral deficit of psychopaths that makes the same assumption that conscience just is the capacity for moral agency is Moseley & Gala (forthcoming).

8 A recent case of an author who does not make this assumption when analysing the moral deficit of psychopaths is Lyons (2009). Lyons develops an integrity-based account of conscience that sees conscience as an individual’s commitment to treat her moral values as overriding. He does not simply assume that the conscience is the same thing as the capacity for moral judgement.
account of conscience. Plausibly, in this case, ‘conscience’ is being used as a purely theoretical term which is intended merely as a useful tool for understanding psychopathy. The point is that, if we want a philosophically robust conception of conscience, we cannot start with that assumption. It cannot be assumed without argument that the moral agency view is the only way of conceiving of conscience.

0.2. Conscience in the Mid-Twentieth Century

Before setting out the precise structure of this thesis it would be useful to describe the current state of field (both in philosophy and cognate fields) when it comes to theorising about conscience. The concept of conscience was a topic of substantial discussion within moral philosophy in the mid-twentieth century. Nearly all of these accounts defined conscience in terms of the moral capacities of the individual, whether all of the individual’s capacities for moral agency were defined as implicating conscience, or only some subset of those capacities. Two prominent accounts of conscience from that era were the accounts of C D Broad and Gilbert Ryle. Since then what meagre discussion of the concept of conscience has occurred has been in political philosophy and bioethics discussion of rights of conscience. I shall begin by discussing the accounts of Broad and Ryle. The focus of Broad and Ryle was in offering an a priori conceptual analysis of conscience, rather than to explain why, say, certain professionals should have an overriding legal right to act in accordance with

It is an interesting question whether the account of conscience I defend, conscience as narrative self-conception, might be a useful explanatory concept for understanding the deficit of psychopaths. Glenn et al. (2010) explain the moral deficit of psychopaths in terms of moral identity. Moral identity, the concept in psychology, concerns the referencing of one’s own self-concept in terms of moral character traits. Glenn et al. argue that psychopaths tend to have weaker moral identity than non-psychopaths. This notion of moral identity has some similarity to the account of conscientious identity that I go on to defend; conscience as narrative self-conception. My focus in this thesis is simply on the concept of conscience insofar as it is useful in justifying rights of conscience, so I leave it open whether or not such an analysis would also be useful in this sort of empirical context.

9 The noteworthy exception is Arendt’s account of conscience as identity which is the object of significant discussion in chapter two.

10 (1940) and (1940), respectively. Other important conscience as moral agency accounts from that period include Wand (1961), Fuss (1964) and McGuire (1963).

11 There have been some important exceptions to this. I devote two chapters to a discussion of Kimberley Brownlee’s recent analysis of conscience. Other important accounts have been offered by Mark Murphy (1997) and Larry May (1983).
their conscience. My focus in this thesis is also on that conceptual question. Having said that, as I shall argue in the next section of this introduction, the right of conscience is still important for my analysis as the ultimate ‘point’ of an account of conscience is to explain how a moral right of conscience might be justified.

Returning to the matter at hand, what were the specific details of the accounts defended by Broad and Ryle? Broad argued that the conscience has three necessary features.\(^\text{12}\) The first feature is cognitive; conscience involves active reflection on the morality of our thoughts, actions and ideals and arriving at moral beliefs. The second feature is affective; conscience involves a disposition to feel certain emotions such as remorse, guilt and approval, in response to morally charged situations.\(^\text{13}\) The third feature is conative; an individual with a conscience must be disposed to have certain motivations. This is to say that the individual must seek the good and shun the bad. Broad’s account is very clearly an account of conscience as moral agency.

We might explain away the cognitive and conative elements by suggesting that they just follow from a common analysis of what it is to make a moral judgement.\(^\text{14}\) The affective element, perhaps, is the element that is more often thought to be a unique feature of conscience rather than simply something that follows from the fact that an individual has made a moral judgement.

One particularly important aspect of this account, for my purposes, is that, on the face of it, it does not involve any particular meta-ethical or normative ethical commitments. There is no assumption made about whether the individual has epistemological access to the moral truth. Nor are there any other meta-ethical assumptions such as the assumption that there is such a thing as moral truth. There are no normative ethical assumptions as there is no assumption that any particular theory of right or wrong or good of bad is correct: this account of conscience does not assume that utilitarianism or Kantianism or any other normative theory is

---

\(^{12}\) (1940), p118. Or, at least, this is the analysis of his ‘wide’ notion of conscience. Broad suggests that there is also a ‘narrower’ sense of conscience where it is taken to involve teleological obligations only ((1940), p122).

\(^{13}\) In emphasising the role of emotions, Broad’s account is similar to the accounts of Hare and Finn & Thagard

\(^{14}\) The term ‘cognitivism’ in meta-ethics is usually used to refer to the fact that moral judgements are necessarily beliefs (rather than, say, desires). And this view is probably the most common view about the nature of moral judgements. The view that moral judgements necessarily motivate the individual to act in accordance with that judgement is often called motivation internalism. Again, this is a common view to hold in meta-ethics (see Rosati (2008)).
true. This neutrality is something that I adopt as a core methodological commitment in my analysis as well.\textsuperscript{15} My approach (which is not to denigrate other approaches) is purely conceptual: If humans have a conscience, I claim, it is something that humans can have irrespective of what the correct meta-ethical and normative ethical theories are.\textsuperscript{16}

Ryle, like Broad, offers an account of \emph{conscience as moral agency}. But Ryle notes an interesting fact about judgements of conscience in particular that does not apply to moral judgements generally. The conscience, Ryle argues, is essentially first-personal; it concerns the individual’s judgements about the morality of \emph{her} own actions. Not the morality of the actual or potential actions of others.\textsuperscript{17} Ryle offers an explanation as to why this is. Like Broad, Ryle argues that conscience is not simply a cognitive capacity but requires possessing certain motivations. To have a conscience is not only to have a belief about some state of affairs, but to have a certain \emph{conviction}. And to have a conviction is to be disposed to act in a certain way. It is because the operation of conscience necessarily involves this disposition to act in a certain way, Ryle claims, that the individual’s conscience does not tell her how she must behave. Your conscience concerns what you are convinced you must do; not what you are convinced anyone else must do. Whatever we may think about Ryle’s argument in support of his claim, I suggest that the first-personal aspect of conscience is an essential element of any analysis and cannot be overlooked.

The analyses of Ryle, Broad and their contemporaries were not particularly influential on later philosophy. Perhaps the most influential theoretical discussion of conscience in the twentieth century has not been in philosophy but in various cognate fields. I have already mentioned that it has played a role in the neuro-scientific discussion of psychopathy. Another area where conscience has been an important theoretical concept is in psychoanalytic theory. In Freud’s theory of the mind he proposes a construct that he calls the ‘superego’. The task of

\textsuperscript{15} This claim that I make no normative ethical assumptions might be surprising given that I appear to assume that \textit{rights} of conscience exist. However, strictly speaking, I make no such assumption. My argument is that if there were a right of conscience, it would dictate an analysis of conscience in terms of the three principles.

\textsuperscript{16} This is an important respect in which my analysis differs from the account of conscience defended by Kimberley Brownlee. Brownlee’s account of conscience (but not her account of ‘conscientiousness’, the two concepts are importantly distinct in her analysis) depends on the truth of moral pluralism.

\textsuperscript{17} (1940), p31. In chapter seven I develop an account of conscience that is consistent with a first-personal requirement. However, the first-personal requirement that I defend is a lot stronger than Ryle’s. I interpret the first-personal requirement as involving a first-personal \emph{perspective}: an account of conscience (or more specifically, conscientious identity) must implicate the individual’s beliefs about the sort of person she is.
the superego, within our minds, is to act as a critical authority. The conscience is a function of this superego. According to Freud, humans have a natural impulse or drive towards aggression. But civilized society prevents the individual from outwardly manifesting this impulse in their actions. As a consequence the conscience is directed towards the individual’s ego. And this tension between the super-ego and the ego is experienced by the individual as guilt; hence why we speak of a guilty conscience. The internal voice of conscience, for Freud, is a sort of self-punishment.

Conscience is also an important concept in developmental psychology. Lawrence Kohlberg theorised that a conscience (or a mature conscience, at least) is the capacity that arises in the final stage of an individual’s moral development. It is the capacity of the individual to reason according to universal and abstract moral principles.

0.3. Conscience in Recent Moral and Political philosophy

I discussed the accounts of conscience offered by Broad and Ryle in the mid-twentieth century, but what role has conscience played in philosophy since then? There has been a surge in interest in the accounts of conscience defended by various historical figures. In the next chapter I will consider some of those commentaries when assessing the conceptions of conscience of Adam Smith, Immanuel Kant and Martin Heidegger.

The other area in philosophy where conscience has been commonly discussed has been in the discussion of rights of conscience within political philosophy and bioethics. For example, John Rawls, Ronald Dworkin and Martha Nussbaum have all defended rights of conscience and in doing so they implicate various accounts of what the conscience is. Rawls defends a right of conscience in Political Liberalism and in doing so offers a particularly expansive

18 (1962), p84. A similar account had previously been defended by Nietzsche (1967). For a philosophical treatment of Freud’s account see Velleman (1999). Velleman argues for an account of conscience that interprets the voice of the Freudian conscience as a categorical imperative.

19 (1973). For a recent attempt to apply the Kohlbergian notion of conscience in explaining child moral development see Kochanska & Aksan (2006).

20 Sometimes the right of conscience is referred to as a right of conscientious action or a right of conscientious refusal. It is worth noting here that conscience need not just proscribe actions and therefore facilitate refusals, conscience can motivate the individual to commit positive acts. I consider this possibility in chapter two when pointing out some of the flaws in Hannah Arendt’s analysis of conscience.
notion of what is protected in a right (liberty) of conscience. For Rawls, it applies to the “religious, philosophical, and moral views of our relation to the world”. Rawls argues that since we often can’t easily abandon these views, and we don’t know which views we will have, if we were in the original position we would agree to a fundamental right of liberty of conscience.

Dworkin mentions conscience when arguing why it is that, sometimes, civil disobedience is justified. Dworkin distinguishes several types of civil disobedience and one of these types of disobedience, integrity-based disobedience, implicates conscience. When an individual disobeys a law on the grounds that she believes following that law would be deeply wrong, that it would affect her integrity, she appeals to conscience.

Nussbaum also defends a right of liberty of conscience. Nussbaum does not claim that the conscience is a peculiarly moral capacity. Nussbaum claims that the conscience is “the faculty in human beings with which they search for life’s ultimate meaning”. In a right of liberty of conscience we protect the individual’s capacity to do this. I discuss liberty of conscience in greater detail in chapter five when I consider the three ‘heads’ of normative justification for a right of conscience.

---


22 (1986), p107. The other types of civil disobedience, according to Dworkin, are ‘justice-based’ and ‘policy-based’ disobedience. The former case involves cases such as segregation protests where it is not the individual’s integrity that motivates the disobedience per se, but rather the belief that the law restricting that behaviour is unjust. Policy-based disobedience occurs when people believe not that the law is wrong as a matter of principle, but as a matter of fact. For example, one might disobey a conscription law not on the basis that one is a pacifist, but on the basis that the particular war in question is exceedingly stupid.

23 It is common for theorists to use the term ‘integrity’ when talking about conscience. It is not clear in this case why the claim that something is deeply wrong is equivalent to the claim that committing that act would bring the individual’s integrity into question.

24 At least, it is not a specifically moral capacity insofar as we hold that morality is concerned with ‘giving others their due’. On a more expansive definition of the term ‘moral’, considerations of meaningfulness might count as moral considerations because they involve categorical value: It might be thought that the meaningfulness of an individual’s actions is independent of an individual’s desires, just like the morality of an individual’s actions.

Most discussion of rights of conscience in recent times has not been in general political philosophy, but in bioethics. There, the discussion has concerned whether healthcare professionals (such as physicians, pharmacists or nurses) ought to have a right to act on their consciences. It is in this context that we see regular reference to the idea that an appeal to conscience is an appeal to identity, albeit in an inchoate form. Mark Wicclair argued that what is protected by a right of conscience is the individual’s “core moral beliefs”. In turn, Wicclair defines “core moral beliefs” as those beliefs that:

Comprise the subset of the agent’s moral beliefs that matter the most to the agent. They are integral to an agent’s understanding of who she is (i.e. her self-conception of identity). Accordingly, acting contrary to core moral beliefs is seen by the agent as an act of self-betrayal.26

Daniel Sulmasy defends a similar view:

Conscience thus arises from a fundamental commitment or intention to be moral. It unifies the cognitive, conative, and emotional aspects of the moral life by a commitment to integrity or moral wholeness. It is a commitment to uphold one’s deepest self-identifying moral beliefs; a commitment to discern the moral features of particular cases as best one can, and to reason morally to the best of one’s ability.27

In Sulmasy’s account we see the same reference to cognitive, conative and affective elements that was present in the conscience as moral agency account of Broad. But in this account they are unified by the notion of integrity. An individual with integrity will not only have a moral judgement but will experience the appropriate feelings and motivations that come with that judgement. But more importantly for our purposes, this commitment to integrity is taken to be the same thing as a commitment by the individual to uphold her self-identifying beliefs.28

---

28 For another account of conscience as identity in bioethics see Jeffrey Blustein:

Properly used, conscience indicates a particular way of seeing moral and other normative demands, a mode of consciousness in which prospective actions are viewed in relation to one’s self and character ((1993), p294).

For an account of conscience as identity in the constitutional law context see Eisgruber & Sager:
As I mentioned earlier, the problem with these appeals is that ‘identity’ is only referred to briefly with little attempt to expand on what exactly is intended by using the term ‘identity’. Given that there has been much discussion of identity in recent moral philosophy, it would be useful for there to be an explanation of how conscience relates to those notions of identity. Such an explanation is a primary goal of this thesis.

Given the wide variety of accounts of conscience that I have mentioned thus far - both within philosophy and in cognate fields - is there any reason to think that there is such a thing as the correct account of conscience? I don’t think there is. I think that there are various ‘correct’ accounts of conscience depending on what goal the theorist articulating a particular account of conscience has in mind. It may be that Freud’s account is correct if one’s goal is to understand why it is that humans experience guilt;²⁹ it may be that Kohlberg’s account is correct if one is trying to understand the way in which a child’s capacity to think morally changes as she ages; it may be that the account implicit in Hare’s discussion is correct if one is trying to understand the moral deficit of psychopaths.

0.4. The Conceptual Methodology of Conscience

In this thesis I outline an account of conscience which is designed to satisfy one ultimate goal; to explain the notion of conscience that is implicit in the moral right to act on one’s conscience. If moral rights of conscience were to exist, what is it that would make them valuable or worthy of respect? However, I shall suggest, in articulating such an account we have to make reference to the functional role that conscience plays in our moral psychology as well.

I attempt to explain what the justification could be for an individual being entitled to act in accordance with her conscience (possessing a right of conscience).³⁰ I argue that there are

---

²⁹ Though, admittedly, Freud’s account does seem dubious.

³⁰ As I have already mentioned. This is not, strictly speaking, a claim about what justifies a right of conscience. It is a conditional argument; an argument about what would justify a right of conscience if there were such a thing.
three principles or ‘heads’ of normativity that explain why it is that individuals’ possess a moral right of conscience. 31 These are the humanistic, psychological benefit and societal value heads of conscience. Using these three heads of normativity, I arrive at three principles of conscience. I suggest that advocates of the conscience as moral agency approach were right to emphasise the importance of morality in an analysis of conscience, but they were wrong to think that it is the definitive mark of conscience. Endorsing the arguments of Kimberley Brownlee I argue that as well as the morality principle there is a communicative aspect to conscience that ought to be captured in a separate communicative principle of conscience. This principle captures the sense in which, the more an individual seeks to engage in reasoned dialogue with others about her convictions, the more conscientious her conviction. Finally I defend the identity principle (which I also refer to as conscientious identity). An important way in which an individual’s thought and actions can be conscientious is through implicating the individual’s identity.

In defending these three principles, I am not attempting to establish any individually necessary conditions for the application of the concept of conscience. I argue that conscience is a cluster concept. A principle, in my analysis, is a consideration which conceptually counts towards some object falling under the concept in question rather than necessitating anything.

In analysing Art as a cluster concept, Berys Gaut offers the following account of what it is for some consideration to ‘count towards’ an object falling under a concept:

First, if all of the properties that are criteria are instantiated, this suffices for an object to fall under the concept; and more strongly, if fewer than all of these properties are instantiated, this also suffices for the application of the concept. So there are jointly sufficient conditions for the application of the concept. Second, there are no properties that are individually necessary conditions for the object to fall under the concept (that is, there is no property that all objects falling under the concept must possess). Third, there are disjunctively necessary conditions for application of the concept: some of the properties must be instantiated if the object is to fall under the concept. 32

---

31 I endeavour to use the term ‘heads’ for these three normative principles in order to avoid confusion between these three principles and the three principles (identity, morality and communicativeness) that are constitutive of the cluster concept of conscience.

32 Gaut (2005), p274.
This analysis of cluster concepts can be used to understand my analysis of conscience. The implication of any one of the three principles of conscience is sufficient for some thought or action to count as a thought or action of conscience (i.e. to count as *conscientious*). And in order to count as conscientious, it is necessary that the thought or action in question implicate at least one of the three principles.

The fact that satisfying any one of these principle is unnecessary for an action to count as conscientious does not mean that the three criteria are of equal importance or weight. I argue that the communicative principle is of the least weight. The weightiest principles are the identity and morality principles. Both of these principles contribute significantly to the conscientiousness of a thought or action. And the reason that they are so weighty, I argue, is that those two principles best capture the three normative heads of conscience. The communicative principle is the least weighty of the three principles, I argue, as this principle is the one least supported by the three normative heads. My emphasis is on the identity principle as it is this principle which, I argue, has been insufficiently explored in contemporary moral philosophy.

My emphasis on the normative import of conscience distinguishes my account from some of the other authors I have discussed in this introduction. While the conceptual analyses of conscience offered by Broad, Ryle and their contemporaries did make reference to the normative purport of an account of conscience, the focus in their accounts was very much on the folk concept of conscience. They were putative accounts of what the folk would and would not be disposed to judge counted as a conscientious thought or action. Is there any room for such descriptive ‘folk’ intuitions in my account? I think there is. While each of the principles I consider is not a necessary condition of conscience, each of those principles can themselves be analysed in terms of necessary conditions or ‘requirements’ of that principle. When discussing the communicative principle I defend Brownlee’s requirements that communicative thought and action be non-evasive and dialogue-promoting. When discussing the morality principle I consider the idea that moral requirements are categorical. When discussing the identity principle, conscientious identity, I set out four descriptive requirements of that principle; the distinctness, first-personal, practicality and endurance requirements. In articulating the latter principle I draw on a literary case of conscience being applied; Victor Hugo’s *Les Misérables*. 

To summarise this rather bewildering methodology: It involves four ‘steps’. To begin with, I suggest there is no one correct account of conscience. The incredible vagueness and ambiguity in the use of the term ‘conscience’ by ordinary people as well as the great variety of theoretical uses of the notion mean that in offering a viable analysis of the concept we have to look closely at what the point of our analysis is. I suggest that the point of our analysis is to capture what it is that would justify a moral/political right of conscience, if such a right were to exist. That is, what would justify an individual’s right to act on their conscience, even if the law prohibits her from doing so? The second step is to give an account of those grounds that potentially justify a right of conscience. I suggest that three broad heads of normativity be established: the humanistic, psychological benefit and societal value heads. The third step is to come up with some principles of conscience that can be justified in terms of these three normative heads. I claim that principles of identity, morality and communicativeness can be so justified. However, I argue, these principles are not all of equal importance; I argue that identity and morality are the weightiest of these principles. The fourth step involves a specification of the individually necessary conditions, the ‘requirements’ of each principle.

0.5. The Structure of the Thesis

Now that I have given some account of the recent history of theorising about conscience, as well as an explanation of the relatively complicated methodology employed in this thesis, it is time to outline the precise structure of this thesis. In chapter one I begin my discussion of prominent historical accounts of conscience. In that chapter I focus on what I take to be the dominant trend in modern philosophy, conscience as moral agency. I describe the conceptions of conscience as moral agency defended by three major early modern philosophers; Adam Smith, Bishop Joseph Butler and Immanuel Kant. Each of these authors emphasises the role of moral judgement in conscience. Then, in chapter two, I go on to consider two significant conceptions of the conscience as identity approach. I describe the theories of Martin Heidegger and Hannah Arendt. Both of these theorists argue that an appeal to conscience is a call to the individual to be ‘true to self’. That completes Part I of the thesis.
In Part II I move on to an evaluation of a recent account of conscientiousness; the account defended by Kimberley Brownlee. Brownlee argues that conscientiousness is grounded in a communicative principle. According to Brownlee, an understanding of this principle supplies us with the necessary conditions of conscientiousness. In chapter three I address the distinction between civil disobedience and personal disobedience and begin evaluating Brownlee’s argument that civil disobedience is a more conscientious act than personal disobedience. I consider the first two requirements of that principle, universality and consistency. I argue that while it is plausible that these are requirements of conscientiousness, they have nothing to do with the communicative aspect of conscience per se. In the case of the universality requirement, I suggest that, while this is a plausible requirement of conscientiousness, it has little to do with the communicativeness of an action, but pertains to the fact that some action involves a moral conviction. This paves the way for my suggestion in chapter four that we ought to distinguish a communicative principle and a moral principle of conscientiousness. The consistency requirement, while again a plausible requirement of conscientiousness, seems to have little to do with the communicativeness of a conviction or action and I suggest that requirement is better captured in yet another principle, the identity principle. In order to evaluate the final two requirements that constitute Brownlee’s communicative principle, the non-evasion and dialogue requirements, I suggest, we need to specify what exactly the normative grounds for a right of conscience are.

In chapter four, drawing on Brownlee’s justification for the communicative principle, I state the argument for the humanistic, psychological integrity and societal value ‘heads’ of normative justification. In making this argument, I draw on not only the arguments of Brownlee, but also the arguments of John Stuart Mill and John Rawls as to the value of conscientious action. With an understanding of those normative heads of conscience, I am then in a position to evaluate Brownlee’s third and fourth requirements of the communicative principle; the non-evasion requirement and dialogue requirements. Following Brownlee, I accept that these are plausible requirements of conscientiousness if and only if they contain ceteris paribus clauses. But once we recognise these clauses, I suggest personal disobedience (what we might often refer to as ‘conscientious objection’) is just as conscientious as civil disobedience.

---

33 Strictly speaking I critique an account of conscientiousness rather than conscience. As I mentioned earlier, Brownlee argues that there is an important difference between these two concepts. Because I don’t consider Brownlee’s account of conscience proper, I use the terms ‘conscience’ and ‘conscientiousness’ synonymously.
I then go on to suggest that we do away with these clauses altogether as their function is better fulfilled by simply denying that the communicative principle is a necessary condition of conscientiousness. I argue that a better way of capturing the fact that some feature of a conscientious action need not hold in every case is by positing that that feature is only a weighty, rather than a necessary, condition of conscience. Framed in that way, it is acceptable for any given case of conscience not to implicate communicativeness (or morality, or identity) as long as it implicates one of the other principles of conscience. At the end of the chapter I suggest that three separate principles communicativeness, morality (these two principles a consequence of severing Brownlee’s communicative principle into two components), and identity, can all be justified in terms of the normative heads of conscience.

In Part III of the thesis I begin an in depth analysis of the identity principle, what I call ‘conscientious identity’. In chapter five I try to articulate exactly what could be meant by an appeal to ‘identity’ in this context. Following other commentators (such as Kimberley Brownlee and Jonathon Bennett) I take literary cases of conscience to be good paradigm cases of the concept at work from which we can extract the particular requirements of conscience (in this case, conscientious identity). In this chapter I describe in detail a famous literary case of conscience in operation; the moral crises experienced by the character Jean point in Victor Hugo’s Les Misérables. This case is useful in bringing out what I consider to be the four fundamental requirements of conscientious identity; the distinctness, first-personal, practicality and endurance requirements.

In chapter six I begin to consider how an account of conscience can be reconciled with various accounts of identity from recent moral philosophy. In this chapter I focus on the accounts of identification defended by Harry Frankfurt and Michael Bratman. I argue that neither account is a viable account of conscientious identity as neither can satisfy the crucial first-personal requirement. However, the notion of a quasi-policy described by Bratman and incorporated into his account of identification does play an important role in the account of conscientious identity that I go on to defend.

In the final chapter, chapter seven, I look at three other accounts of identity: I consider neo-Aristotelian character, Bernard Williams’ deep character and Marina Oshana’s self-conception. I argue for an account of conscientious identity grounded in Oshana’s notion of self-conception, though I add to that idea a narrative constraint taken from accounts of
narrative identity. On my analysis of conscientious identity, conscience is a capacity to evaluate thoughts and actions in terms of (a) the individual’s beliefs about those of her ideals that contribute to the meaningfulness of her life, and (b) the extent to which those beliefs generally cohere with her thought and action. My articulation of narrative self-conception, just like the communicativeness and morality principles, can be justified in terms of the three normative heads of conscience.
Chapter One: Conscience as Moral Agency

After acting, we sometimes have an experience of conscience. Sometimes this is a painful feeling, the so-called 'pangs' of conscience. Sometimes it is an 'inner voice' of conscience that condemns our actions. Sometimes, on reflection, we feel no psychological pain nor hear any voice of condemnation and say that we have a 'clear' conscience. On some occasions conscience affects our practical deliberation; we experience conscience before acting. In those cases our conscience can 'prick' us into not performing an action. In this chapter and the one that follows I look at two distinct ways of thinking about the phenomena of conscience that have arisen over the last few centuries. In this chapter, my focus is on the first way of thinking about conscience, an approach I call conscience as moral agency. In broad terms, this is the view that conscience is the 'voice of right and wrong' within the individual. I will describe the conceptions of conscience as moral agency defended by Bishop Joseph Butler, Adam Smith and Immanuel Kant. The second way of thinking about conscience is the view that the conscience calls us to be ‘true to self’. I call this the conscience as identity approach and this is the approach I focus on in chapter two.

The purpose of this chapter is two-fold. First, I want to describe some historically influential accounts of conscience as moral agency in order to show what the historical ‘opposition’ to accounts of conscience as identity look like. Second, I want to show that conscience as identity is not as foreign account as it might at first seem. Elements of the conscience as identity approach are evident in some of the most influential historical accounts of conscience (specifically, in the accounts of Smith and Kant).

This chapter shall proceed as follows: In the first section I set out more clearly what the distinction between the two approaches is. In this section I also briefly discuss the origin of both approaches to conscience in medieval philosophy.

In the second section I consider the first candidate conception of conscience as moral agency; the account of Bishop Butler. Butler argued that the conscience is the ruling capacity of our moral agency and exercises authority over other aspects of our moral psychology. In section three I describe Adam Smith's 'empirical’ account of conscience. For Smith, judgements of conscience form a subset of our propriety judgements and involve an appeal to how an
idealised version of ourselves would feel in certain circumstances. In section four I consider and evaluate Kant’s conception of conscience. Conscience, for Kant, only makes a special sub-class of moral judgements; second-order moral judgements concerned with whether or not the agent has been scrupulous in her exercise of Practical Reason. In the accounts of both Smith and Kant we see reference to the ‘self’ as an important element of conscience and in this respect, elements of conscience as identity are present. This provides a useful entry point for considering the prominent accounts of conscience as identity which are the subject of chapter two.

1.1. The Moral Agency/Identity Distinction

In this chapter I focus on just two ways of thinking about conscience. But the moral agency/identity distinction is just one of the ways in which we might distinguish theories of conscience. For example, John Skorupski divides accounts of conscience into the following two categories: conscience as self-judgement and conscience as moral knowledge.¹ The first way of thinking about conscience concerns the way in which the individual can evaluate her own actions against the moral principles that she is committed to.² The second way of thinking about conscience sees it as the individual’s capacity to access the moral truth.

On my definition, both of Skorupski’s categories would count as conscience as moral agency. Each category simply focuses on a different aspect of our moral agency. We can think of many more possible distinctions as well. For example, we could distinguish theories of conscience by origin; some theories define conscience as an evolved mechanism, others define it is God-given access to the divine truth.

Another approach is to distinguish those accounts that emphasise the conscious quality of conscience and those that do not. This might be thought of a contrast between etymological and non-etymological approaches to conscience. As I shall discuss at the end of this section, the Latin term ‘conscientia’, from which the words ‘conscience’ and ‘consciousness’ are both derived, did not refer to a specifically moral capacity of the individual, but simply the

¹ (2010), p552.

² This notion has some similarities with Kant’s conception which I shall discuss shortly.
individual’s capacity for awareness. Christine Korsgaard describes a consciousness account of conscience: “Conscience is the psychological faculty by which we aware of and respond to the moral character of our own actions”.\(^3\) This approach can be contrasted with accounts that hold that the conscience is a necessarily non-conscious. We see this second approach in the transcendental accounts of conscience defended by Kant (see section five of this chapter) and Heidegger’s account (discussed in chapter two). For Kant and Heidegger, the conscience is not something that we are consciously aware of but something that is a necessary pre-requisite to our having certain psychological capacities.

Another way of distinguishing accounts of conscience is to distinguish affective and cognitive accounts. Affective accounts define conscience as a capacity for feeling a certain way. In the introduction I mentioned that an affective approach to conscience has been implicated in discussions of the moral deficit of psychopaths; it is assumed that the emotional deficiencies of psychopaths are deficiencies in conscience.\(^4\) We also saw that an affective component was an important element of several other accounts of conscience.\(^5\) Cognitive accounts, by contrast, define conscience in terms of an individual’s beliefs. We saw an emphasis on the cognitive component of conscience in the accounts of Rawls and Dworkin where conscience was defined in terms of the individual’s religious/moral/philosophical beliefs.\(^6\) The conception of conscience as identity that I go on to defend, conscience as narrative self-conception, is also a cognitive account of conscience as it concerns the individual’s beliefs about what would or does make her life meaningful. Given that there are so many different ways of distinguishing accounts of conscience, why in this thesis do I concentrate on the moral agency/identity distinction?

As we shall see, conscience as identity, the idea, is present in inchoate form in some of the most well-known philosophical treatments of conscience. While divorcing conscience from

---

\(^3\) (1996), p110. See also Nicholas Dent: “To have a conscience involves being conscious of the moral quality of what one has done, or intends to do.” (1998).

\(^4\) E.g. Hare (1995) and Moseley & Gala (2014).

\(^5\) The emotion of guilt was the consciously felt element of conscience in the accounts of both Freud and Nietzsche (1967). Emotions also played an important role in Thagard & Firth’s conception of conscience as emotional consciousness and in Broad’s account of the conscience as involving cognitive, conative and affective capacities of the individual (see introduction, sections one and two).

morality is, prima facie, a radical proposition, the presence of conscience as identity in some form in historically prominent accounts of conscience such as Kant’s and Smith’s lends it more plausibility. Focusing on the moral agency/identity distinction helps us to see that conscience as identity has historical precedent.

Before moving forward to discuss the first formulation or conception of the conscience as moral agency approach it is necessary to make this vague distinction between moral agency and identity more precise. I define moral agency as the possession of the capacities which are essential to moral thought. This could involve, for example, the capacity to judge acts to be right, wrong, obligatory, permissible, praiseworthy or blameworthy. It might involve some combination of those judgements. It is not just judgements, however, that constitute moral thought. It also involves having a particular feeling in response to morally-salient situations. It might require that the individual have certain desires in relation to her moral beliefs.

Being a moral agent is often thought to be a pre-condition to being morally responsible. An individual who was not a moral agent, might well end up behaving in a way that is morally valuable. An android that devotes its life to vaccinating children might well be doing something that is morally good, but they are not moral agents and wouldn’t generally be thought to be morally responsible for their actions.

It may be that moral agency only involves one discrete psychological capacity. For example, we can imagine a very crude utilitarian for whom moral agency might only require the capacity to weigh up which of several proposed actions will maximise utility. But more, plausibly, as I suggested earlier, moral agency requires several different psychological capacities working together. As well as the capacity to apply moral criteria (such as perceived utility), it requires correctly assessing the factual situation. For example, a competent moral agent is not only one who accepts a moral principle such as ‘give to those in need’, she must also judge that some particular factual situation involves a person in need.

Full moral agency also requires having certain motivational states; the moral agent experiences at least some degree of motivation to perform an action on recognising that that action is morally required. Plausibly, agency also requires the possession of certain affective

---

7 There are several different ways in which this connection between conscience and moral motivation could be made. One possibility is that it is a necessary truth that motivation follows from a moral judgement (at least in the normal case). For example, Michael Smith defends such a view:
capacities; the ability to have the appropriate feelings for a particular factual situation. An example might be the capacity to feel sad upon seeing a photo of a starving child.

Being a moral agent is probably not an all or nothing affair; one can be a deficient moral agent while possessing, to some extent, the capacities of a moral agent. Importantly, full moral agency requires judging the morality of both past and future actions of the agent. It requires judging not just what one intends to do but one’s own past actions as well.

Conscience as moral agency holds that the conscience is partly or wholly constitutive of moral agency. Different formulations or conceptions of this approach define the conscience in terms of a different aspect of moral agency. We have already seen how some formulations say that the conscience is a capacity for moral knowledge. Other formulations restrict the operation of conscience to just the moral judgements themselves, often with the addendum that the conscience is only concerned with moral judgements about the agent’s own conduct. Still others identify the conscience with the guilty feeling that follows an agent’s realisation that she has acted wrongly. In nearly all cases, possessing a conscience is not a sufficient condition for full moral agency, but it is a necessary one.

Conscience as identity holds that the conscience is the capacity to judge that a proposed or past action is or would be true to self. What do I mean by such an opaque expression? Perhaps the most prominent twentieth century theorist who discussed this notion of being ‘true to self’ or ‘authentic’ is Charles Taylor. Taylor describes the evolution of our idea of authenticity. He refers to the idea that morality is a voice within us and says that "authenticity develops out of a displacement of the moral accent in this idea". The idea of being true to oneself is a matter of following one’s internal voice telling oneself not simply to do the right thing, but to be oneself.

Perhaps the most straightforward way of thinking about authenticity is in terms of a self-conception. My self-conception is my set of beliefs concerning the sort of person I am. I appeal to my self-conception when I ask myself, ‘what sort of person am I’? I consider Marina Oshana’s account of self-conception in detail in chapter seven where I argue that

---

It is conceptually necessary that if an agent judges that that she morally ought to \( \phi \) in circumstances \( C \), then she is motivated to \( \phi \) in \( C \), or she is practically irrational (2007, p211).

conscientious identity is partially constituted by the individual’s self-conception. Though this might be the most straightforward way of thinking about conscience as identity, this approach does not require that the conscience is defined in terms of an individual’s psychological capacities. For example, in the next chapter I discuss Heidegger’s account of conscience as identity which does not purport to describe a psychological capacity at all. For Heidegger, conscience as identity is formulated as an ‘ontological’ or ‘existential’ notion which is offered as a pre-condition to the psychological capacity that we often call the ‘moral conscience’.

The question naturally arises, are these two approaches to conscience really distinct? Couldn’t conscience as identity be taken to be just one formulation of the conscience as moral agency approach? Presumably, one could argue that he capacity to judge or act in a way that is ‘true to self’, to be authentic, is a component of full moral agency. Christine Korsgaard offers an argument along these lines which I discuss in detail in chapter five. In short, Korsgaard argues that the individual’s possessing a self-conception (or a ‘conception of her practical identity’ as Korsgaard calls it) is the only possible basis for any normative judgement. I want to leave open this possibility. The two approaches to conscience discussed in this chapter do not pick out mutually exclusive categories; and as we shall see, both the accounts of Smith and Kant seem to involve elements of both approaches.

Given the apparent dissimilarities in these two approaches to conscience it might be wondered how they both came to be known by the name 'conscience'. Arguably, the roots of this divergence can be traced to the first millennium. Inspired by the apostle Paul's description of conscience in the New Testament, Jerome gave the following analysis:

> [it] is that which the Greeks call synteresis: that spark of conscience which was not even extinguished in the breast of Cain after he was turned out of paradise, and by which we discern that we sin, when we are overcome by pleasures or frenzy and meanwhile are misled by an imitation of reason.

---

9 Oshana, (2010).

10 In brief, Korsgaard argues that an individual must value her humanity, and valuing one’s own humanity requires valuing the humanity of others on pain of inconsistency. And this is what it is for an individual to have a moral identity ((1996), p121). I discuss this argument in detail in chapter five.

In this passage we see Jerome affirming a conception of the moral agency approach. Conscience here is the capacity to retrospectively judge that one has acted wrongly due to the individual letting her emotions get the better of her. But more important than this is another contribution that Jerome made (perhaps unwittingly) to theorising about conscience. In translating the Bible from Greek to Latin Jerome translated the Greek term ‘synteresis’ (sometimes spelt ‘synderesis’) using the Latin term 'conscientia'. While, as in the above description, 'synteresis' was a term commonly recognised in the Greek language as referring to an exclusively moral capacity, that was not the case with 'conscientia'. 'Conscientia' was a term used to refer to conscious thought generally and had no necessary connection to morality at all. Consequently, this translation of ‘synteresis’ left open the possibility for theorising about the conscience in a way that made no reference to the standard capacities of moral agency. This influence is particularly obvious in Hannah Arendt's account of conscience which I discuss in chapter two. Arendt, plausibly inspired by the etymology of conscience, sees the conscience as a side-effect of the capacity to think. There is no reference to anything that resembles a moral capacity.

More generally, this translation left the notion of conscience relatively open-textured and led to the flourishing of an incredibly diverse range of theories of conscience. Aquinas, for example, distinguished two different conscience-like capacities that were supposedly constitutive of moral agency. Synteresis, for Aquinas, was that capacity within us by which we had access to the most general and formal practical principles; principles such as 'avoid the wrong and do the right'. Conscientia was a capacity of practical reason, a capacity to apply moral principles in particular situations. While an agent might be aware of her duty to aid those in need, she also needs to know that that principle is to be applied when she comes across a homeless person. Despite the flourishing of conceptions of conscience in the

---


14 By practical reason here I simply mean the capacity of the agent to work out what she ought to do, not the Kantian faculty of 'Practical Reason' that I discuss in section three of this chapter.

15 There is an obvious gap in Aquinas' account of moral agency as I have described it thus far. Synteresis only gives very general principles, not principles with sufficient content to ground the particular judgements that are the work of conscientia. To round out his account of moral agency Aquinas posited that we learn the more specific content of the moral principles through experience guided by the virtue of prudence (see Langston (2001).
medieval period, it is in the early modern period that conscience is first defined in terms of what we would now think of as the central capacities of moral agency.

1.2. Butler, Smith and ‘Empirical’ Conscience

The concept of conscience plays a more important role in Bishop Butler's moral philosophy than it does for, perhaps, any other philosopher of the early modern period. Butler’s account is a paradigmatic account of conscience as moral agency; in giving an account of conscience Butler takes himself to be giving an account of what others might call the moral sense or moral reason.\(^{16}\) Butler describes a moral psychology that involves the operation of several different faculties or capacities. Butler focuses on four capacities; particular passions, self-love, benevolence and conscience.\(^{17}\) Our particular passions are our everyday inclinations and aversions. Self-love is a capacity that concerns a higher-order desire to satisfy a set of desires that constitutes our happiness. Benevolence is also higher-order as it involves our disposition to care about the happiness of other people. So what is the conscience? It is the capacity to assess and reflect on proposed and past actions and assess them by appealing to moral principles. Sometimes Butler restricts the conscience to the assessment of the agent's own conduct only (a point later made by Adam Smith) whereas at other times he sees it as applying to the assessment of the conduct of others as well.

Both Butler and Smith develop moral psychologies from an “empirical understanding of human nature”.\(^{18}\) Both authors aim to describe our observed psychology. In this respect, Butler and Smith differ from Kant and Heidegger. Kant’s account of conscience, the subject of section three of this chapter, is a transcendental account of conscience; it is an account of the way the conscience must be structured if we are capable of being reliable moral agents. In the next chapter I argue that Heidegger’s ‘ontological’ account of conscience is supposed to

\(^{16}\) For example, Butler talks of "Conscience, moral reason, moral sense, or divine Reason, as a sentiment of the understanding, or as a perception of the heart" (Dissertation on Virtue, Section 1, (1736)).

In this section and the one that follows when I refer to the work of both Butler and Smith I will reference their work by section number rather than page number, as is the convention.

\(^{17}\) In my analysis here I follow Hill (1998a) p27-31.

\(^{18}\) Hill (1998a), p27.
be an account of how the conscience must be structured if we are capable of being bound by moral norms.

For Butler, the distinction between benevolence and conscience is somewhat analogous to Aquinas’ distinction between *synteresis* and *conscientia*. Butler acknowledges the existence of something akin to Aquinas’ faculty of *synteresis* by holding that individuals do have innate moral tendencies. But, according to Butler, that ‘work written in their hearts’ (i.e. benevolence) needs to be distinguished from the ‘witness of conscience’.

By the ‘work written in their hearts’ Butler means:

the natural disposition to kindness and compassion...that part of the nature of man treated of in the foregoing discourse which, with very little reflection and of course, leads him to society, and by means of which he acts a just and good part in it, unless other passions and or interest lead him astray.\(^{19}\)

With regards to ‘conscience’, Butler says:

there is a superior principle of reflection or conscience in every man which distinguishes between the internal principles of his heart, as well as his external actions; which passes judgment upon himself and them; pronounces determinately some actions to be in themselves just, right, good; others to be in themselves evil, wrong, unjust.\(^{20}\)

According to Butler, the conscience stands in judgement of both the individual and her actions and assesses them in moral terms. Importantly, there is a reference to conscience being a ‘superior’ principle. This is a reference to the *authority* of conscience. The claim here is not that the conscience is what ultimately motivates the individual. Butler’s moral psychology allows for the possibility that individuals don’t act in accordance with their conscience. Rather, the claim is that the conscience has a rational authority over the other practical considerations that the individual might appeal to or that might influence her decision.\(^{21}\) One interesting implication of this view is that it would lead to a very strong right of conscience for the individual. If the individual’s conscience, for some reason or another,

---

\(^{19}\) *Fifteen Sermons*, Sermon II.

\(^{20}\) Ibid.

\(^{21}\) Mark Murphy defends a similar claim when he articulates ‘the conscience principle’. This principle holds that the individual ought never to act contrary to her conscience (1997).
always had authority over other normative considerations then, plausibly, the individual should be entitled to act in accordance with her conscience. In this thesis I argue only that the conscience is a weighty practical normative consideration; I leave it open how exactly considerations of conscience should be assessed against other normative considerations.\textsuperscript{22} For Butler, it is this capacity rather than any of the others that influence practical deliberation, which is definitive of moral agency.\textsuperscript{23} Conscience is also the capacity (though he often calls it a ‘principle’) which is constitutive of and distinctive of our nature; it is this capacity of conscience, rather than any of the others, that distinguishes us from the brutes.

Butler's account of conscience is a paradigm case of \textit{conscience as moral agency}. It takes conscience to be the 'governing' capacity in agents’ determining what they are morally required to do. Adam Smith’s account, while still falling squarely within the category of ‘moral agency’ also contains elements indicative of the \textit{identity} approach. Furthermore, conscience does not play as prominent a role in Smith’s moral psychology as it does in Butler’s account. This is partly a result of the different goals that Butler and Smith have in developing their accounts of conscience. Butler has a broader theoretical aim, to explain what it is that distinguishes humans from the brutes. Smith's goal is more like an early attempt at ‘empirical’ psychology. He wishes to describe that part of our moral psychology that we often refer to using the term 'conscience'. For Smith, the conscience is just one element in our moral psychology and not one that has any peculiar authority over our other moral psychological capacities.

Central to Smith's moral psychology, and hence his account of conscience, is the notion of \textit{sympathy}. I will briefly explain Smith’s account of sympathy before explaining how it grounds Smith’s account of conscience. For Smith, sympathy is a capacity by which we imagine ourselves as another person and imagine how they would feel in certain circumstances. We employ this capacity for sympathy in two different types of moral judgement. Smith claims that we can evaluate another person's actions either in terms of the

\textsuperscript{22} It is worth noting, however, that the claim that considerations of conscience are weighty does tell us \textit{something} about how they interact with other normative reasons. For example, it tells us that conscience is not a practical authority in the sense defended by Joseph Raz. It doesn’t work by \textit{excluding} other reasons from practical consideration (Raz (1975)).

\textsuperscript{23} As Butler puts it "It is by this faculty, natural to man, that he is a moral agent" (\textit{Fifteen Sermons}, Section Two).
merit of their actions or the propriety of those actions; both of these evaluations are a sort of moral judgement. Judgements of propriety are judgements about what makes people do what they do; "the sentiment or affection of the heart from which their action proceeds."24

Judgements of merit, on the other hand, concern the outcome of an agent's action; whether it is good or bad. According to Smith, we judge the propriety of someone's action by putting ourselves in her place and asking whether we would respond as she does. If we would respond as she would, on noticing this we approve of her action and judge it to be proper. If we wouldn't, we disapprove of her action and, noticing this, judge it to be improper. In order to judge the merit of someone's action we ask who would be affected by that action and put ourselves in the place of that individual or individuals and ask whether we would have gratitude or resentment towards the perpetrator of the act. If we feel gratitude we approve of that action and judge the person who performed that action praiseworthy. If we feel resentment, then we approve of their being resentful and see the cause of that action as being blameworthy.

Smith notices a serious problem with the account of sympathy that underlies his analysis of judgements of propriety and merit. Sometimes people approve of things and disapprove of things when, intuitively, we would judge that they ought not to. Consequently, it would be unreliable of us to place too much emphasis on what, as a matter of fact, people would or would not approve of. The criteria for what counts as proper and meritorious judgements needs to be idealised to some extent. Smith proposes this: we rule out all those either (a) misunderstand their circumstances or, (b) have a personal stake in what has happened or is going to happen. In other words, we restrict ourselves to only considering the reactions of those who are suitably informed and impartial. Smith calls such an individual the ‘impartial spectator’.25

So much for Smith’s general account of sympathy, how does conscience fit into his moral psychology? Smith holds judgements of conscience are a subset of our propriety judgements. The sub-set of propriety judgements that the conscience restricts itself to, are those

24 *Theory of Moral Sentiments*, II.i.3.5
judgements assessing the agent's own conduct. My conscience tells me only what it would be proper for me to do, not what it would be proper for you or anyone else to do.26

Consider the following two passages:

We begin, upon this account, to examine our own passions and conduct, and to consider how these must appear to them. . . . We suppose ourselves the spectators of our own behaviour, and endeavour to imagine what effect it would, in this light, produce upon us. This is the only looking glass by which we can, in some measure, with the eyes of other people, scrutinize the propriety of our own conduct.27

When I endeavour to examine my own conduct . . . it is evident that . . . I divide myself, as it were, into two persons; and that I, the examiner and judge, represent a different character from that other I, the person whose conduct is examined into and judged of. The first is the spectator . . . The second is the agent.28

In a judgement of conscience I imagine myself as a suitably idealised spectator of my actions (whether actions future or past) and ask whether, as spectator, I would act in that way. If I would not, then my conscience forbids that action. If I would, then conscience permits that action. D. D. Raphael in summarising Smith’s account of conscience says: “The voice of conscience reflects what I imagine that I, with all my knowledge of the situation, would feel if I were a spectator instead of an agent”.29

The most notable feature of Smith’s account (for my purposes, at least) is the way in which Smith divides the individual agent into several distinct ‘persons’. The individual makes her judgements of conscience by ‘standing’ as a spectator on her actions and asking how she would feel. This idea of the individual being ‘split’ into several different entities is central to the accounts of conscience defended by Kant and Heidegger which I go on to describe in this

26 Here we see perhaps one of the earliest statements of the first-personal constraint on conscience. I mentioned in the introduction that this constraint was defended by Ryle (1940). In chapter three I argue that Brownlee is wrong to simply identify conscientious convictions with moral convictions for this reason; conscientious convictions can only potentially be a sub-set of our moral convictions.

27 Theory of Moral Sentiments iii. 1. 5.

28 Theory of Moral Sentiments iii. 1. 6.

chapter and in the next. I suggested earlier that there are elements of *conscience as identity* in Smith’s account. Recall that the hallmark of conscience according to *conscience as identity* is that it involves the individual judging whether or not some past or future action is or would be ‘true to self’. In some sense, Smith’s account of conscience does require that the individual be ‘true to self’ – it is just that the self in question is an ‘idealised spectator’ self. The ‘pure’ *conscience as identity* accounts that I consider in chapter two differ from Smith’s account in the way that they completely divorce conscience from moral agency, though this is not to say that the conscience, for those authors is not of considerable normative, and perhaps moral, significance.

1.3. Kant and the 'Inner Court' of Conscience

For both Butler and Smith, conscience was essentially reflective; a capacity within us by which we reflect on our actions and assess their morality. For both authors conscience was not the sole capacity of moral agency. In both cases their moral psychology was supplemented by other capacities (such as ‘synteresis’ for Butler and judgements of ‘merit’ for Smith). In Kant’s moral psychology, a similarly reflective capacity also plays a central role. Though for Kant, importantly, this reflective capacity does not go by the name ‘conscience’ but, rather, ‘Practical Reason’. Practical Reason, the most important component of Kant’s moral psychology, is our capacity to review the facts and reflect on the considerations that favour one moral conclusion rather than another. In particular, Practical Reason reflects on whether the agent’s actions are consistent with the categorical imperative. One way of describing this reflection is that it is reflection on whether the maxim underlying some action can be consistently willed as a universal law.\(^{30}\)

Practical Reason for Kant, just as conscience, for both Butler and Smith, concerns itself with first-order moral judgements; judgements that pertain to what the agent morally ought to do. Conscience, for Kant, concerns itself with *second-order* judgements only; it stands not in judgement of actions but stands rather in judgement of Practical Reason itself.

---

\(^{30}\) Kant famously gives several different formulations of the categorical imperative, but it is this universal law formulation that is often thought to provide the individual with a practical decision procedure (Guyer (2000), p161).
Kant's most illuminating discussion of the conscience occurs in two separate sections of the *Metaphysics of Morals (MM)*.\(^{31}\) In the first place, he discusses conscience as one of the four feelings that an individual must be able to experience if she is to be susceptible to duty.\(^{32}\) Kant thinks that, along with certain rational capacities, an agent must possess certain affective capacities in order to count as a moral agent. Conscience is one of these capacities. His second discussion of conscience in *MM* comes under the heading of ‘On a Human Being’s Duty to Himself as His Own Innate Judge’.\(^{33}\) Here conscience is thought of as a reflective capacity by which the agent reflects on whether she has exercised due diligence in her use of Practical Reason.\(^{34}\) We have a duty to exercise our conscience, according to Kant, as it is only by doing so that we can be justified in thinking that we have made the correct moral judgements. It is not immediately clear exactly how the feeling/capacity analyses of conscience are to be reconciled, but I will argue shortly that the ‘feeling’ of conscience can be interpreted as the natural end result of exercising the reflective capacity of conscience.

When Kant discusses the feeling of conscience he seeks to describe how and why the feelings that we associate with moral thought and action are central to our moral psychology. In doing so he shows how his moral psychology can be reconciled with the emphasis that some other authors (such as Smith) put on moral feelings. In doing this, his discussion aims to correct a common misunderstanding of Kant as a theorist who divorces morality from feeling or sentiment. The claim of moral sentimentalists which Kant wishes to reject is not that feelings can influence moral action but rather that a *certain class* of feelings are involved in moral action. Kant wishes to deny that moral worth is attached to actions that are motivated by *empirical* feelings (such as sympathy). Empirical feelings are ones that arise passively through being receptive to empirical impulses. For example, perhaps we naturally feel sympathetic when we are asked for money by the homeless. For Kant, that feeling ought not to influence your assessment as to whether you ought to give to the homeless. The feelings involved in moral action, for Kant, are not the empirical feelings that might cause certain

\(^{31}\) Kant also defends this notion of conscience in *Critique of Practical Reason* ((1997), p82, 5:78 – 5:79)) and *Religion within the Boundaries of Mere Reason* ((1998), p179, 6:186). 

\(^{32}\) (1996), p159, 6:399.


\(^{34}\) Thomas Hill, in his analysis of Kant on conscience, suggests that the conscience also has another function; it assesses whether or not the agent has contravened her own judgement as to what she believes to be right. ((1998a), p 16).
action but rather the rational feelings that are a consequence of proper moral thought and action. If we weren’t susceptible to these feelings, not only would we not be moral, we would be incapable of being a moral agent. Kant makes this point when he talks of feelings as “subjective conditions of receptiveness to the concept of duty, not as objective conditions of morality”\(^{35}\)

For Kant there are four types of rational feeling. The first is moral feeling. This is the agent’s susceptibility to feel pleasure or displeasure as a result of her awareness that she has acted either consistently with or contrary to the requirements of duty. This notion is similar to that traditional notion of conscience identified by Jerome as an agent’s feeling of guilt attendant on the perceived recognition of wrongdoing. It is moral feeling that makes us susceptible to be moved by “pure practical reason”. Why does Kant think this? Why does he think that such a feeling is necessary in order to be moved by considerations of duty? The reason is, I think, that this is often the way in which we are consciously aware of whether we have done as we ought. It might be supposed that we would not need such a capacity if we were always automatically motivated by duty. But, importantly for Kant, moral thought and action is impossible without genuine free will. If an individual were automatically motivated in that way she would not be acting freely and consequently her actions would not count as rational and thus moral.\(^{36}\) Therefore, continued moral thought and action requires that the individual is susceptible to this feeling.

Kant bridges his discussion of moral feeling with that of conscience by remarking that conscience, as well, is not something that we have a duty to acquire, but is something, rather, that we naturally have and must have in order for us to be susceptible to duty at all. Conscience is not the feeling that makes us aware of our own wrongdoing, rather it is a feeling of either pleasure or displeasure attendant on the recognition that we have not been sufficiently diligent in the exercise of Practical Reason. The reason offered by Kant as to why we must be susceptible to this feeling is similar to the justification given in the moral feeling

\(^{35}\)(1996), p159. 6:399.

\(^{36}\) The reason that moral requirements must be adhered to freely rather than automatically (say through a natural inclination to do the right thing) is that moral requirements are requirements of Practical Reason. And an agent that judges in accordance with Practical Reason, must judge freely. Consider the following claim of Kant’s:

Reason must view herself as the authoress of her principles, independent of alien influences, and must consequently, as practical reason, or as the will of a rational being, by herself be viewed as free (2013, p58, 4:447-4:448).
case. Given that we sometimes do wrong, and can easily get into a habit of doing so, we need 
some sort of inbuilt mechanism by which we continually question our exercise of that 
capacity pertaining to our own actions.

Intuitively, such a feeling does not seem necessary in the same way that moral feeling might 
be. It seems possible for us to be moral agents who are always diligent in our exercise of 
Practical Reason. However, given our natural inclination to self-deception – our tendency to 
be too light or too heavy on ourselves- conscience does seem to be an essential check. Kant 
goes on to suggest that talk of someone ‘not having a conscience’ is purely metaphorical. 
What we mean when we talk in this way is that the person in question pays no heed to the 
verdict of conscience.

The next section of MM where Kant discusses conscience is in Part 1 of the Doctrine of the 
Elements of Ethics under the heading ‘On Duties to Oneself as Such’. Kant is interested in 
the duty that is involved in the operation of conscience. Conscience involves a duty owed to 
one self and one that is owed in virtue of one’s being a Moral Being. Here Kant describes 
conscience not as a feeling per se but as the “consciousness of an internal court in man”. Here I take it that Kant does not mean consciousness of all the operations of this internal 
court, if it did then this would be a clearly implausible account of conscience. Presumably we 
do not explicitly engage in the peculiar process that Kant describes. It is more plausible that 
the consciousness involved is consciousness of the verdict of that court.

Kant’s internal court of conscience is a psychological process, different aspects of which 
Kant represents figuratively with the officers of a conventional law court. 
The prosecutor lays the charge against the agent and argues that she has not done her duty. 
A defence counsel argues on behalf of the agent’s self-love, responding to the prosecutor. A 
Judge makes the final determination and gives a verdict of acquittal or condemnation. What 
exactly is the agent charged with in this process of conscience? Recall that the conscience 
does not charge that the agent has failed to act morally per se.

Consider:

Conscience does not pass judgement upon actions as cases that stand under the law, for this is what reason does insofar as it is subjectively practical. Rather, here reason judges itself, whether it has actually undertaken, with all diligence, that examination of actions (whether they are right or wrong), and it calls upon the human being himself to witness for or against himself whether this has taken place or not.\(^39\)

Using modern parlance, we could say that an act of Practical Reason and the accompanying feeling, moral feeling involves the agent assessing herself against a moral norm. If the agent judges that she has acted wrongly, she feels guilty. If the agent judges that she has acted rightly, she feels satisfied. Conscience, as distinct from Practical Reason, assesses the agent according to what we would in modern parlance call an epistemic norm. It asks whether the agent’s moral judgements have been justified rather than correct. There are several ways in which we could interpret the agent as failing in this regard. The most straightforward case would be a misunderstanding or misapplication of the categorical imperative. For example, the agent might mistakenly judge that she can will a universal law whereby an individual can break a promise as long as breaking that promise would be to her financial advantage. Another way in which the individual could fail is that the individual’s understanding of morality is correct, but her understanding of the facts is not. For example, the agent might acknowledge that it is wrong to break a promise to repay someone for money that has been lent, but mistakenly believe that the lender has forgiven the debt and neglect to pay on that basis. In either of these cases, the conscience does not determine whether or not the agent got things wrong. The conscience is concerned, rather, with whether or not the agent has been sufficiently scrupulous in her exercise of Practical Reason, whether in her original commitment to general moral principles or the application of those general principles in particular cases.

To summarise Kant on conscience: The conscience for Kant, in contrast to, say, Butler’s account, is not intended to explain any part of our general capacity to assess actions for their rightness and wrongness. Kant describes the conscience in two apparently distinct ways. The first way in which Kant describes conscience is as a certain type of rational feeling. The second way of describing it is as a meta-capacity which checks whether we have taken due diligence in our application of Practical Reason. This meta-capacity is modelled by Kant as an internal court whereby the individual stands as prosecutor, defence counsel and judge in

her own case. The two different ways of describing conscience can be reconciled because the final verdict of the internal ‘court’ of conscience is a feeling of either guilt or a clear conscience.

We are not consciously aware of the operations of the ‘internal court’ of conscience. Often the only aspect of conscience that we will be consciously aware of is the feeling of conscience. Given that we are not aware of the operation of this internal court, why does Kant think that we must have such a capacity? I think Kant’s view is the following: Moral agency, for Kant, requires that we are not automatons. It requires that we don’t automatically do the ‘right’ thing all the time. Moral agency requires, rather, that we are beings ‘free’ to deliberate on and decide for ourselves what we ought to do. Given this, if we are going to rely on our moral agency we must possess some capacity to regularly check on the operations of our Practical Reason to ensure that we are justified in following its deliveries. This guards against the natural tendencies that we have to be either too lenient or too harsh on ourselves.

Kant’s conscience is still an instance of the conscience as moral agency approach as it is a capacity that, according to Kant, along with the faculty of Practical Reason and other capacities of the individual (such as the capacity to have certain feelings) is partly constitutive of the individual’s moral agency. However, as was also the case with Smith’s account, we see elements of the conscience as identity approach in Kantian conscience. As in Smith’s account, ‘person-splitting’ is present in Kant’s account. In the ‘court of conscience’, the individual is split into several different ‘persons’ that fulfil a role. Furthermore, the object of the conscience is not the actions of the individual that are assessed for their morality but the actions of the individual. As the individual stands in judgement of herself, in following conscience we might say that the individual is being true to self. The next account of conscience (that I describe in chapter two), while very squarely a conscience as identity account, relies on several aspects of Kantian moral psychology.

In this chapter I have described three accounts of what is, in my, opinion, the dominant approach to conscience, even today; conscience as moral agency. I have intentionally chosen these three historical figures as I think that each account tells us something important about the moral agency approach. The first conception I consider, Butler’s, is important because it is a paradigmatic account of conscience as moral agency. Butler argues that conscience, while it is not the only capacity that constitutes our moral agency, it is the supreme capacity in determining what the individual morally ought to do. The second conception shared with
Butler’s account the idea that conscience is a capacity of moral judgement. But according to Smith conscience only picked out a certain subset of our moral judgements. It picked out only those judgements of the agent that are concerned with what *she herself* ought to do. In Smith’s account like in many contemporary accounts of conscience, moral feelings or emotions play a particularly important role in conscience. In the third account of conscience I considered, Kant’s account, conscience did not involve a moral judgement *per se*, but was a capacity that stood in judgement of the individual’s moral judgements. Nevertheless, Kant’s account is still a *conscience as moral agency* account of conscience as it is, according to Kant, an essential element of our moral agency. Without a conscience we would be unable to rely on our Practical Reason to lead us towards acting morally. But importantly, the accounts of both Smith and Kant also involve elements of the *conscience as identity* approach. In both cases conscience involves the individual standing in judgement of herself and attempting to arrive at a view which is, in some sense, ‘true to self’.
Chapter Two: Conscience as Identity

In the introduction to this thesis I described Conscience as identity as the view that to think and act in accordance with conscience is a matter of being ‘true to self’. I suggested that the main rival to this approach is conscience as moral agency.1 In Chapter One we saw elements of this approach in the accounts of conscience defended by Adam Smith and Immanuel Kant. On Smith’s account, acting in accordance with my conscience is a matter of ‘being true’ to some version of myself; conscience is a matter of feeling the way that a suitably idealised version of myself would feel. For Kant, Conscience is a matter of ‘being true to oneself’ insofar as the feeling of conscience is a sentence passed by the self when wearing the ‘hat’ of the Judge in the individual’s own internal ‘court’ of conscience.

Given that elements of the conscience as identity approach are present in the conscience as moral agency accounts just mentioned, are these two approaches really distinct? Consider again the definition of those two approaches. Conscience as moral agency holds that conscience is an element (if not the totality) of our moral agency. For Butler, conscience is our all-things-considered moral judgement. For Smith, the conscience refers to a certain subset of our moral judgements. For Kant, conscience is a capacity to make moral judgements, though a very peculiar sub-set of those judgements.2 It is the capacity to make judgements about whether we have succeeded or failed in our moral duty to ourselves to scrutinise the operation of our Practical Reason and to know ourselves. Accounts of conscience as identity, ...

---

1 Though this is not the only rival approach. In chapter three I discuss Kimberley Brownlee’s communicative approach to conscience. While this approach could be seen as an instance of the conscience as moral agency approach, I shall argue in that chapter that it does not fit neatly within that category.

In “Four conceptions of conscience”, Thomas Hill describes (but does not endorse) an account of conscience he calls the “extreme cultural relativist” (ECR) conception of conscience. According to this view, conscience is simply the internalisation of cultural norms that serve to promote social cohesion, whether or not these norms are genuinely moral norms ((1998), p 22). This account, too, is not in any obvious way, an instance of either approach.

2 At least, they are a sub-set of moral judgements insofar as Kant identifies the categorical with the moral. Plausibly, one could deny this identification.
by contrast, deny that the conscience is some component or the totality of our moral agency.\(^3\) This negative claim is, of course, not sufficient for some account to count as an instance of conscience as identity.\(^4\) Conscience as identity needs some positive account of what it means exactly to have a conviction, or to act in a way that is ‘true to self’.

In this chapter I describe two particular accounts of the conscience as identity approach which both give quite a different account of what it is for an individual to act ‘true to self’. I look at the accounts of Martin Heidegger and Hannah Arendt. Why do I focus on these two accounts specifically? In the first place, they are, perhaps, the two most prominent philosophers of the twentieth century for whom conscience was an integral component of their broader philosophy.\(^5\) Any attempt to be thorough in a discussion of recent philosophy of conscience would be remiss to ignore their accounts. Second, both the accounts are helpful in understanding the account of conscience that I go on to defend. I argue that there are three principles of conscience with the identity principle being one of the most important of these principles. There are some important similarities between the account of identity that I defend and Arendt’s account in particular.

In section one, I begin my discussion of Heidegger’s account of conscience. I argue that, despite the way it may at first appear, Heidegger’s account of conscience is not intended to be

---

\(^3\) Accounts of conscience as identity can, however, hold that the conscience, while not a component of our moral agency, is of huge moral significance. In my own account, thoughts and actions which implicate conscientious identity are of moral significance because they capture what is valuable in a right of conscience (through the three heads of normativity).

\(^4\) In the introduction I mentioned Martha Nussbaum’s account of conscience which makes the same negative claim but is not, on the face of it, an instance of conscience as identity. According to Nussbaum the conscience is the faculty with which the individual that searches for meaning. Although the account of conscientious identity that I defend in chapter seven does incorporate aspects of Nussbaum’s account into a conscience as identity framework. I argue that an individual’s narrative self-conception is constituted by her beliefs about which of her ideals make her life meaningful.

\(^5\) Another historical figure who seems to endorse something like a conscience as identity approach is Hegel. For example:

Conscience … is that deepest inner solitude within oneself in which all externals and all limitation have disappeared – it is a total withdrawal into the self ((1991/1820), pp163–4).

There certainly seem to be strong affinities between that account and Heidegger’s, however, in my view Hegel’s account is a more problematic exemplar of the conscience as identity approach as his account also seems to be an account of moral agency.

For example:

Conscience expresses the absolute entitlement of subjective self-consciousness to know in itself and from itself what right and duty are, and to recognise only what it thus knows as the good; it also consists in the assertion that what it thus knows and wills is truly right and duty ((1991/1820), p164).
a rival account to moral agency conceptions of conscience. It is not even an attempt (let alone a successful attempt) to capture that capacity implicated in our practical psychology and in rights of conscience.\textsuperscript{6} Heidegger is quite explicit that he is not trying to capture an ‘ordinary interpretation’ of conscience. Rather, his account is an attempt to explicate existential conscience, which, I argue, is supposed to serve two functions within his broader philosophy.

The first function of conscience for Heidegger is to explain how it is that humans can be separate from their social world. The second function, presumably inspired by Kant, is to explain how it is that humans can be bound by norms. In section one I briefly discuss Heidegger’s project in \textit{Being and Time} in order to provide the necessary background for understanding Heideggerian conscience. I discuss Heidegger’s phenomenological method and define the key Heideggerian terms involved in his analysis include ‘Dasein’, ‘das Man’, ‘authenticity’ and ‘guilt’. Then in section two I describe Heidegger’s account of existential conscience which analyses the notion of conscience in terms of those previously defined notions. Briefly, and in language that might be opaque for the moment, conscience consists in the individual in one guise calling on herself in another guise to be ‘authentic’; a call to recognise for herself that she is not really free and is, rather, ‘thrown’ into her social world. I then outline Heidegger’s argument that it is this function of conscience that individuates humans from their broader social world. In section three, I discuss Rebecca Kukla’s argument that Heidegger’s account of conscience is also intended to ground normativity in \textit{Being and Time}. I argue that existential conscience is not an account of ‘ordinary’ conscience – it is not intended to be a rival account to the accounts of conscience as moral agency discussed in chapter one. But it is still morally relevant in the sense that that concept is a necessary prerequisite to the existence of moral norms.

Furthermore, Heidegger’s obscure remarks about the connection between his existential conscience and the more commonplace conscience that we are all familiar with suggests a different way of thinking about ordinary conscience, a different way of thinking about conscience that is a potential rival to the moral agency conceptions. In section four I consider Hannah Arendt’s account of conscience which I consider to be an example of such an account. For Arendt, conscience is a by-product or ‘side effect’ of thinking. And because

\textsuperscript{6} In the introduction to this thesis I specified this as the goal of the accounts of conscience in this thesis that I am interested in. The goal is to give an account of what is protected in a moral right of conscience and then, secondarily, explain that element of our practical psychology that we call ‘conscience’.
thinking involves ‘talking to ourselves’, conscience is an appeal to identity. In the words of Larry May, Arendtian conscience involves “the thoughtful experience of bringing before the mind one’s projects and judging them vis-à-vis the consistency and integrity of self.”

According to Arendt, in conscience, we don't appeal directly to whether some action would be right or wrong but rather to whether or not we can be the sort of person who commits that action. The agent is working out what action would be ‘true to self” and in that sense her account is a paradigmatic conception of the conscience as identity approach. I propose that Arendt’s account is an example of the revised account of ordinary conscience envisioned by Heidegger in his analysis. The identity principle that I defend in chapter seven, conscience as narrative self-conception, has some affinities with this conception of Arendt’s.

2.1. Heidegger and the Project of Being and Time

On the face of it, Heidegger is an unusual figure to contrast with other authors (such as Butler, Smith and Kant) on matters of ethics or moral psychology. Being and Time is generally thought to have little to say about morality and the psychological capacities associated with morality. Although Heidegger repeatedly uses the term ‘conscience’ in Being and Time, it doesn't necessarily follow that he is referring to the moral conscience or even an account that can be usefully compared with such a notion. Some descriptions of Heideggerian conscience might give the impression that the conscience for Heidegger is a radically individualistic variation on conscience as moral agency, where conscience is not the capacity to make moral evaluations but the capacity to ‘follow one’s heart’ or ‘be one’s self’.

Consider, for example, John Skorupski:

---

7 (1983), p57.

8 Heidegger’s project in Being and Time is a project in fundamental ontology. Ethics, on the other hand, is what Heidegger refers to as an ‘ontic’ domain which presupposes the sort of investigation he is carrying out in Being and Time. Consider, for example, Stephen Mulhall on this point:

as we define “ethics” as a theoretical study of one aspect of human beings and their common life, it counts as what Heidegger calls an ontic science; like anthropology or biology, it aims to construct a body of knowledge about a distinct domain of reality (2010), p241).

9 Or, at least, that is how the term Heidegger uses ‘gewissen’ is always translated.
Again conscience is withdrawal into the self, this time from “lostness in the ‘they’” – that is, the pressures and expectations of the crowd.\(^{10}\)

Or the following description offered by Allen Wood:

> In this respect, Kant’s conception of conscience stands in striking contrast, for example, to Heidegger’s in Being and Time. For Heidegger, the call of conscience is not articulate, or properly articulable, but consists in a discourse of silence. Conscience for Heidegger is Dasein calling itself back out of the public realm (which Heidegger dismisses as the realm of das Man) to its own uncanny authenticity, which recognizes no public standards.\(^{11}\)

I think there is an easy way of interpreting these passages (and thus Heidegger) as saying something like the following: The conscience calls the individual *not* to act simply in the way the rules and conventions of society ask of her but to act in a completely free and authentic way. The conscience is seen as something like the *opposite* of Thomas Hill’s ‘extreme cultural relativist’ (ECR) conception of conscience.\(^{12}\)

For an even stronger case, consider Schalow:

> The call of conscience is radically individuating. It makes its appeal by circumventing Dasein’s social involvements and yet returns the self to a stance where it first becomes disposed to act in the situation Dasein’s social involvements and yet returns the self to a stance where it first becomes disposed to act in the situation.\(^{13}\)

In this passage this interpretation seems even more obvious. To act according to conscience is to act in an individualistic way, rather than simply going along with the expectations of the social world. This interpretation, however, is incorrect. Heidegger's project in *Being and Time* is primarily what he calls an ‘ontological’ inquiry but what in modern parlance we might call *meta-ontological*. His aim is to work out what it means, in the most general sense, for something for something (anything) to be. And according to Heidegger, the way to work this out is through a careful investigation of the human way of life, or as Heidegger refers to

---

\(^{10}\)(2010), p556.

\(^{11}\)(Forthcoming), p8.

\(^{12}\)(1997), p21-22. See footnote one of this chapter for a description of that account.

\(^{13}\)(1998), p146.
it 'Dasein' (there-being). One matter, Heidegger notes as central to our way of life - what defines us - is our capacity to ask the ontological question itself; what does it mean to be? And given that it is in the nature of Dasein to be constantly concerned with the meaning and nature of being, Heidegger suggests that Dasein operates with a limited and shallow understanding of being (a 'pre-ontological' understanding). Given that Dasein begins with a pre-ontological grasp of being, Heidegger suggests that in answering the meta-ontological question there is no better place to start than with the activities and engagements of Dasein itself.

For many purposes, we can just treat Dasein as a synonym for the human entity and as equivalent to terms like 'agent', 'self' or 'person' employed by other philosophers. Heidegger insists on this special term in order to emphasise that Dasein is not to be interpreted, at least initially, as something separate from the world around it. The scientific interpretation of humans as biological entities operating in concert with other biological entities in a physical environment does lend support to that classical 'severable' notion of the human. However, Heidegger argues, that scientific notion already assumes a particular understanding of humans. It is this more basic, pre-supposed, understanding that 'Dasein' is used to represent.

What methodology does Heidegger use in his analysis of Dasein? He employs a phenomenological approach. That is, he is not directly investigating our concepts or trying to intuit the 'essence of things' as we might understand many other philosophers to be doing. In the introduction to this thesis I suggested that my methodology is a mixture of normative and descriptive conceptual analysis. I engage in normative conceptual analysis as my starting point is the moral right of conscience; I am interested in the conception of conscience that is presupposed in such a right (if there were to be one). My methodology involves descriptive conceptual analysis as I argue to understand the nature of the principles of conscience, such

---

14 This is Hubert Dreyfus's interpretation of the meaning of 'Dasein' in his influential commentary on the first division of Being and Time (1990). John Haugeland has a different view. He argues that Dasein is the distinct kind of entity that humans are (Haugeland ((2005), p 422).

This scholarly disagreement (as with many disagreements about the meaning of Heideggerian terms) need not concern us here as it has little bearing on whether Heidegger’s account of conscience plausibly involves an appeal to ‘identity’ or ‘self’.


16 I discuss the nature of this right in considerable detail in chapter four. The right we are talking about is not a legal right, but a defeasible moral right that protects an individual’s conscientious actions to a certain degree.
as conscientious identity, we need to look at our intuitions about which sorts of mental acts count as operations of conscience. In light of this, chapter five of this thesis is devoted to discerning the four descriptive requirements of conscientious identity from Victor Hugo’s description of conscience in Les Misérables.

Heidegger’s methodology, by contrast, is phenomenological: He proceeds by a careful analysis of the phenomena of conscience (whatever ‘phenomena’ might turn out to be). Husserl’s phenomenology had identified the phenomena with mental phenomena that we have conscious access to. Phenomenology cannot proceed by interrogating mental phenomena, Heidegger argues, because both mental states and the practices incorporating those states often operate in a way that cover up their ‘true essence’. And it is this underlying ‘essence’ that is the phenomena that we are investigating. The covering up can happen in either one of two key ways. The phenomena can simply be covered up insofar as the phenomena are undiscovered. The actual thought and action of humans as they go about their business in the world can distract us from seeing what is really going on. But the phenomena in question were there all along waiting to be discovered. It is this approach of discovering that is the dominant approach in Division I of Being and Time.

The second way in which the phenomena can be covered up is that phenomena that was once straightforwardly available to us in our everyday practices can be buried over and disguised. This means that “it has at some point been discovered but has deteriorated to the point of being covered up again”. This second phenomenological method, of trying to uncover what is disguised in the everyday world, is crucial in understanding what Heidegger calls ‘conscience’ and what Heidegger calls ‘guilt’ as both ideas are ones that, according to Heidegger, have been buried over and disguised.

This interpretation of conscience, as something that is ‘buried over’ goes a significant way towards explaining why the Heideggerian concept that is named ‘conscience’ is so at odds with any of the ways in which we use the term ‘conscience’ in everyday life and the notion

---

17 Husserl, (1913).
18 (1962), p60.
19 (1962), p60.
20 (1962), p60.
articulated by other theorists. According to Heidegger the ordinary ‘moral conscience’ that we are all familiar with *masks* existential conscience. We shall see shortly Heidegger’s explanation as to why existential conscience would do this. It is worth noting before proceeding any further that, in Heidegger’s view, the status of being ‘buried over’ is not ‘bad’. The fact that existential conscience is covered over is not a state of affairs that needs to be ‘fixed’.

2.2. Dasein, Existential Conscience and Existential Guilt

Now that we have an understanding of Heidegger’s guiding question and his chosen methodology we are in a position to articulate his account of conscience. Heidegger introduces conscience in chapter II, Division II, of *Being and Time* as part of an argument as to how it is possible for Dasein to be *authentic* (or as I frame authenticity, as part of how the individual can be ‘true to self’). For the most part Dasein is caught up in *das Man* or ‘the They’. This is the everyday social world, the public realm, which we all inhabit. Dasein, in this respect is a ‘they-self’. Dasein, when in this mode is carried along by the laws and conventions of the everyday world and does not even realise that it is essentially constrained and without real freedom. In this respect, Dasein is *inauthentic*: “So dasein makes no choices, gets carried along by the nobody, and thus ensnares itself in inauthenticity”.

But Dasein does not have to be like this, it has the potential to act in an *authentic* mode. Heidegger claims that this potentiality for being authentic is attested to by what the folk sometimes refer to as the ‘voice of conscience’.

Before further articulating this mysterious voice, Heidegger forestalls a sceptical response to the notion. There is considerable disagreement as to the nature and origin of this voice. Some argue that it is the voice of God within us. Others argue that it is the Kantian categorical imperative. The sheer degree of theoretical disagreement as to what exactly this voice is might lead us to doubt whether there is, in fact, a voice with a discernible nature in operation. Heidegger claims that inference would be a mistake. These different interpretations are ‘perversions’ and simply result from

---

22 (1962), p312.

23 (1962), p313.

24 In the introductory chapter I mentioned Velleman’s recent argument that the conscience is, in fact, an internal voice communicating the categorical imperative to the individual (1999).
the fact that ‘ordinary conscience’, the one that is implicated in our everyday thought, is constantly fighting against its true nature in various different ways.

Heidegger characterises the voice of conscience, as a *call*. Conscience calls Dasein out of the realm of ‘the They’ and out of its ‘they-self’ mode. What exactly is Dasein called to be?

> the call of conscience has a character of an appeal to Dasein by calling it to its ownmost potentiality – for-Being-its-self; and this is done by way of summoning it to its own most being-guilty.\(^ {25} \)

Conscience calls the agent to be herself, and being herself involves recognising that she is fundamentally ‘guilty’. We have already seen in this thesis that it is common for theorists to make a conceptual connection between the notions of conscience and guilt. For example, the displeasure from the verdict of condemnation in Kant’s conception of conscience is what we might ordinarily call guilt.\(^ {26} \) According to psychological and philosophical discussion of psychopaths, the psychopath’s lack of conscience is evidenced in her inability to feel guilty when a normal person would do so.\(^ {27} \)

As mentioned earlier, Heidegger’s conception of guilt here is ‘existential’ - it is not supposed to be identified with the ordinary feeling of moral guilt referred to by those other theorists. I will come to an analysis of this notion of guilt shortly. For now let’s consider the nature of this *call*. Who is doing this calling? In the ordinary interpretations of conscience I have considered thus far I have considered the possibility that the voice present in conscience is variously the voice of God, or perhaps the super-ego or the categorical imperative. But these interpretations are not available to Dasein. Those notions are not what Heidegger calls ‘ontological’ notions. Heidegger posits that the ‘caller’ in conscience, is, and must be, Dasein itself.

How can we make sense of Dasein calling itself? Recall that the Kantian conception of conscience (discussed in section three, chapter one) requires that, in some sense several

\(^ {25} \) (1962), p314.

\(^ {26} \) As I mentioned in the introductory chapter, Nietzsche and Freud also interpreted guilt as the most important phenomenon of conscience. For both of those authors the conscience was analysed as the individual’s imposition of guilt upon herself as an outlet for natural aggressive impulses that had no other outlet.

\(^ {27} \) See introduction, section one.
different persons are present in the human agent in order to put the agent on trial; there is the person <as agent>, person <as prosecutor>, person <as Judge> etc. I also suggested in that chapter that Adam Smith’s account also involved such ‘person-splitting’ as conscience involved the individual imagining how some idealised version of herself would feel in certain circumstances. Heidegger adopts a similar approach. Heidegger doesn’t speak of several separate individuals within the one individual, but rather of several different ways in which Dasein can be. What is the caller’s way of being?

The caller is Dasein in its uncanniness: primordial, thrown Being-in-the-world as the “not-at-home” – the bare “that-it-is” in the “nothing” of the world. Uncanniness for Heidegger is the most philosophically significant of moods or emotions, though it is controversial whether this should be identified with any affective state or whether, like Heideggerian conscience and guilt, it is an ontological concept. Uncanniness (in German, unheimlichkeit- translated literally as ‘unhomeliness’) is the realisation of Dasein that while it has been ensconced within ‘the They’ it is not ‘at home’ there. It is Dasein, in the mode of the ‘they-self’, that realises that it is not completely determined by ‘the They’, by the social world. This is the caller in conscience.

Heidegger referred to a voice of conscience, so what does the conscience say? According to Heidegger it says nothing. This seems a strange way to characterise a ‘voice’, though recall that even in the Kantian conception of conscience, for example, not all aspects of the operation of conscience are consciously experienced. In Kant’s case it may be that the only conscious manifestation of conscience is the feeling of satisfaction or displeasure that is the result of its verdicts. Similarly, it may be that the only conscious manifestation of the voice of conscience for Heidegger is that unsettling feeling that Dasein has that it is not at home in the world. Though we may sometimes have an experience of a voice giving us specific instructions, this is still with a voice issuing from ‘the They’ and is therefore not genuine conscience. That voice could not be the voice of existential conscience.


Or, if uncanniness is not an actual affective state, there may be no empirical trace of the existence of Heideggerian conscience.
Now that we have an understanding of whom the call of conscience issues from and issues to we need to account for the content of that call. In the quotation I referred to above the conscience is supposed to summon Dasein back to its own most being—guilty. What does Heidegger mean when he talks of 'being-guilty'? Heidegger claims that a connection between conscience and guilt is contained in nearly every ordinary interpretation of the conscience. Recall that in the case of conscience, Heidegger’s starting point was the notion of a ‘voice’ of conscience, something that is present in the everyday interpretation of conscience (although, Heidegger then quickly departed from the everyday interpretation). It is the same for guilt. Heidegger begins with some of the everyday features of guilt-talk. Two important senses of guilt that he emphasises are the sense of owing/indebtedness and responsibility. Someone is guilty only when they have failed in some perceived obligation; when they feel that they owe something. Guilt also involves the idea of responsibility, according to Heidegger. To feel guilty for something is to feel that you are responsible for some action that you have failed to perform. One does not feel guilty (or shouldn’t) for something that is completely out of one’s control.

But the everyday notion of guilt will only get us so far, Heidegger claims. Being guilty in the everyday sense, involves judging that one has not done what one ought to. That the agent’s behaviour is lacking in some way. Heidegger provides a rather obscure argument that Dasein can never be characterised as lacking as we can’t make sense of ourselves as beings that are fundamentally not something. If guilt is to be an ontological notion then it is something that dasein possesses necessarily. But if guilt refers to an acknowledgement that we ought to have done something else, then this implies that it is possible for us to have acted in a way such that we were not guilty. Existential guilt must refer to something that we always are.

In Life of the Mind, Hannah Arendt puts the point in the following way:

The main point in Heidegger’s ‘idea of guilt’ is that human existence is guilty to the extent that it ‘factually exists’; it does not ‘need to become guilty of something through omissions or commissions; [it is only called upon] to actualize authentically the “guiltiness” which it is anyhow’. . . . The concept of ‘being thrown into the world’ already implies that human existence owes its existence to something that it is not itself; by virtue of its very existence it

---

31 (1962), p327.

32 (1962), p326.

33 Ibid., p329.
is indebted: Dasein – human existence inasmuch as it is – ‘has been thrown; it is there, but not brought into the there (“da”) by itself’.

And what we always are, according to Heidegger, is acting to close off other possible actions: “in having a potentiality-for-being it always stands in one possibility or another: it constantly is not other possibilities”. Dasein is fundamentally indebted to itself because, in acting, it has closed off possibilities for action. It is always an opportunity cost of doing one thing that we do not do another. The call of conscience is a call to recognise that in pursuing one action we rule out other actions.

Given that conscience is an ontological notion, it is something that humans necessarily have. However, what we can choose, according to Heidegger is wanting to have a conscience. “What is chosen is having a conscience as Being-free for one’s ownmost Being-guilty”. What exactly does this mean? What it means, I think, is that while we are not free to have a conscience, we are free in whether we recognise what the conscience is telling us. And it is this understanding that allows us to see how the existential conscience is supposed to play a meta-normative role; it explains how it is that we are capable of being bound by norms (this point is the subject of the next section).

Heidegger says that “the appeal calls back by calling forth: it calls Dasein back to its thrownness so as to understand this thrownness as the null basis which it has to take up in existence”. ‘Thrownness' for Heidegger is another way of characterising our absorption within 'the They'. It is not that conscience calls us to literally be something different and independent of our social world: We can't help but be absorbed within that world. However, it is open to us to recognise this 'thrownness'. Dreyfus accepts that while Heidegger talks of possibilities of the self as if Dasein can have possibilities of its own, this cannot be what Heidegger intends.

---

37 As I shall suggest in the next section, such an argument is offered by Rebecca Kukla.
38 (1962), p333.
But this talk of possibilities of its self cannot mean Dasein's "individual concrete content" or defining commitment. Rather it must mean that authentic Dasein projects public possibilities in a way that reveals rather than covers up what it is to be a self.\textsuperscript{39}

Authenticity on Dreyfus' interpretation is not a matter of acting independently of 'the They'. It is simply acting with a realisation that one is 'thrown' in with 'the They'. There is no option but to act from one's own possibilities.

2.3. Existential Conscience and Morality

Thus far I have been focusing on the ontological purport of conscience in Heidegger's project: conscience is there to explain the (limited) sense in which the agent can be individuated and seen as independent from 'the They'. It is a fairly plausible requirement of any fundamental metaphysics such as Heidegger's to be able to explain the sense in which we are all 'individuals' distinct from other aspects of our world. Now I want to move on to a consideration of the meta-normative purpose of conscience for Heidegger. I mentioned in the previous section that while it is not open to us to choose whether we have a conscience, and whether or not we are therefore ‘free’, it is open to us to want to have a conscience. And it is the freedom to do this that opens up the possibility for conscience grounding normativity.

Heidegger makes somewhat obscure reference to the meta-normative role of conscience when he says "wanting to have a conscience is rather the most primordial existentiell presupposition for the possibility of factically coming to owe something".\textsuperscript{40} What does this mean? Several times thus far I have described several Heidegerrian concepts using the term ‘existential’. Here Heidegger uses the term ‘existentiell’. An existentiell understanding of something is an understanding of the ontic domain; understanding things in terms of their existence.\textsuperscript{41} An existential understanding is the one that is presupposed by that domain. So when Heidegger is talks about wanting to have a conscience being an existentiell presupposition he means that it is a presupposition in the ordinary ‘ontic’ domain. He means that it is a presupposition to owing someone something in the ordinary sense. So here the

\textsuperscript{39} Dreyfus (1991), p312-313.
\textsuperscript{40} (1962), p288.
\textsuperscript{41} (1962), p33.
suggestion is that the notion of ‘wanting to have a conscience’ is a pre-requisite to the actual existence of a certain normative obligation.

At another point Heidegger makes the more general observation that some ontological investigation of Dasein must underlie our normative or evaluative notions:

> Even the theory of value, whether it is regarded formally or materially, has as its unexpressed ontological presupposition a ‘metaphysic of morals’ - that is, an ontology of dasein and existence.\(^{42}\)

Rebecca Kukla articulates an argument as to what this connection between existential conscience and normativity is supposed to look like. It is an argument that supposes Heidegger as mirroring Kant's famous argument in the *Groundwork for the Metaphysics of Morals* that freedom is a transcendental pre-condition to moral agency.\(^{43}\)

We have established already that Dasein is carried along by ‘the They’ and is thus ‘fallen’. However, if the individual is merely carried along by the rules of ‘the They’ then the individual could not be genuinely normatively responsive. The individual would be following rules as if they were ‘laws of nature’, acting on blind impulse, and that is not what it is to be normatively responsive. The legitimacy of a norm must be something that I can, at the least, call into question.

In Kukla’s view, it is the individuated, authentic Dasein (as defined in the previous section) that grounds normativity. Kukla argues that Dasein in its guise as the ‘they-self’ cannot be normatively responsive. As a ‘They self’, Dasein is just carried along by laws and rules of the world as if it were an automaton. Genuine normative responsiveness requires *taking responsibility* with regard to the acceptance of the norm in question. It is not about actually acting independently of the ‘They self’ (that is impossible), but rather with a realisation that one is the cause of one's own actions. It is not a matter of having possibilities of acting that are independent of das Man (‘the they’). It is a matter of acting on the recognition facilitated by conscience. It is the *wanting* to have a conscience that is a pre-condition to normativity.

\(^{42}\) (1962), p293.

\(^{43}\) Kukla (2002). Kant’s argument that freewill is a pre-requisite to being normatively bound is given in section three of the *Groundwork* (2012).
To summarise where we have got to thus far, In *Being and Time* Heidegger’s project is what he calls ‘ontological’; he is interested in working out what it means, in the broadest sense, for something to exist. The way to do this, Heidegger claims, is through an analysis of the ways in which the human being - Dasein - exists. One of the ways in which Dasein exists is as an individual human being, not a being that is merely ensconced within a social world; within *das Man*. This individuated Dasein is *authentic* Dasein. The conscience, for Heidegger, is identified with a voice that summons the human being to recognise that it always operates within a public realm and is necessarily *guilty* (in an existential sense) because whenever it acts, it acts on some possibility and not others. It is this failure to perform certain actions that constitutes primordial or existential guilt. Plausibly, this is not *just* an argument as to how Dasein is individuated but also an argument as to how normativity is possible. The call of conscience, seen as a call from the individual (conceived in one way) to the individual (conceived in another way), calls the individual to recognise that she is the author of her actions. And it is in *wanting* to have a conscience, in taking responsibility for this way the way in which we are ensconced within ‘the They’, that the individual comes to be genuinely bound by norms.

From the account given thus far we don't see an account of conscience which looks like any sort of a rival to the moral agency accounts that were the subject of the previous chapter. In Heidegger’s view the moral agency accounts would simply be *ontic* analyses of conscience, which are answering a different question that his account of conscience is supposed to be. Heidegger’s existential conscience is not the sort of notion that could be captured in a moral right of conscience or that could potentially be engaged in our practical thought.

However, given that Heidegger used the term ‘conscience’ to describe this concept which, on the face of it, has little connection to any familiar analysis of conscience it would be useful to say more about why Heidegger would call this the ‘conscience’ and what connection it has to an ordinary conception of conscience. I mentioned earlier Heidegger’s discussion of the ‘ordinary’ notion of guilt (which is, of course, closely tied to conscience). The connection between the existential notion of guilt and a related ‘ordinary’ notion of guilt makes sense, I suggest, given the nature of the phenomenological approach. In this section I will draw out the connection between the existential and ordinary notions of conscience which point the way towards a revisionary Heideggerian account of ordinary conscience.
In reference to the connection between existential concepts like conscience and their everyday counterparts, Heidegger says:

All ontological investigations of such phenomena as guilt, conscience, and death, must start with what the everyday interpretation of Dasein says about them.\(^{44}\)

I mentioned in section one, that Dasein has a pre-ontological and ‘shallow’ understanding of the notions in question. Given this, it makes sense that while Dasein’s everyday understanding of those notions will not be authoritative it tell us *something* about the notions in question. It is the fact of our ‘fallenness’ (the way we are absorbed within ‘the They’) which means we don’t have access to the concepts as they really are. Nevertheless, “the ordinary experience of conscience must somehow – pre-ontologically – reach this phenomenon”.\(^{45}\) It is important for any explanation of the connection between existential conscience and ordinary conscience, not just how the everyday concept gives a reference point for the existential concept, but also *why* the everyday concept is misleading in its representation of the ordinary concept.

Heidegger acknowledges that it is a potentially legitimate criticism of his account of conscience that it is too far removed from the everyday conception: can any account that is so far removed from the everyday conception of conscience legitimately go by the tag ‘conscience’? Heidegger lists several central tenets of ordinary conscience in order to see how that notion and existential conscience come apart. He begins with the distinction in the ordinary interpretation between the good conscience and the bad conscience.\(^{46}\) Good conscience, Heidegger suggests is a *clear* conscience. It involves feeling certain that one has done nothing wrong. But, says Heidegger, this is nothing like the conscience (in its existential form). Existential conscience could never be certain it has *not* done something. Existential conscience calls the agent to recognise that she is always guilty. There is no room for recognising that one is not-guilty. Heidegger suggests that this 'reversal' between the existential and ordinary conscience is completely plausible given the tendency of human nature to cover over and disguise what it finds deeply unsettling.

\(^{44}\) (1962), p 326.

\(^{45}\) (1962), p336.

Does Heidegger have a view on what exactly a correct account of ordinary conscience is? Heidegger's response is intentionally vague here; several different interpretations of ordinary conscience might be legitimate depending on what ‘the They’ - our everyday world of social interaction - requires. However, Heidegger does hint, I think, at the plausibility of a revisionary account of ordinary conscience grounded in the conscience as identity approach.\(^47\)

the interpretation which is more primordial existentially also discloses possibilities for a more primordial existentiell understanding, as long as our ontological conceptualization does not let itself get cut off from ontical experience.\(^48\)

What does this mean? Earlier, in the context of guilt, I mentioned Heidegger’s existential /existentiell distinction which captures the distinction between the underlying concept (such as existential guilt and existential conscience) and its everyday, real-world counterpart. Heidegger seems to be recognising that there could be a legitimate account of ordinary conscience which is in some sense ‘modelled’ on his account of existential conscience. Donovan Miyasaki offers an interpretation of Heidegger whereby Heidegger’s existential conscience can be seen as a model for ordinary moral conscience.\(^49\) Briefly, Miyasaki argues that we act authentically and in accordance with conscience whenever we act on an understanding of all the practical possibilities that lie open to us. Wanting to have a conscience becomes wanting to act from the recognition that we are not identical with 'the They'. While I don’t think that there is a sound textual basis for thinking Miyasaki’s interpretation is the same as Heidegger’s account of conscience, it is a gesture at the way in which a Heideggerian conception of ordinary conscience might look. I argue that the next account of conscience that I consider, Hannah Arendt's account, can also be seen as an account of ordinary conscience modelled on Heidegger’s existential conscience.

In summary, Heidegger’s account of conscience is quite unlike any of the conceptions of conscience I have considered up until this point. Heidegger’s account of conscience is not intended to capture some aspect of our moral or practical psychology. It is not intended to

\(^{47}\) It would be a revisionary account because, as I mentioned earlier, the current ontic conception of conscience is a conscience as moral agency approach which disguises and covers over the way things really are.

\(^{48}\) (1962), p341.

\(^{49}\) Miyasaki, (2007).
capture the capacity of humans that is protected by a moral right of conscience. The concept that Heidegger refers to using the term ‘conscience’ is an existential concept which serves two core theoretical questions in *Being and Time*. One, how is it that we can be individuated from the social world in which we are ensconced? Two, how is it that any rule or principle could be genuinely normative given that we are ensconced within that world and therefore not truly free?

Heidegger argues that the account of conscience that satisfies these questions is function of Dasein. It is that function of Dasein to call itself to be authentic. It calls itself to acknowledge that it is always guilty (in an existential sense) by recognising that in acting it is always closing off practical possibilities. In wanting to have a conscience the individual takes responsibility for the path that she has gone down. If we are consciously aware of any element of conscience it would be in the feeling of uncanniness; the feeling of not being at home in the world. Of course, in this thesis I am not interested in working out whether or not Heidegger’s account of existential conscience is correct. But nor has my interest in Heidegger’s account been purely historical. Heidegger’s existential conscience is an instance of the conscience as identity approach. Conscience for Heidegger is emphatically not a capacity of moral agency. Also it does, in some sense, analyse conscience as a matter of being true to self. The conscience is the call for the individual from herself, to herself, to be herself. But my real interest is in whether an ordinary account of conscience, an account that does capture some aspect of our practical psychology, can be an instance of the conscience as identity approach. In the next section I consider such an account.

2.4. Arendt and Conscience as 'Living with Oneself'

Hannah Arendt’s account of conscience is the last entry in this intellectual history of conscience. Arendt’s interest in conscience is initially motivated by practical concerns rather than the theoretical concerns which motivated Butler, Smith, Kant and Heidegger. Arendt first articulates her account of conscience in an attempt to explain the moral deficit of those who collaborate with or are generally involved in morally abhorrent political regimes. Presumably, some are involved because, after reflection, they wholeheartedly sign up to the

---

50 This would be impossible to do without giving a detailed defence of Heidegger’s ontological project in *Being and Time*. 
pernicious 'moral' code of that regime. Others, perhaps, go along with the regime out of fear of failing to co-operate. But a sizeable group (which, in her view, includes Adolf Eichmann), Arendt suggests, collaborate because they just don't think. The possession of a conscience, Arendt suggests, is just an inevitable consequence - a 'side-effect' or 'by-product' - of being someone who thinks properly. Arendt's account is a conception of conscience as identity, I argue, because Arendt analyses thinking in terms of an internal dialogue that the agent has with herself. So for Arendt, to possess conscience is just to engage with and be true to that self.

In *Life of the Mind*, Arendt expresses surprise at Heidegger’s rejection of the ordinary interpretation of conscience and suggests that his own analysis is “unaccounted for by the phenomenological evidence”. Arendt's account does have some affinities with Heidegger's account. Arendt shares a generally phenomenological approach; what matters to Arendt is understanding this phenomenon of conscience. And, like Heidegger, Arendt thinks that the true structure of conscience is, to some extent, obscured in our everyday engagements. The activity of thinking and the related concept of conscience need to be carefully distinguished from related capacities and states with which they are sometimes confused. The capacity to think, for Arendt, is importantly distinct from other mental capacities such as the capacities to judge, know and be conscious.

But while her account does not replicate our everyday understanding of conscience, it is a conception of conscience as identity that is much closer to our ordinary experience of conscience than the notion that Heidegger articulates. For Arendt, the phenomenon of conscience is that which is captured in the familiar locution, 'could I live with myself if I were to do that?'. Arendt is attempting to capture the ‘ordinary’ conscience rather than trying to provide a primordial or existential concept as Heidegger did.

Arendt offers her initial idea that evil is often characterised, not by bad thoughts per se, but the absence of any thought at all, in *Eichmann in Jerusalem*. However, it is in her more
abstract work *Thinking and Moral Considerations*\(^{54}\) (and its subsequent reworking in *Life of the Mind*\(^{55}\)) that she provides the theoretical underpinning for that idea. It is by understanding what thinking is that we can understand what conscience is and how it can constitute a bulwark against evil-doing.

Arendt's starting point in her analysis of thinking is the Kantian insight that there is a divide between thinking and knowing.\(^{56}\) It is the faculty of Reason that is responsible for thinking and the faculty of intellect that engages in knowing. Though Arendt departs from Kant's claim that thinking is restricted to any particular subject matter (i.e. matters of faith that are in principle unknowable). The knowing/thinking distinction is crucial, Arendt claims, because if thinking is to be a bulwark against evil, it must be ubiquitous. And knowing is not an activity that everyone is necessarily capable of engaging in. None of this is to say that there is not an important connection between thinking and knowing. Plausibly thinking is a pre-requisite to knowing.

According to Arendt, the defining characteristic of thinking is its ability to *interrupt* whatever activity we are engaged in:

> For it is true that the moment we start thinking on no matter what issue we stop everything else, and this everything else, again whatever it may happen to be, interrupts the thinking process; it is as though we moved into a different world.\(^{57}\)

This capacity of thinking to interrupt explains why, except where thinking is serving some other end, it can feel so unnatural. It makes our otherwise straightforward navigation through things more difficult. Thinking also, in some sense, interrupts itself. It is of the nature of thinking that it is always questioning its own conclusions and propositions. It is because of this that Arendt claims:

> we cannot expect any moral propositions or commandments, no final code of conduct from the thinking activity, least of all a new and allegedly final definition of what is good and evil.\(^{58}\)

\(^{54}\) (1971).

\(^{55}\) (1978).

\(^{56}\) (1978), p62.

\(^{57}\) (1971), p423.
Arendt takes this point from Socrates’ observations about the nature of thinking:

thinking inevitably has a destructive, undermining effect, on all established criteria, values, measurement for good and evil, in short on those customs and rules of conduct we treat of in morals and ethics.\(^{59}\)

Here we see a significant departure from Kant in Arendt’s conception of Reason. For Kant, thinking is destructive, but also constructive. While thinking will cause us to question the moral views we otherwise adhere to, for Kant, it was also through thinking (through Practical Reason) that we construct the correct moral propositions in accordance with the categorical imperative. We might wonder whether Arendt is overstating the case here; while there is a tendency of thought to constantly call into question the propositions we believe in, we could still hold on to these propositions, albeit fallibly.

Arendt admits that this tendency of thought to constantly call into question is potentially dangerous. If misunderstood, it can lead to nihilism.\(^{60}\) But Arendt seeks to emphasise that it is non-thinking that is really dangerous. It is non-thinking that results in the blind acceptance of 'moral' codes thereby facilitating atrocities like the holocaust. Non-thinkers are those who never make up their own minds.

Thus far there has been no mention or reference to self - something crucial for a purported account of conscience as identity. This aspect of thinking (and thus, conscience) lies in its underlying structure as an internal dialogue between two different 'selves' (what Arendt calls the 'two-in-one'). Thinking is talking to oneself - albeit talk that we rarely vocalise. Thinking always aims to harmonise these two different selves - the talker and the one being talked to - in order to achieve internal harmony. Arendt claims that it is this underlying structure of thought, which explains Socrates' famous dictum “it is better to be wronged than to do wrong”. Doing the wrong thing involves the agent becoming a wrongdoer. And when an agent is considering doing something wrong - thinking- she is considering whether or not she wants to occupy the same space as a wrongdoer. The agent is deliberating as to whether or not she can 'live with' that person.

\(^{58}\) (1971), p425.

\(^{59}\) (1978), p175.

\(^{60}\) (1978), p176.
It is this underlying structure of thought as a 'two-in-one' that explains the etymology of the term 'consciousness': A word which literally means "knowledge by one's self". Though, Arendt claims, consciousness is not itself thinking: consciousness is a necessary precondition to thinking, in thinking the internal dialogue inherent in consciousness is brought explicitly before the mind.

Arendt claims that the phenomena of conscience is just a 'side-effect' of thinking when it is analysed in the way that I have described. It is that experience we have when considering an action we have either performed in the past or may perform in the future where we judge that the performer of that action is, or would be someone that we cannot live with. For Arendt, the conscience does not give any particular prescriptions. It does not say 'do x' or 'do not do y'. Those sorts of imperatives are the domain of the 'faculty of judging particulars', not conscience.

With this understanding of the underlying structure of thought and conscience we are in a position to see how it is that the conscience can constitute a bulwark against evil action. It does not give us any instinctual access to the moral truth, nor is it the capacity by which we adhere to our own moral code. Thought and its by-product conscience interrupt our regular activities and judgements (including moral judgements) and call them into question. And this interruption is itself somewhat peculiar. In Personal Responsibility under Dictatorship (1964), Arendt argues that those who refused to collaborate with the Nazi regime did not merely ask 'do I agree with the code of rules that I am being asked to follow?'

Their criterion, I think, was a different one: they asked themselves to what extent they would still be able to live in peace with themselves after having committing certain deeds. And they decided to do nothing; not because they thought that the world would be changed for the better, but simply because only on this condition could they go on living with themselves.61

In Arendt's terms, to have a conscience, to be conscientious, is simply to be disposed to live with oneself. It is not simply to judge that things would be better if one rule rather than another was followed. Those who do not possess a conscience are those who do not question whether they could live with themselves if they were to perform a certain action.

As Larry May put it:

---

61 (1964), p44.
The insight is the voice of conscience which begins from the realization that it means more to the self to be consistent with itself, to achieve what Arendt called inner harmony than to fall into line with the other selves that it encounters.\textsuperscript{62}

Thus far Arendt’s account of thought and conscience has been, in some sense, descriptive. A description of what it is, at a deep level, to think and to have the capacity of conscience. But in “Personal Responsibility under Dictatorship”, Arendt makes it clear that we \emph{ought} to act in accordance with the conscience so-construed.

those who cherish values and hold onto moral norms and standards are not reliable. Much more reliable will be the doubters and the sceptics, not because scepticism is good or doubting things wholesome but because they are used to examine things and make up their own minds. Best are those who will know only one thing for certain: that whatever else happens, as long as we live we shall have to live together with ourselves.\textsuperscript{63}

We all have this capacity to engage with our conscience, but we don’t always \emph{actually} do so: It is what we ought to do. Perhaps even more readily then we ought to engage our ordinary moral reasoning. There are clear limitations to Arendt’s account of conscience. It seems as though some evil people might be perfectly willing to live with their evil selves. But, this does not mean that conscience is of \emph{no} value in getting people to do the right thing. As Paul Formosa puts it: “In any case, conscience can have a conditioning effect, by creating an interest in one’s own moral integrity, for at least those people who cannot live with an evildoer”.\textsuperscript{64} Another limitation of Arendt’s account of conscience is that it is purely negative. It proscribes certain conduct, but it doesn’t make any positive recommendations.\textsuperscript{65}

While, perhaps, the most common examples of acts of conscience are refusals to do certain things, an account of conscience should also allow for the fact that an individual’s conscience can mandate that she perform certain positive acts as well. I consider the distinction between these two different applications of conscience in the next chapter.

In this chapter I have continued my summary of some of the more important philosophical analyses of conscience. In this chapter I looked at two accounts of \emph{conscience as identity}. The

\textsuperscript{62} (1983), p58.
\textsuperscript{63} (1964), p45.
\textsuperscript{64} (2010), p95.
\textsuperscript{65} A point emphasised by Formosa (2010), p99.
idea that the concept of conscience ought to be analysed in terms of identity is the guiding idea of this thesis. My primary focus in this chapter was Heidegger’s account of existential conscience. Heidegger’s account is a difficult one to fit into any study of conscience. Heidegger’s account is an account of existential conscience which plays a functional role in his ontological approach in *Being and Time*. By contrast, the account that I am trying to defend is supposed to be a concept that is captured in moral rights of conscience and is implicated in our practical psychology. Nevertheless, I argue, Heidegger’s account is relevant to the analysis that I will give. Heidegger’s account points the way for a non-existential analysis of conscience. I suggest that Arendt’s account of conscience is a conception of conscience modelled on the Heideggerian analysis. Arendt argues that the conscience is a side-effect of an individual *thinking*. To possess a conscience is for an individual to decide that she could not live with herself if she were to do something. But there are questions that remain, what does it mean to be willing to ‘live with oneself’? In chapter seven I offer my own conception of conscience a narrative self-conception which tries to make sense of this very idea.
Part II Communicative Conscience
Chapter Three: Brownlee and the Communicative Principle

When a suitably qualified clinician refuses to perform an abortion on the grounds that abortion is immoral we often characterise that clinician as conscientiously objecting to the provision of abortion services. We make a similar characterisation of draftees who refuse to participate in a war when the law demands their participation. In Conscience and Conviction, Kimberly Brownlee refers to these sorts of practices as ‘personal disobedience’. Personal disobedience is a refusal to do as the law requires on the grounds that one morally objects to that law. This practice is called ‘personal’ disobedience in order to distinguish it from civil disobedience as the reason the agent disobeys the law is simply to follow her personal moral conviction rather than to express, in a public way, the agent's condemnation of that law. In Brownlee's view it is civil disobedience that has the better claim to be an exercise of an agent's conscientious conviction.¹ This chapter is a description and evaluation of Kimberley Brownlee’s argument that civil disobedience is a more conscientious act than personal disobedience.

In the first two chapters of this thesis I looked at two prominent historical approaches to conscience; conscience as moral agency and conscience as identity. In the first chapter I suggested that there were elements of the conscience as identity approach in the conceptions of conscience defended by Smith and Kant. In the last chapter I focused on the elaborate but baffling account of conscience as identity defended by Heidegger. In this chapter and the next, I shift focus slightly. I describe and evaluate a recent philosophical treatment of conscience, Kimberley Brownlee’s Conscience and Conviction: The Case for Civil Disobedience. Thus far I have looked at the way in which conscience can implicate an individual’s moral agency and an individual’s identity. Brownlee’s account of conscientiousness is, in some sense, a ‘third way’. Her account emphasises the communicativeness of conscientious convictions. I largely accept Brownlee’s argument as to

¹ The claim is not that personal disobedience is unconscientious. Rather, conscientiousness is best seen as a ‘sliding scale’ with some practices we can legitimately describe as ‘conscientious’ being very conscientious and some other practices we can also describe using that term being less conscientious. According to Brownlee, civil disobedience is the most conscientious practice, then non-evasive personal disobedience and finally, evasive personal disobedience is the least conscientious ((2012), p 27).
the importance of communicativeness, but will eventually argue, in chapter five, that it is not
the only principle of conscience. I argue that there are three principles of conscience which
all have a certain weight, but none of which are necessary.

This chapter is divided into several different sections. In section one I describe Brownlee’s
account of the different practices that might count as conscientious: civil disobedience (in
direct and indirect forms), assistive disobedience, radical protest and personal disobedience.
In section two I describe the first requirement of the communicative principle, the
consistency requirement. I suggest that consistency is a plausible requirement of
conscientious action, but that it seems to have little to do with the potentially communicative
nature of conscientiousness. In section three I describe the universality requirement and agree
that this is a plausible requirement of conscientiousness. However, I claim it is in tension
with the supposed ‘first-personal’ requirement of conscience. I suggest that this tension is
best resolved by supposing that the first-personal aspect of conscience pertains to one of the
principles of conscience (the identity principle) whereas the universality requirement pertains
to a different desideratum (the morality principle). In section four, I begin my discussion of
the non-evasion requirement of conscientiousness. I argue that we cannot settle the question
as to whether this is a genuine requirement of conscience, whether the conscientious agent
must do something positive in support of her convictions, without considering what the
normative grounds for conscience are. It is that topic which I discuss at the beginning of
chapter four.

3.1. Varieties of Disobedience

In Conscience and Conviction Brownlee distinguishes concepts of conscience and
conscientiousness and argues that each concept facilitates distinct normative protections:
conscience and conscientiousness. For Brownlee, conscientiousness is “a largely descriptive
property of deeply held though possibly mistaken moral commitment”. This can be
contrasted with the concept of conscience of which Brownlee says:

---

2 I continue to use the terms ‘conscience’ and ‘conscientiousness’ interchangeably.

3 (2012), p51.
conscience is an evaluative property. It is a set of practical moral skills that stem from an inward knowledge of the workings of our own mind and heart. Having conscience means not just taking morality seriously (conscientiousness), but also being genuinely, self-consciously morally responsive. The moral reality that conscience makes us responsive to is, in my view, a fundamentally pluralistic one.4

In Brownlee’s view both notions ought to be respected by the law, though in separate ways. In this chapter I am only interested in Brownlee’s account of conscientiousness rather than her distinct account of conscience. The reason that I do not discuss her account of conscience is, as I suggested in the introduction, I wish to defend an account of conscience that is largely normatively and meta-ethically neutral. My question in this thesis is about the nature of the concept of conscience implicated in rights of conscience, if there are to be such rights. I aim to offer no argument as to why the possession of a conscience is a good thing, or why convictions and acts that are the result of conscience should be protected. Consequently, I do not consider whether conscience gives us access to the (pluralistic) moral truth or is an ideal that we should cultivate as Brownlee argues. Which is, of course, not to say that such an account of conscience is incorrect; it is simply not the sort of account of conscience that I will evaluate in this thesis. For the remainder of this thesis I treat the terms ‘conscience’ and ‘conscientiousness’ as synonymous.5

In this chapter, I focus on one of the key arguments that guides Conscience and Conviction; the argument that the act of civil disobedience is more conscientious than the act of personal disobedience. This is not a trivial argument that the term ‘conscientious’ really applies to one sort of behaviour and not the other. Plausibly, the conscientiousness or otherwise of an agent’s actions affects how society and the state ought to deal with that agent.

How do we go about answering the question ‘what is it for an act to be conscientious?’ Brownlee endorses a method similar to the one that I advocated in the introduction. We don’t just look at the various descriptive ‘platitudes’ implicit knowledge of which is essential for

---


5 Or, more often, I treat the conscience as the psychological capacity that results in conscientious convictions and actions. I accept that the terms ‘conscience’ and ‘conscientiousness’ are often used to refer to quite separate notions. For example, Charles Pigden refers to ‘conscientiousness’ as the “widespread tendency to live up to one’s precepts” ((2009), p89). But it would be strange to ever define ‘conscience’ in such a way.
mastery of the folk concept referred to as ‘conscientious’ or ‘conscience’. We also look to the normative significance of the concept in question. On the face of it, this approach seems to get the cart before the horse: surely the natural approach is to work out what the conscience is (a descriptive matter) and then determine what normative significance (if any) that concept has. However, in my view, we do need some account of what the eventual goal of our account of conscientiousness is, from the beginning, so that we know precisely what we are working towards/trying to achieve. There is no use in working out the normative significance of conscience in political philosophy if the concept that we are really interested in is the capacity of conscience lacking in the moral thought of psychopaths.

Our starting point, whether or not we take into account normative constraints, is to look at the central features of the practices that we (the folk) commonly describe using the term ‘conscientious’ as well as practices that seem similar to the practices so described.

As mentioned earlier, a practice that we typically judge to be conscientious is a professional’s refusal to do something they seem otherwise required to do on the basis that that professional has a personal or moral objection to performing the act in question. It may be a qualified clinician’s refusal to perform euthanasia in a jurisdiction where patients are legally entitled to it; or, a pharmacist’s refusal to dispense the emergency contraceptive pill. We commonly refer to these objections as ‘conscientious objections’ and they are often referred to as such in legislation that protects these objectors. One distinctive feature of these acts is that they involve disobedience on behalf of the agent. Furthermore, it is disobedience that is motivated by some specified mental state (or set of mental states) of the agent. Without

---

6 For example, this is the approach to analysing moral judgements defended by Michael Smith (1994).

7 This isn’t to say that the supposed normative import of a concept cannot be captured by the approach of traditional conceptual analysis. Larry Sumner emphasises the normative constraints on the concept of happiness in his analysis of that concept (1996). Kristie Miller argues that in the case of social and political concepts, after some conception meets the barest descriptive constraints we need to look at the normative importance of that conception in order to determine if it is ‘correct’ (2009).

8 For example, that is Robert Hare’s use of the term ‘conscience’ (1993). For more discussion of that approach see the introductory chapter of this thesis.

9 It may be that some practices that are not called ‘civil disobedience’ still count as civil disobedience.

10 It is worth noting that this disobedience need not necessarily involve disobedience of the law. Sometimes the law allows for individuals to conscientiously object. In those cases there is still a sense, I think, in which that act counts as a form of disobedience. Perhaps it is disobedience of what is generally acceptable or expected in a profession that is at issue in such a case.
specifying yet what exactly that mental state or set of mental states must be, we can all agree that we would never describe as conscientious a murderer who was motivated to kill by bloodlust.

This sort of disobedience, that the clinician and pharmacist are engaging in, Brownlee calls personal disobedience. This is the refusal to follow a law, injunction or directive on the grounds of perceived personal conviction.\(^\text{11}\) A related sort of disobedience is civil disobedience. When Rosa Parks refused to give up her seat to a white person in 1955, she was a civil disobedient. We may often refer to this sort of act using the term ‘conscientious’ as well; though, perhaps, not as readily as we would in the case of the clinicians and the pharmacist. What is it for an act to be one of civil disobedience? Without claiming to give sufficient conditions, Brownlee suggests:

> It must include a deliberate breach of the law taken on the basis of steadfast personal commitment in order to communicate our condemnation of a law or policy to a relevantly placed audience.\(^\text{12}\)

This aim, to communicate condemnation has both backward-looking and forward-looking dimensions. It is backward-looking in that it aims to communicate our disavowal of and dissociation from the protested law or policy as well as our reasons for doing so. Civil disobedience, necessarily, has a communicative intention.

Brownlee suggests that the act of civil disobedience must be carried out in order to communicate condemnation of a law or policy.\(^\text{13}\) However, the law that the agent is breaking need not be the same law that she sets out to condemn. Direct civil disobedience is disobedience of the very law that we oppose (such as in the Parks’ case). Indirect disobedience is a breach of a law other than the one that we oppose: For example, protestors

\(^{11}\) Brownlee (2012), p27. Technically, Brownlee uses the term ‘personal objection’ to refer to disobedience of a law, directive or injunction, and reserves ‘personal disobedience’ to refer to disobedience of the law only. I use the term ‘disobedience’ for all cases because, as I stated above, intuitively, ‘disobedience’ is not a term that refers to law-breaking only.

\(^{12}\) (2012), p18.

\(^{13}\) Or at least paradigmatically, that is what civil disobedience involves. What exactly constitutes a paradigm case turns out to be an important element of Brownlee’s argument that civil disobedience is more conscientious than personal disobedience. I discuss the nature of a paradigm case in detail in the next chapter. In chapter five I draw out the requirements of the identity principle of conscience through the treatment of Hugo’s Les Misérables as a paradigm case of conscience.
who occupied the draft board in protest at conscription engaged in indirect disobedience. The law that they breached (the law against trespass) was not the same as the law that they sought to condemn. The constraints mentioned, on whether some act counts as 'civil disobedience', have implications for the way in which civil disobedience must be carried out. It means that that the civil disobedient cannot be overly radical in the way in which she communicates her conviction. To see lasting change in the law or policy that we protest we need to try and rationally persuade our interlocutors that our own view is the correct one. It is not acceptable to try and force others to change their mind. For example, plausibly, those who block access to abortion clinics are not bona fide civil disobedients; their behaviour has a dialogue-stifling rather than dialogue-promoting effect.

Also problematic are acts that target one party in order to ultimately communicate condemnation towards another party. For example, an agent might organise a sit-in of a government building in order to signal condemnation of the military action of that government. In doing so the protesting agent might prevent a civil servant from carrying out her job. The permissibility of such an action will depend on whether (1) the action drowns out the moral appeal of the effort (2), it fails to respect the targeted person as an end in herself and (3), and the action is consistent with the agent admitting that she might be mistaken. Under those circumstances, the act could still count as civil disobedience.

Obviously, violent disobedience will be problematic. However, it is still possible for violent behaviour to count as civil disobedience. It may well be that minor and limited violence is consistent with the communicative constraints that apply to civil disobedience. Any of this conduct alluded to, that might otherwise be referred to as ‘civil disobedience’ but lacks the communicative constraints outlined by Brownlee, is referred to as ‘radical protest’. This behaviour still aims to communicate condemnation of a law or policy as civil disobedience.

---

14 This distinction is related to but different from Ronald Dworkin’s distinction between integrity-based and justice-based disobedience (as discussed in the introductory chapter to this thesis). Integrity-based disobedience, according to Dworkin, occurs when the individual’s conscience simply will not permit the individual to follow the law in question. Whereas justice-based disobedience occurs when the individual wants to make a broader point about the injustice of a law.

15 (2012), p20. Though this is somewhat speculative. It may be that the actions of those activists does enhance dialogue on the issue. I consider this possibility in greater detail in the final section of this chapter.

does, but it is not sufficiently communicatively constrained. Its failure to be communicatively constrained might lie: (a) in the way that it is unduly violent; (b) in the way that it treats others as mere means; or, (c) in some other way where the behaviour in question stifles dialogue.

Another type of disobedient act that is similar to civil disobedience is what Brownlee calls *assistive disobedience*.\(^\text{17}\) Civil disobedience is an act done *for the purpose* of communicating condemnation of a law or policy. Another way of putting this is that it is an act for which this communicative aim is the primary objective. But an action can communicate condemnation of a law or policy even where that communication is not the primary objective or purpose of that act. Brownlee offers the example of a clinician performing an illegal assisted suicide on a terminally ill patient. The act is not done (or should not be) for the purpose of condemning the law that prohibits the procedure; it should be done for the purpose of aiding the patient. Nevertheless, the act is still communicative and conscientious as long as that is an intention of the disobedient. Brownlee suggests that much of what she goes on to say in support of civil disobedience similarly the case for assistive disobedience.\(^\text{18}\)

Now that we have an understanding of the sort of practice with which personal disobedience is supposed to be contrasted it is worthwhile describing personal disobedience in greater detail. First, necessarily, civil disobedience involves a deliberate breach of the law. Personal disobedience need not.\(^\text{19}\) Like assistive disobedience, in personal disobedience, the breach of the law might be completely incidental. What matters to the agent is that she act in accordance with her personal conviction, whether or not that action is a breach of the law. Of course, by definition, personal disobedience requires that the agent actually acts contrary to the law but it need not be deliberate. The deliberate/non-deliberate distinction marks a deeper difference between the two types of disobedience. Civil disobedience requires that the agent have a communicative aim to change the law via reasoned discourse. Personal disobedience

---

\(^{17}\) (2012), p24.


\(^{19}\) (2012), p24.
has no such requirement.\textsuperscript{20} If one does not set out to change the law then it is no surprise that the breaking of the law can be completely incidental.

Personal disobedience can itself be divided into two categories; it can be evasive or non-evasive. Evasive Personal disobedience occurs when the agent acts surreptitiously in an attempt to ensure no interference from others with her disobedience.\textsuperscript{21} Often such behaviour is motivated by concern for one's own self-protection. An example might be a US draft-dodger who flees to Canada in order to avoid being punished. Non-evasive disobedience occurs when we break the law in order to act in accordance with our convictions but do not mind whether we are seen doing it or not. It differs from evasive personal disobedience as when we act we are not attempting to hide what we are doing. And it is not a communicative form of disobedience as our primary objective is not to communicate to others our condemnation of a law. Plausible examples of non-evasive disobedience include, a civil registrar who swaps assignments with a colleague so that she does not have to register same-sex partnerships, a clinician who refuses to be involved in the provision of abortions, or a Judge who refuses to hear a gay couple's application for an adoption.

Thus far we have looked at a variety of different practices that all involve disobedience of the law (or, perhaps, some other non-legal directive); civil disobedience, radical protest, assistive disobedience and personal disobedience. Two of these practices (civil and assistive disobedience) were defined so as to be communicative acts. Communicative disobedience is disobedience where the agent acts for the purpose of communicating condemnation of a law or policy. Given that the communicativeness of an act is, according to Brownlee, what determines the conscientiousness of an act, the robustness of this distinction is crucial. Why must we accept that an act is only conscientious if it is has a communicative intention? In the next section of this paper I will evaluate Brownlee's refinement of the notion of communicativeness. I argue that, according to Brownlee's communicative principle, personal

\textsuperscript{20} The wording that Brownlee uses is that personal disobedience does not 'necessarily' have this communicative intention, which seems to leave open the possibility that contingently personal disobedience might have this intention. This can't be right. Neither of the two possibilities that Brownlee goes on to describe, non-evasive and evasive personal disobedience permits the agent to have such an intention. And this makes sense; personal disobedience with such an intention present would be, according to Brownlee's taxonomy, assistive disobedience.

\textsuperscript{21} (2012), p28.
disobedience is just as communicative and therefore just as conscientious, as civil disobedience.

3.2. The Consistency Requirement of the Communicative Principle

I have already mentioned that in both Brownlee’s approach and my approach, conceptual analysis must make reference to normative considerations. The concept of conscience cannot simply be given by the descriptive intuitions of the folk about what does and what does not count as conscientious. As the concept that we are interested in is a normative concept we also need an account of normative importance of the concept in order to arrive at our final conceptual analysis. In chapter four I will consider Brownlee’s argument in defending the non-evasion condition of conscientiousness that a conscientious agent must sometimes do something positive in support of her conviction not because that follows, in some sense, from folk intuitions about what counts as 'conscientious', but because it is only those sorts of conviction that would be worthy of respect in the way that conscientious convictions are worthy of respect.

Putting to one side those normative considerations, for now, the starting point for Brownlee’s analysis of conscientiousness is to look at what we would judge to be paradigm cases of conscientious behaviour and then abstract away from those cases, the features of the case that make it an instance of conscientiousness. We then test those features as putative conditions of conscientiousness by seeing whether they have any counterexamples - by seeing whether there are any instances where we would still be inclined to judge that the agent is conscientious even though those features were not present. I take a similar approach in chapter five where I begin to discuss my idea of conscientious identity. In that case I use examples from Victor Hugo’s Les Misérables as my paradigm case of conscientiousness.

In the first chapter of Conscience and Conviction, Brownlee describes her purpose as to "outline a conception of conscientious moral conviction as a descriptive property of sincere and serious, though possibly erroneous, moral conviction". At times Brownlee treats conscientiousness as not simply a feature of sincere moral conviction but a necessary

22 (2012), p16.
feature. I will suggest in the first section of chapter four that that is implausible. There are various situations where it is appropriate to judge that an agent possesses a moral conviction where it would be inappropriate to judge that she possesses a conscientious conviction.

Brownlee develops the initial idea, that being communicative is something to do with engaging others in reasoned dialogue into something that she calls the communicative principle. The communicative principle breaks down the notion of conscientiousness into four key requirements. To be communicative (and thus conscientious) is to be consistent in one’s convictions, to apply them universally, to be non-evasive in one’s conduct when acting on a conviction, and to engage in reasoned dialogue with others on the merits of one's convictions. I will now consider the plausibility of this principle by examining each requirement of the communicative principle in turn. In this chapter my focus is on the first two requirements. I suggest at the end of the chapter that the only way to assess the legitimacy of the last two requirements is to have an account of what the normative basis for conscientiousness is (which is the subject matter of the first half of the next chapter).

The first proposed element of the communicative principle is consistency. This condition requires that we be consistent in our judgements, motivations and conduct to the best extent we are able. For example, we would doubt the sincerity of an agent's claim to have a conscientious conviction against participation in war if she voluntarily joined the Marines. The inconsistency between the agent’s claimed convictions and her actions is suggestive that she does not hold the convictions she claims to have at all. Some version of the consistency requirement, is often defended as an essential element of conscience. This consistency condition is similar to a thesis defended about moral judgements sometimes known as ‘motivation internalism’. This is the view that once an agent has judged that some action is

23 For example, Brownlee describes the communicative principles as developing a "general principle of sincere moral conviction" ((2012), p16). And elsewhere Brownlee says “This is what I call the communicative principle of conscientiousness. This principle lies at the core of my analysis of what counts as ‘sincere moral conviction’. ” ((2012), p29). Again, the implication is that this is a feature of all sincere moral convictions.


25 For example Ryle claims that it is a necessary condition of an agent possessing a conscientious conviction “That he regularly and readily behaves in accordance with it, on occasions when it is relevant” ((1940), p33). Or, for C D Broad: That he has and exercises the conative disposition to seek what he believes to be good and to shun what he believes to be bad, as such, and to do what he believes to be right and avoid what he believes to be wrong, as such. ((1940), p118)
wrong, she must, *ceteris paribus*, be motivated by that judgement and act in accordance with it.\(^\text{26}\) I say that Brownlee's consistency thesis and moral motivation internalism are 'similar' rather than 'identical' as one could endorse one of the theses and reject the other. One might agree with Brownlee that in order for an agent to have a *conscientious* judgement she is necessarily motivated to act in accordance with that judgement, but one could still deny that this is the case with all moral judgements. Furthermore, one might think that all moral *convictions* motivate while denying that all moral *judgements* motivate; plausibly, possessing a conviction, the states of being *convinced* of something, is intrinsically motivating in a way that simply possessing a judgement is not motivating.\(^\text{27}\)

An important element of this proposed requirement is the qualifier 'to the best extent we are able'. Sometimes, Brownlee calls this qualifier a 'ceteris paribus' clause. What is the meaning of this qualifier? Which cases is it supposed to cover? Brownlee, mentions some of the unusual circumstances that might affect the agent's ability to act consistently.\(^\text{28}\) For example, someone with severe depression can be a conscientious friend even though during depressive episodes she might be unable to act on that conviction. Or the inability might be to do with the circumstances the agent finds herself in rather than her own mental states. We can imagine peculiar circumstances where an agent must temporarily fail to adhere to one conviction in order to cohere with another moral imperative. Suppose that a devout Muslim woman has a sincere conviction that she must always leave the house with her head covered. But her child's life is at risk and because of that she feels compelled to leave the house uncovered.\(^\text{29}\) Is this woman still satisfying the *ceteris paribus* clause? We might say that it does satisfy the clause as the requirement for her to act consistently with her conviction to keep her head covered only holds all things being equal – it does not apply when another more serious moral requirement applies to the situation at hand. One might argue that there are ways of explaining such a case without resorting to the 'to the best extent possible/\textit{ceteris paribus} clause. One could argue that the agent's original conviction (against leaving the house with her head uncovered) has an exception clause built into so that fully spelt out the

---

\(^\text{26}\) For example, Michael Smith defends such a thesis in (2007), p211.

\(^\text{27}\) In the introduction I described Gilbert Ryle’s argument (1940) that it is because conscience involves 'convictions’ rather than mere judgements that conscience is a necessarily first-personal capacity.


\(^\text{29}\) This was the situation for Lilian Ladele who was fired from her job as a registrar for failing to register those partnerships. Her case eventually made it to the European Court of Human Rights where her complaint was dismissed (*Eweida and Others v. The United Kingdom* [2013] ECHR 37)
conviction would be 'always cover one's head when leaving the house unless there is some compelling reason not to'.

The examples offered by Brownlee suggest two quite different ways in which the *ceteris paribus* clause can be spelled out. In the depression case we can imagine this being extended to all sorts of psychological situations that result in practical failure. An agent might act contrary to her best moral judgement out of rage, or laziness. We might expect the circumstantial inability exception to cover all cases where the agent's conviction is normatively defeated by another conviction.\(^{30}\) Is there any way to spell out *all* the cases where the 'to the best extent possible' qualifier might apply? There seem to be two possibilities that present themselves. One, it is possible that psychological and circumstantial possibility are just two entries in a finite list of exceptions that, potentially, could be spelt out. In that case, the 'ceteris paribus' clause operates as a shorthand for all those potential exceptions. Or, perhaps, the *ceteris paribus* clause could be a pragmatic device covering an indeterminate number of situations that are incapable of being explicitly listed (I consider these two possibilities in detail in section seven of chapter four). I conclude there that the communicative principle should not incorporate *ceteris paribus* clauses at all. Possible exceptions to a proposed requirement are better accounted for by holding that we are not dealing with a necessary condition of conscientiousness at all. I argue in chapter four that the communicative principle, just like the other two principles of conscience (morality and identity) should be treated not as necessary conditions of conscience, but as weighty desiderata which count in favour of some conviction or action being conscientious.

7.3. The Universal Normative Scope of Conscientious Conviction

The second putative requirement of the communicative principle is ‘universality’. According to Brownlee, the possession of a conscientious conviction entails making a judgement that purports to have universal normative scope.\(^ {31}\) What does this mean? A conviction can have a

\(^{30}\) In the case described by Brownlee a conviction is defeated by being 'outweighed' by another moral consideration. But there are various other ways in which a conscientious conviction might be defeated by other normative considerations. For example, it may be that other moral considerations can cancel or exclude a conscientious conviction from one's rational deliberation (the most influential discussion of these possibilities perhaps is by Joseph Raz in his (1975)).

\(^{31}\) Brownlee, (2012), p34.
restricted or a universal normative scope. An example of a conviction with a restricted scope is my conviction that hokey pokey ice cream is superior to vanilla or chocolate. My conviction, being convinced that hokey pokey is superior, is simply a matter of having an idiosyncratic preference for that flavour. My conviction in such a case is not supposed to govern the conduct of others. The normative scope of the judgement is restricted to myself, the individual making the judgement (or perhaps others who happen to share the sentiment).

Brownlee argues that a conscientious conviction is not similar to that sort of conviction. A conscientious conviction does not involve a judgement with restricted scope. When I have a conscientious conviction against engaging in some act I judge not only that I ought not to engage in that act, I also judge that anyone else in the same situation ought not to engage in that act. If I am a clinician who has a conscientious conviction against euthanasia, I judge not only that I ought not to be involved in it, I judge also that a similarly situated clinician also ought not to be involved. This notion of universality is most famously, perhaps, advocated by Kant when he claimed that the requirement of a moral imperative is to "Act only according to that maxim whereby at the same time you can will that it should become a universal law".  

Two points need to be made about this requirement. First, the notion of universality is commonly accepted as part of the analysis of moral judgement/reasons/imperatives in contemporary meta-ethics. And given that conscientiousness (according to Brownlee) is a property of moral convictions, is it plausible to hold that the universality of conscientious convictions has anything to do with the communicative principle? Recall the structure of Brownlee's analysis of conscientious conviction: The concept of conscientiousness is analysed in terms of the communicative principle. The communicative principle is, in turn, analysed in term of four distinct requirements: the consistency, universality, non-evasion and dialogue requirements. In order to constitute a requirement of the communicative principle, each putative requirement needs to be some way of explaining how a particular act or practice is communicative. Even if we accept that it is an essential feature of conscientious conviction that the judgements contained within those convictions have universal normative

---


33 For example, some version of this is captured in R M Hare’s universal prescriptivism (1952). Some version of this is also captured in Richard Joyce’s claim about the ‘inescapable authority’ that moral judgements claim: moral judgements are supposed to apply to everyone irrespective of their wants and needs ((2006), p191).
scope this is more plausibly a consequence of the conviction's moral nature than its communicative nature.

This point may sound trivial; what does it matter whether a conscientious conviction commits an agent to a universal judgement in virtue of the communicative nature of that conviction or in virtue of something else? It is important because, potentially, the communicative nature of conscientiousness convictions is controversial in a way that the universality of those convictions is not. It is important for anyone who wants to deny that conscientious convictions must be communicative that in doing so they need not give up the claimed universal normative scope of those convictions. It is this point, in part, which leads me to divide Brownlee’s communicative principle into separate principles of communicativeness and morality.

The second point is that there is no explicit argument given support of this condition. The thought is, perhaps, that this is supposed to be an obvious truth about conscientiousness. However, there is a substantial tradition of theorising about conscientious convictions which casts doubt on that claim. Arguably, a conscientious conviction is different from many other moral convictions precisely in the fact that it only pertains to what I ought to do. It does not pertain to what you or anyone else ought to do. It makes sense for a draft-dodger to say to himself 'My conscience simply will not allow me to participate in a war' whereas there is something unusual about him saying 'my conscience will not allow you to participate in a war'. This is not to deny that there is a universality condition for conscience. In chapter four I argue that there is such a condition but it is best captured in a separate moral principle of conscientiousness, rather than in the communicative principle. And even in that case, universality is not, strictly speaking, a necessary condition as the moral principle is only a weighty desideratum of conscience.

34 As discussed in chapter one, this can be traced back to Adam Smith's discussion of conscience in The Theory of Moral Sentiments and was also argued by Gilbert Ryle in his (1940).
3.4. The Non-Evasion Requirement

The third requirement that Brownlee posits as constitutive of the communicative principle is non-evasion.\(^{35}\) Recall that Brownlee introduced the evasive/non-evasive distinction in order to distinguish different types of personal disobedience. Personal disobedience was non-evasive when the agent merely acted without a communicative intention; she neither intended to be seen to be disobedient nor did she intend \textit{not} to be seen. Personal disobedience was evasive when the agent acts on her conviction in a surreptitious manner and has a positive intention not to be seen to be acting as she is. The difference between the two is the difference between the absence of an intention to do something and the presence of an intention not to do something. In the discussion of the communicative principle, Brownlee makes this distinction more precise in order to bring to the fore exactly why being non-evasive is so important.

In this section Brownlee describes the non-evasion condition as the requirement that we bear the risks of honouring our convictions.\(^ {36}\) What does it mean to honour our convictions? There are two separate elements: First, it means not seeking to evade the consequences of our convictions and second, it means sometimes doing something positive in support of that conviction. I call the first element the ‘negative element’ and the second the ‘positive element’. The way the condition is articulated here, of course, lines up with the way it was articulated earlier; it requires the absence of surreptitious motives. But it also adds to that earlier characterisation by requiring that the agent sometimes act positively in support of her conviction.

Let’s consider each element in turn. The first element asks that the agent not to try and evade the consequences of acting on her conviction. What exactly are the consequences we are supposed to consider? Brownlee again refers to the example of Sir Thomas More.\(^ {37}\) The consequence of his acting on his conviction was that he was put to death. But surely the test for non-evasion can’t be to do with the \textit{actual} consequences of acting on one's conviction. We

\(^{35}\) Brownlee (2012), p37.

\(^{36}\) Ibid.

\(^{37}\) (2012), p38.
are dealing with an agent's mental states before acting, after all, and the agent won't know with certainty what those consequences will be. More plausibly, it is about the anticipated consequences of acting on one's conviction. In the case of More, it is the fact that he anticipated that he would be put to death his behaviour which means that he satisfies this element.

If we accept that the negative element is concerned with anticipated consequences rather than actual consequences, what exactly needs to be anticipated? Obviously consequences as severe as those that faced by More will be sufficient, but was it necessary for More to accept those consequence to count as conscientious? This seems like too onerous a burden for conscientiousness. Brownlee accepts this point. It is permissible for an agent to seek to avoid punishments that are guaranteed, but they mustn't seek to avoid all risk of being punished or socially censured. Consider how this lines up with the original examples of evasive and non-evasive personal disobedient: The personal disobedient clinician will meet the non-evasion condition, even if she is legally permitted to refuse to perform an abortion (as legislative 'conscience clauses' often permit), as long as there is some anticipated social censure for her doing so. Likewise for the civil registrar who refuses to register same-sex marriages and as a result of this refusal swaps assignments with a colleague. By contrast, consider one of the examples of a non-evasive disobedient; the US draft-dodger who flees to Canada. That agent, does not meet the non-evasion condition as he acts to avoid being prosecuted as a draft-dodger in the US by doing so; he acts purely out of self-protection.38

The positive element of the non-evasion requirement holds that the agent must sometimes do something positive in support of her conviction. What counts as an agent 'doing something positive in support of her conviction'? For example, we wouldn't want to count the civil registrar's swapping of assignments as a positive act, presumably; that seems better characterised as an omission to do something rather than positively doing something. Similarly, a clinician's refusal to offer or be involved in abortion services; it seems better characterised as an omission (if anything does) rather than a positive act. On the other hand, some cases that Brownlee thinks (and we ought to think) are clear cases of civil disobedience

38 There is something unusual about it being a requirement of conscientiousness that an agent risk certain negative social or legal consequences. Remember, there is no requirement that conscientious acts involve a breach of the law; there is only a requirement that a subset of conscientious acts (civil disobedience, assistive disobedience etc.) involve a breach of the law.
seem to only involve an omission. For example, Rosa Parks' refusal to give up her seat seems just as much an omission as in the civil registrar case; in simply sitting in her seat she is hardly 'doing a positive act'. But it would be absurd to hold that her act did not count as conscientious, or even that it didn’t count as communicative.

Brownlee provides some illustration of the sort of behaviour intended to be covered by this requirement when she claims that "little weight could be put on the asserted 'convictions' of someone who does not think that, at least in some cases, she should be willing to act positively to prevent others from doing what she regards as seriously wrong." This suggests that what Brownlee means by a positive act is that the individual in question must behave in such a way so as to try and change the conduct (and beliefs) of others in accordance with her convictions. In the case of the clinician, her refusal to provide abortions does not in itself count as a positive act; it is not intended to stop others from doing what she judges to be wrong. The civil registrar does not count as doing something positive in support of her conviction for the same reason; her behaviour in assigning her work to other colleagues is not, in itself, an attempt to get others to do what she believes ought to be done. And this interpretation of a 'positive act' explains how it is that Rosa Parks' behaviour is consistent with the communicative principle and counts as conscientious. Her behaviour was intended to get others to change their beliefs and actions.

While we may well accept that it is praiseworthy, and all-things-considered better, when an agent does something positive in support of her conviction, is this really an essential component of the communicative principle? Brownlee remarks that the non-evasion requirement is a good practical test for conscientiousness. This sounds right; the fact that an agent engages in positive acts in support of her asserted conviction constitutes evidence that she actually holds that conviction. But that does not make it a necessary condition or requirement.

Brownlee considers three possible objections to her claim that this positive act is necessary. First, it seems overly onerous to require that an agent pursue a burdensome option when a non-burdensome option could just as effectively communicate her conviction. Let's consider

---

the case of a pacifist drafted into military action: Suppose that fleeing the country to avoid being drafted would just as adequately communicate that pacifist’s objection to war as staying in the country and refusing to participate. Perhaps major newspapers will run a story on his departure communicating his pacifist views. Let's also suppose that this is genuinely less burdensome than staying in the US and trying to escape the draft (suppose that he won't get detained by authorities etc. in Canada). Wouldn't this be an acceptable way of communicating his conscientious conviction? Wouldn't such an agent still count as conscientious? According to the non-evasion requirement, as stated by Brownlee, this behaviour would certainly not count as communicative or conscientious. The pacifist is not honouring his convictions as he is evading the negative consequences of his behaviour (being pursued by legal authorities) that would result if he was to stay in the US.

Given that such an action appears to be contrary to the non-evasion condition, how could it be a necessary condition of conscientiousness at all? Brownlee accounts for this by maintaining while non-evasion is a necessary condition, like the consistency requirement, it is a condition with a *ceteris paribus* clause. When things are not equal, the agent does not need to honour her convictions. As in the case of the consistency requirement this raises the question of when, if ever, *ceteris paribus* clauses are legitimate. Again, this is the focus of chapter four, section seven.

Another objection is that it is asking too much of an agent that she engage in such burdensome conduct in order to count as conscientious. It is simply too stringent a requirement on an agent that in order to be conscientious she must act positively in support of her conviction. One way of framing this objection is to point out the ways in which the non-evasion condition is contrary to the common use of the term 'conscientious'. In the introduction to this chapter I pointed out that a common use of the term 'conscientious' occurs in describing professionals who morally object to doing something that would otherwise be part of their job. In almost all cases, this involves a *refusal* to do what is, in some sense, required of the agent: Whether correctly or incorrectly, we describe the pharmacist who refuses to dispense the emergency contraceptive pill on moral grounds as someone who conscientiously objects to dispensing it. We class the clinician who refuses to perform abortions on moral grounds in a similar way. Whereas the cases of civil disobedience that Brownlee holds to be paradigm cases of conscientiousness aren't often described using the term 'conscientious'. Consider the example Brownlee used of the assistive disobedient; the clinician who illegally performs euthanasia. I suspect we would be less inclined to call this
more 'active' form of disobedience 'conscientious' than we would the more passive form that is distinctive of personal disobedience.

Whether or not people do in fact refer to some practice using the term 'conscientious' and they do not use that term in other cases is not, of course, determinative. This is even more pressing given that, as I have repeatedly pointed out, the sort of conceptual analysis we are engaged in here is fundamentally normative. However, the fact about linguistic usage of the term 'conscientious' does constitute evidence in support of the idea that the 'positive act' element of conscientiousness is too strong. It is important for Brownlee to demonstrate that the positive element of the non-evasion condition is required as this is one of the central reasons for thinking that civil disobedience counts as conscientious whereas personal disobedience does not.

In support of the idea that no positive act is required in order to count as conscientious, consider some common locutions associated with conscientious conduct: Sometimes as part of her rationale for acting in accordance with her conscience rather than acting in accordance with the law, the agent reasons 'I just couldn't live with myself if I were to do that'.\textsuperscript{41} Such locutions make no reference to doing something positive in support of one's actions. In order to act in accordance with conscience, on this reading, requires only that the agent \textit{not} do certain things, not that she get others to share her view.\textsuperscript{42}

To bolster the argument for the positive element, Brownlee offers the remark I mentioned earlier: we would place little weight on the 'asserted convictions' of someone who was not at least sometimes willing to prevent others from doing what she judges to be wrong. Let's evaluate this claim: why would we place little weight on such a person's convictions? One possibility, as indicated by the use of inverted commas, is that to claim conscientiousness without sometimes performing some positive action in support of that conviction is

\textsuperscript{41} As I discussed in chapter two, Hannah Arendt emphasised this locution when forming her analysis of conscience (1978). But it is a locution which seems to motivate many of those theorists who have emphasised the connection between conscience and identity or integrity; for example, Wicclair (2012), Sulmasy (2008), May (1983) and Lyons (2009).

\textsuperscript{42} Again, as I discussed in chapter two, it is an interesting aspect of Arendt's account of conscience (and perhaps a flaw of that account), that it treats the conscience as only requiring the agent refrain from doing certain things, rather than positive action.
incoherent. In making such an assertion the agent reveals that she does not understand the straightforward descriptive meaning of the word 'conscientious'. But the claim certainly doesn't seem incoherent. A better interpretation of Brownlee's claim, perhaps, involves reference to the normative component of conceptual analysis. Having a conscientious conviction is not simply a matter of satisfying some of the descriptive constraints on the meaning of the word 'conscientious'. In order to have a conscientious conviction an agent must also have a conviction that is \textit{worth} something. So we can interpret the claim in the following way: an agent must be willing, at least \textit{sometimes}, to do something positive in support of her convictions, for her conviction be worthy of respect in the way that conscientious convictions are worth respecting.

On the interpretation just given, is Brownlee's claim true? That depends on what the normative grounds for conscientiousness are. Let's assume that it is more admirable when an agent acts positively in support of her conviction than when she fails to so act. It doesn't automatically follow that there is a requirement of conscientiousness for two reasons: One, such behaviour could be supererogatory with respect to conscientiousness: It's good for an agent to so act, but it is not a moral requirement of the conscientious agent. Two, the normative value of a conscientious act need not be exhausted by the conscientiousness of that act. We can imagine a conscientious act being good in ways that have nothing to do with the act's conscientiousness. Consider the following, somewhat contrived, example: There are two ways, perhaps, in which an observant Catholic counsellor could communicate her objection to abortion. She could do so by spending her time protesting outside abortion clinics and writing letters to the government in order to protest the provision of abortions, or, she could do so by offering free counselling services to women considering abortion to help them see the viability of having a child. Let's say that both actions are equally communicative and therefore, according to Brownlee, equally conscientious. It seems plausible to me to say that the second option is, all-things-considered, the morally better option because it is beneficent or compassionate in a way that the other option is not. In order to answer come to a determinate answer in such cases we need to take a step back and consider what the particular normative grounds of conscientious conviction are.

Where have we got to so far? In this chapter I shifted focus from historical accounts of conscience to a contemporary account. Instead of emphasising the moral or identity components of conscience, Brownlee emphasises the \textit{communicative} nature of conscientious convictions and actions. In the first section I described Brownlee’s distinction between
different types of disobedience such as civil and personal disobedience. Then, in section two, I began my assessment of the communicative principle used by Brownlee in order to argue that acts of personal disobedience are less conscientious than acts of civil disobedience. I start with the consistency requirement: the requirement that an individual act consistently with her conscientious convictions. I suggest that while this is plausibly a requirement of conscientiousness, it is less clear that this is an element of the communicative principle *per se*. And ultimately, I suggest, the status of consistency as a requirement of conscientiousness depends on the legitimacy of *ceteris paribus* clauses; a matter I come to a final determination on in section seven of the next chapter. In section three, I considered the universality requirement. Again, like the consistency requirement, I suggest that the legitimacy of this requirement depends entirely on whether the putative exceptions to that requirement are best dealt with by a *ceteris paribus* clause. In chapter four I argue that all of Brownlee’s *ceteris paribus* clauses should be jettisoned in favour of treating communicativeness as a weighty, but non-necessary, desideratum of conscientiousness or conscience. Finally, in section four I considered the non-evasion requirement. I suggested that whether or not this is a requirement of conscientiousness depends entirely on what the normative grounds of conscientiousness are. In light of this, I begin the next chapter on that very subject.
Chapter Four: The Communicative Principle and the Normative Heads of Conscience

The primary normative consequence of an agent possessing a conscientious conviction is that the presence of such a conviction affords that agent certain moral rights. Most importantly, perhaps, it affords a right of conscientious action; a right for the agent to act on her conscience, irrespective of what the law demands.\(^1\) Though, importantly, this right is a defeasible one. Whether or not the individual is, all things considered, entitled to act on their conscience will depend on the circumstances of their case and is beyond the scope of this thesis. For example, I hold that a clinician has a moral right to refuse to be involved in the provision of abortions where that would be contrary to her conscientious judgement. But I do not argue that, all-things-considered, a clinician is permitted to act on that judgement. Plausibly, a patient’s right to receive the clinical intervention of her choosing trumps that clinician’s right. My focus in this thesis is purely on a *pro tanto* right of conscience.

In this chapter, I continue where chapter three left off. At the end of that chapter we were at the point where we needed an explanation of what is valuable about conscience in order to explain how the non-evasion requirement could be a requirement of conscientiousness. The first three sections of this chapter are devoted to explaining what exactly is valuable about conscience. I argue there are three normative ‘heads’ of conscience, which potentially justify a moral right of conscience: the humanistic head, the psychological benefit head and the societal value head. Each head contributes to our understanding of how it is that an individual might possess a moral right of conscience. In brief, an individual has a moral right to act in accordance with her conscientious conviction as those convictions are the expression of her

---

\(^1\) It is this right that I refer to when I speak generically in this thesis of a ‘right of conscience’. There is also a different potential right of conscience: What Brownlee calls a ‘moral right to inner control and free thought’ ((2012), p128). It is this right, perhaps, that seems to be envisaged in various human rights instruments (such as the Universal Declaration of Human Rights Article 18: “Everyone has the right to freedom of thought, conscience and religion”) where a distinction is made between the right to freedom of thought and the right to manifest those thoughts in actions. Arguably a right to act on one’s conscientious conviction is of little use if the agent doesn’t also possess a right to freely form one’s thoughts. The nature of this right is beyond the scope of this thesis and any reference to a ‘right of conscience’ should be taken to be a right of conscientious action.
self-determined ideals (whether moral or otherwise), acting in accordance with one’s ideals is psychologically beneficial and such action is valuable to society in general.

Then in section four I return to Brownlee’s non-evasion requirement: In particular, I focus on the positive element: the requirement that an agent must do something positive in support of her actions to count as communicative and thus conscientious. In section five I consider the final putative requirement of conscientiousness offered by Brownlee, the dialogue condition. This was the requirement that an individual endeavour to enter in to reasoned dialogue with others about her convictions. Again, I agree that such a requirement is a component of the communicative principle and conscientiousness more generally. And the justification for that requirement, just like the justification for the non-evasion requirement can be grounded in the normative heads of conscience. I suggest that, in fact, all of the requirements of the communicative principle (including the consistency and universality requirements) can be justified by reference to those three normative heads.

In section six I return to the distinction between civil disobedience and personal disobedience. I argue that in terms of the normative heads of conscientiousness, personal disobedience is just as conscientious an act as civil disobedience, due to the presence of ceteris paribus clauses mitigating the stringency of Brownlee’s four requirements.

Finally, in section seven, I consider a possible objection to this account of conscience. Personal disobedience only counted as conscientious in virtue of ceteris paribus clauses which in many cases are methodologically dubious. In this section I suggest that we eliminate ceteris paribus clauses in the communicative principle by denying that this principle is a necessary condition of conscience at all. Instead, I suggest that we treat communicativeness as a non-necessary but weighty factor in determining whether a given conviction or act is conscientious. Not only is the communicative principle a weighty principle of conscience I suggest that it be supplemented with principles of morality and identity. These three principles together constitute the ‘criteria’ or desiderata in terms of which any conviction or action ought to be assessed for its conscientiousness.
4.1. The Humanistic Head of Conscience

What is the function of the 'right' in the moral right of conscientious action? According to one popular analysis of rights, it functions to protect the interests of the right-holder.\(^2\) I assume that while it is necessary that a right protect an individual’s interests it can also function to protect or promote the interests of society.\(^3\)

Which interest of the agent precisely is supposed to be protected the moral right of conscientious action? Brownlee argues that is "rooted in a humanistic principle of respect for our agency and dignity as persons"\(^4\) This is respect for our status as reasoning, feeling beings capable of forming moral convictions, convictions that often have a non-evasive and dialogic component that we wish to express.\(^5\) Some version of this principle is defended by several other authors such as Ronald Dworkin, Joseph Raz and Martha Nussbaum, though often articulated slightly differently. In Taking Rights Seriously, Ronald Dworkin claims that to treat a person with respect is to treat them “as human beings who are capable of forming and acting on intelligent conceptions of how their lives should be lived”.\(^6\) Dworkin’s formulation is slightly more inclusive than Brownlee’s. Plausibly an individual’s moral convictions implicated only a subset of the individual’s beliefs about how life should be lived. We might say that while Brownlee’s emphasises the moral convictions of the agent, Dworkin emphasises the ethical views of the individual.\(^7\) This distinction may turn out to be important.

---

\(^2\) As an example of this account consider Neil MacCormick:

> The essential feature of rules which confer rights is that they have as a specific aim the protection or advancement of individual interests or goods ((1977), p192)).

This approach to rights can be contrasted with the will theory. This theory holds that the function of a right is to give the holder power over the duties of another (e.g. Hart, (1982), p183).

\(^3\) Though plausibly the protection of an individual’s interests is necessary for it to be a right at all.


\(^5\) (Ibid.).

\(^6\) (1977), p272.

\(^7\) For now I will assume that the mark of the moral is that it is categorical; moral imperatives apply whether or not that imperative aligns with the individual’s desires or interests. Whereas the ‘ethical’ is an Aristotelian notion picking out a broader domain that includes moral concerns but also other matters that contribute to the goodness and meaningfulness of the individual’s life.
If a conscientious conviction necessarily involves the agent having certain moral beliefs, then
the normative principles underlying conscientious conviction must pick out the agent’s moral
beliefs as peculiarly worthy of protection.  

I will eventually argue in chapter seven that the humanistic principle grounds an identity
principle as well as the communicative principle. This idea, that there is some essential
connection between the humanistic principle and identity, is expressed in inchoate form in
Raz’s articulation of the humanistic principle in *The Authority of Law*.

Raz claims that:

> humanism calls for respecting the autonomy of persons, that is, their right and ability to
develop their talents and tastes and to lead the kind of life they are committed to. The area of
a person’s life and plans which have to be respected by others are those which are central to
his own image of the kind of person he is and which form the basis of his self-respect.

In Raz we see an account of the humanistic principle that is, perhaps, narrower than the
account of Brownlee or Dworkin. The principle pertains directly to the commitments of the
individual that are “central to his own image of the kind of person he is”. In chapter seven I
develop an account of what it is to have a self-conception or ‘image’ of oneself in
considerable detail. While I argue that the humanistic head does ground some notion of
identity, I claim that it also grounds other commitments of the individual.

There isn’t much of a positive argument that needs to be given for the humanistic principle.
Insofar as one embraces liberalism, one should accept the value in individuals having the
freedom to determine for themselves what they should do rather than to have their actions
and convictions mandated by someone else. While the principle might itself be intuitive, so
too are some objections to that principle. One potential ground for objection is that our
convictions will often conflict with the convictions of others. How can we have a right to act

---

8 If the principle did not make any reference to the agent's moral convictions, then the principle in question
would not necessarily ground protection of an individual’s moral convictions any more than it would protect her
mere preferences. I shall argue at the end of this chapter that, in fact, the humanistic head doesn’t just ground
moral convictions but can, in fact, ground non-moral communicative and identity-implicating convictions.

9 Raz (1979), p280.

10 This thought is, perhaps, best captured in John Stuart Mill’s ‘harm principle’. Mill claims that liberty allows
“framing the plan of our life to suit one’s own character” as long as it does no harm to others ((2010), p11).
on our own beliefs, values and convictions when others are equally entitled to act on beliefs, values and convictions that may conflict with our own? A second potential problem is the fact that some of our convictions may be flat-out contrary to our obligations to others. The standard way in which these problems are resolved is by treating this as a ‘conflict of rights’ case. The moral right we are talking about here is a defeasible right, and sometimes it will be all-things-considered wrong for an agent to act on that right. To work out what the agent ought to do, all things considered, requires working out which moral considerations defeat others.11

Brownlee claims that there is a stronger connection between the humanistic principle and our obligations to others than the connection I have described. Brownlee claims that the humanistic principle must be subject to an egalitarian constraint; in other words, the humanistic principle requires that only those convictions which are consisted with the needs and dignity of others.12 By way of example, Brownlee claims that it is not in accordance with the humanistic principle for a civil registrar to refuse to marry homosexual couples as it does not respect the dignity of those couples. In fact, not only would refusing to marry such couples be inconsistent with that principle, so too would referring that couple on to a willing registrar as the referral would be discriminatory.

The notion that an agent’s respect for her own humanity requires respect for the humanity of others is commonly accepted. It is most famously captured in Kant's categorical imperative. Recall that the universal law formulation of the categorical imperative requires that we "act only in accordance with that maxim through which at the same time you can will that it become a universal law".13 If I judge that my own dignity ought to be respected, then I must also judge that the dignity of all others ought to be respected. A recent statement of this view is offered by Christine Korsgaard in Sources of Normativity where she holds that "valuing ourselves as human beings involves valuing others that way as well, and carries with it moral

11 This ‘ad hoc’ approach holds that rights can potentially conflict with each other (and with other normative considerations), and it is then a matter of working out what, all things considered, ought to be done. This might be contrasted with the definitional approach which would define the scope of the right in terms of what all things considered ought to be done. The comparison between these two approaches is common in constitutional legal theory (e.g. Ahdar & Leigh (2013), p192).

12 (2012), p149.

13 Kant, (2012), 4:421.
obligations." However, even if we were to accept that there is a strong connection between respecting one's own humanity and respecting the humanity of others (as suggested by these authors) this does not have the implication that Brownlee suggests; it does not follow that our own convictions need to be curbed by the interests of others.

There are several reasons why I deny that the humanistic principle has these implications (I use the terms ‘principle’ and ‘head’ interchangeably): First, when we make a claim of conscience, it is often in the context of disagreement about what dignity requires or whose dignity is relevant. There is certainly a sense in which the conviction of the reluctant registrar fails to respect the dignity of the gay couple. On one influential account of dignity, Kant’s account, respecting the dignity of persons involves respecting their ability to set their own ends. On this interpretation, the reluctant registrar fails to respect the dignity of the gay couple by disregarding the fact that their decision to be civilly partnered is a consequence of their end-setting capabilities. But, of course, the registrar may well disagree that this is what dignity requires; perhaps she is a natural law theorist and thinks that the gay couple in getting civilly partnered would be acting contrary to their natural ends. An account of conscience must allow for an individual to have a conscientious conviction in cases where people disagree about what dignity requires, otherwise people would very rarely have conscientious convictions.

A similar problem arises if one defines the egalitarian constraint in terms of ‘needs’ rather than dignity. Perhaps the gay couple could argue that companionship is a basic human need and they cannot get the full societal benefits of companionship without being able to be civilly partnered. But even if this is a basic need, recognising such an entitlement should not be a pre-requisite to possessing a conscientious conviction. I think that there is another important reason why there shouldn’t be such an egalitarian constraint. It would sever the connection between the humanistic principle and an agent’s personal autonomy. The humanistic principle recognises that there is something valuable in an agent using her

\[\text{Reason thus refers every maxim of the will as universally legislating to every other will, and also to every action towards oneself, and it does so not for the sake of any other practical motivating ground or future advantage but from the idea of the dignity of a rational being that obeys no other law than that which at the same time, it itself gives} (2012), p46, 4:34.\]

---


15 For this interpretation of Kant see Korsgaard (1986). For a passage in Kant that is suggestive of this interpretation consider:
capabilities to arrive at an ethical conviction. What is valuable is the free exercise of those capabilities, not simply exercising of those capabilities which result in the correct ethical or moral convictions. The capacity to come to one's own conclusions on such matters, the *freedom* to do so, is plausibly good in itself.

Finally, one might argue that an egalitarian constraint is inconsistent with the societal value of conscientiousness. I argue in section three of this chapter that one of the heads of conscience pertains to the societal value of an individual acting on her conscience. I discuss there an argument that the presence of a false moral beliefs in the public sphere prevents us from forgetting the reasons for our (correct) moral beliefs: If there is not at least some representation of false beliefs in the public sphere then there is a risk that the correct moral views will become dead dogma; views that we accept on blind faith rather than on the basis of reasons which support those beliefs.

In light of these considerations, I don't think it is plausible to subject the humanistic principle to an egalitarian constraint. We should accept that an agent has a moral right of conscientious action no matter what the content of the moral belief contained in that conviction. The only intuition which leads us to support such a constraint is better captured not by watering down the humanistic principle but by pointing out the circumstances where some moral considerations normatively defeat the *pro tanto* right of conscience.

In grounding the right of conscientious action in the humanistic principle we attest to the value in humans being free to develop their capacity to form moral convictions and act on them. Another way of putting this is that to deny agents a right of conscientious action is to place an absolute and unjustified amount of weight on the value of being a law-abiding citizen. Even if we believe that agents ought to always follow the law, we should only accept that this is true because some normative considerations defeat the value of autonomous action. We should not do so on the basis that such action is of no moral value.

The humanistic principle, I suggest, is best interpreted as a statement of the value of an agent acting on convictions that they have formed for themselves. Even though we have been dealing primarily with moral convictions, they could be convictions of a different sort as well. At the end of this chapter I will consider some of the other sorts of conviction, besides moral convictions, that might be protected or promoted by the humanistic principle.
4.2. The Psychological Benefit Head of Conscience

The humanistic principle is not the only normative basis for the right of conscientious action that Brownlee offers. Brownlee explores two other possibilities. The second possibility concerns the psychological consequences, for the agent, of never being permitted to act in accordance with her convictions. In the introduction I mentioned some theories of conscience that make reference to the negative psychological impact of conscience. For example Nietzsche and Freud both emphasised the phenomena of bad conscience: the guilt that we feel as a result of self-punishment. Both placed great weight on the psychological impact of conscience, in the form of guilt, for the agent. In this section I consider not only the psychological distress of being forbidden from acting on one’s conscience but also the positive psychological benefits of being permitted to so act.

Let’s call the view that agents are never allowed to act in accordance with their convictions, ‘slavish law-abidingness’. There are different ways in which slavish law-abidingness is, in my view, too psychologically costly for the individual. The first way is that it forces the agent to experience the psychological anguish, the guilt, of acting contrary to her moral convictions. On any plausible view of when agents are allowed to act on their convictions, of course, the agent will sometimes have to experience guilt. There is no plausible theory of when an agent can act on their convictions which would allow that an agent was always permitted to act on their convictions no matter what they are.

Granrose also defends this principle as a normative justification for conscience (1970). Swan & Vallier also consider the possibility that psychological distress might be the normative ground of conscience before dismissing it in favour of a grounding in social contract ((2013), p8-10). I argue that both psychological distress and contractualist reasons can ground a right of conscience.


Which does not mean that there aren’t, what are in my view, implausible theories to that effect, consider for example, Robert Paul Wolff’s philosophical anarchism:

On the basis of a lengthy reflection upon the concept of de jure legitimate authority, I have come to the conclusion that philosophical anarchism is true. That is to say, I believe that there is not, and there could not be a state that has a right to command and whose subjects have a binding obligation to obey ((1969), p607).
Another way in which the requirement to be slavishly law-abiding may harm the agent, Brownlee argues, is that the agent’s very ability to practically reason atrophies if not regularly exercised. There is less need for the agent to agonise over what she morally ought to do, if there is a law on the matter that prescribes what we ought to do, then that is what we should do.\(^9\) A third way in which slavish law-abidingness is psychologically costly, Brownlee suggests, is that behaving in that way may have the consequence that the agent loses respect for the law, and therefore fails to respect it in the cases that really matter.\(^\)\(^\)\(^2\) It is more important to follow some laws than other ones. For example, it is more important to follow the law which proscribes murder than to follow the law that requires the filing of tax returns. Yet, if the agent always deliberates merely by reference to what the law requires, she may lose the ability to distinguish the relative moral importance of following different laws.

All the considerations mentioned pertain to the psychological costs of acting contrary to conscience. A right of conscience psychologically benefits the individual insofar as she is not subject to these significant psychological costs. But there might be an even stronger argument as to the psychological benefits of acting on conscience insofar as this autonomous action might enhance the individual’s well-being.\(^\)\(^\)\(^2\) As an example of what this argument might look like, consider the claims of Timothy Macklem in defending a right to act on one’s religious beliefs:

> What is more, their opportunity to exploit their knowledge of what is good for them, by acting on their best sense of what is good for them, is a fundamental part of what it means for them to be authors of their own lives, a role that I take to be central to the well-being of every person.\(^\)\(^\)\(^2\)

Here, we see, as with Dworkin, reference to the ethical nature of the convictions that the humanistic principle can ground, the agent’s ‘sense of what is good for them’. According to Macklem, acting on one’s convictions is not only good simpliciter (as suggested by the humanistic principle), but good for the agent herself as it contributes to her well-being. When

---

\(^9\) Brownlee (2012), p167. Perhaps, though, this point should not be over-stated. Most of the day-to-day practical reasoning that we engage in involves deliberation about matters that the law says nothing about.

\(^\)\(^\)\(^2\) (2012), p168.

\(^\)\(^\)\(^2\) Brian Leiter has recently endorsed such an argument in (2013), p17.

\(^\)\(^\)\(^2\) Macklem (2006), p129.
an individual gets to act on her conscience rather than being forced to do whatever society requires of her, this will, contingently, make her life go better. What exactly an individual’s well-being consists in is controversial. But one prominent account, hedonism, holds that, fundamentally, well-being is constituted by pleasure. So one way of thinking about how conscientious action contributes to an individual’s well-being is the sense in which that action makes the individual’s life incrementally more pleasurable.

Is there any connection between the humanistic principle and the psychological benefit principle? Both emphasise the badness or wrongness of slavish law-abidingness. The humanistic principle identifies the sense in which this would involve a strong form of compulsion on the agent; her personal autonomy would not be respected by a requirement of slavish law-abiding. We might say that the humanistic principle pertains to what is intrinsically wrong about curbing someone’s autonomy, whereas the psychological benefit principle pertains to what is instrumentally wrong about curbing such action. Or putting it slightly differently, the humanistic principle concerns deontological considerations that support a right of conscience and the psychological benefit principle emphasises consequentialist considerations.

Interestingly, the psychological costliness of such a rule (of slavish law-abidingness) would not only harm the agent in question but would also, according to Brownlee, have negative consequences for broader society. It could potentially reduce the amount of practical reasoning going on in society and also, somewhat paradoxically, mean that agents don’t end up acting in accordance with the law as often. In the next section I move on to considering this societal value of conscientiousness. I realize that it is relatively unusual to think that an individual has a ‘right’ to do something because that sort of action is of interest to broader society rather than merely protecting the interests of the individual, nevertheless I suggest, that such considerations can ground a moral right as long as they are combined with considerations that protect the interests of the individual. In section three I will also consider some of the other respects in which recognising a right of conscience would be of societal value.
4.3. The Societal Value Head of Conscience

I have just mentioned one way in which a right of conscientious action, potentially, facilitates both a private and a public good. It facilitates a private good insofar as acting in accordance with conscience psychologically benefits the individual. It facilitates a public good insofar as the failure to accord such a right to individuals has broader negative effects on society. In this section I want to consider some other ways in which a right of conscientious action has public or societal value. Brownlee calls this a 'double-harmony' between the individual's own interests in acting on her convictions and the broader interests of society. It is good for society if individuals are skilled in practical reasoning. It is good for society if individuals can reliably distinguish the relative importance of following some laws over others. But there are other ways in which a right of conscientious action is conducive to the public good. This is the value to society of out of having various different views, as represented by conscientious acts carried out in the public sphere. Brownlee mentions several possibilities here: One, a conscientious conviction could represent views which are not represented in the media and thereby encourage those who defend the law as it stands to articulate the reasons that exist in favour of that law. Two, if society is wrong it provides a good opportunity for society to change its views. This societal value of conscientiousness supports my earlier argument that there are no egalitarian constraints on the humanistic principle. There is value to society (if not to the individual) in inegalitarian action. The reluctant registrar who refuses to register same-sex marriages performs an act that is valuable to society insofar as it causes people to respond 'that action is morally unacceptable, it is discriminatory'.

I think these are two plausible ways in which a right of conscientious action is valuable to society and they mirror two of the arguments that John Stuart Mill offered in support of society tolerating the expression of a variety of different views. Since these arguments

23 (2012), p146. This notion was originally articulated by Raz (1986).

24 Of course, this is not to say that, all up, the registrar was entitled to do as she did. To say that a particular conscientious conviction is societally valuable is not to say that, without qualification, the individual should act on that conscientious conviction.

25 Of such a right Mill says:

If the arguments of this present chapter are of any validity, there ought to exist the fullest liberty of professing and discussing, as a matter of ethical conviction, any doctrine, however immoral it may be considered (Mill, (2010), p14).
concern the extent to which the true moral views are represented in broader society and the
presence of justification for those views, it might be thought these are all arguments for the
epistemic, rather than the moral value of conscientious action. They are reasons why
conscientious action can help move society towards the truth. But, as Brian Leiter has pointed
out, this type of argument is ultimately an argument about what is morally valuable as well as
it is morally valuable for society (particularly society as represented in the state) to possess
the correct moral views.\textsuperscript{26} A society that has the correct views about what is morally good are
more likely to do what is morally good. I suggest that all of Mill's arguments in support of a
right of freedom of expression can also be seen as arguments in support of a right of
conscientious action. This is not to say, that the arguments are \textit{exactly} analogous in both
cases. For example, speech acts don't have the same effects on others as actions, so one might
be more inclined to accept an argument supporting the toleration of the former than the latter.
I think those differences, however, are best accommodated not by suggesting that
conscientious action has less of a claim to being a ‘right’ as free speech, but simply by
holding that it is more common for individuals to be morally restricted from \textit{acting} on their
right of conscientious action.

I will consider each of Mill's arguments in turn: First, the law might simply be wrong.\textsuperscript{27} I
mentioned that one consequence of an agent acting on her (correct) conviction would be that
it increases the chance that society will change, the law will change, and adopt the correct
view. But also, even if acting on that correct conviction does not result in society or the law
changing its view, it still means that the amount of right action in the world is increased. It is
an even stronger consideration in support of the idea that an agent should act on their
convictions rather than simply express them in speech. While it may be good to express an
opinion that the law is wrong, but then simply do as the law requires of you, it is even better
to actually do the right thing.

Mill's second argument is that the law might be incorrect insofar as it only contains part of
the truth.\textsuperscript{28} An individual’s conscientious action can help change the law such that it is
completely morally correct. For example, let's say there was a law proscribing all abortion on

\footnotesize
\textsuperscript{26} Leiter (2013), p15.
\textsuperscript{27} Mill, (2010), p14.
\textsuperscript{28} (2010), p31.
the grounds that embryos are sentient. And let’s assume for the sake of argument that possession of sentence is a necessary condition for being morally entitled to live; there is nothing wrong with killing a non-sentient being. Assume also that a clinician believes (correctly and on good epistemic grounds) that a foetus is not in fact sentient. When asked to provide an abortion to a patient, the clinician does so. Perhaps there are several morally good features of this clinician’s action. But one good-making feature of his action, I think, is the way in which it could communicate to broader society that the law was half-wrong. The hypothetical law is incorrect insofar as it is based on incorrect facts. The embryo is not sentient. But it is correct in so far as it gets the criterion of a being's moral worth correct. The being is worth protecting just in so far as they are sentient.

Mill's third argument is that even if the law is correct, it being contested by others, and the presence of that disagreement prevents the law from becoming 'dead dogma'; something that is believed on the basis of mere prejudice without understanding of the rational grounds that support the law.\(^29\) This was the second reason offered by Brownlee as to why conscientious action is socially valuable. Consider the following case: Imagine that euthanasia is morally permissible and that we are in a jurisdiction where it is legally permissible. It is one benefit of some agents conscientiously opposing that law that citizens would be reminded of the rational justification for euthanasia. And perhaps this would ensure that the safeguards surrounding the practice of euthanasia are fiercely protected. This point coheres nicely with Brownlee's argument about the society effects of the psychological cost of slavish law-abidingness. All laws become 'the same' in some sense. The agent becomes incapable of distinguishing the laws which are really important to follow and the laws that are less important to follow.

Mill's fourth argument, related to the last one, is that the very meaning of a law can be lost if it is not subject to regular scrutiny.\(^30\) It might not be immediately obvious how this argument differs from the previous one. To illustrate, consider the following hypothetical case: As a society we decide that euthanasia should be legally permissible. We do not do so on the basis that human life is completely unworthy of respect. Rather, we retain respect for human life, but believe that in some cases the extreme pain for an agent means that on those occasions, at

\(^{29}\) Mill (2010), p25.

\(^{30}\) Mill (2010), p35.
the request of the agent, a clinician is allowed to help the agent to die. Perhaps, after this law has been in force for a certain period of time there is a risk that society forgets what respect for life involves. It is not just that we forget why this is a case where the principle of respect for life is defeated by other moral considerations; it is that this will eventually have the effect of weakening the very principle of respect for life. The idea of 'respect for life' will no longer have the normative ‘oomph’ that it once did. Clinicians refusing to carry out euthanasia on the grounds that it did not respect life might counteract this possibility.\textsuperscript{31}

These Millian considerations all capture, in various ways, the\textit{ democratic} value of conscientious action. Conscientious action is another way of achieving what free speech aims to achieve, though arguably, conscientious action is an even more powerful way of communicating the agent's views than mere speech. It is the\textit{ democratic} value of conscientiousness because it describes the way in which a democratic society can benefit from agents acting on their conscientious convictions. But, I suggested, the societal value of conscientiousness is not exhausted by the extent to which it aids democratic aims. For example, if the agent's moral conviction is correct and the law is incorrect, then part of the public value of the agent acting on that conviction lies simply in the fact that it means there is more right action happening than otherwise would be.

Another way in which a right of conscientious action might have democratic value is if such a right is justified by the\textit{ social contract}. Rawls argued that equal liberty is one of the two fundamental principles of justice that rational individuals would choose in the ‘original position’; that is, the situation where individuals are choosing the basic principles of justice for governing their societies but they are unaware what position they will have in that society and what their beliefs and preferences will be. Rawls claims:

\begin{quote}
31 Alasdair Macintyre offers a good example of a concept that has lost its meaning by falling into disuse, with the ancient Greek conception of the good:

The society reflected in the Homeric poems is one in which the most important judgments that can be made about a man concern the way in which he performs his assigned social function (Macintyre (1966), p4).

The modern notion of goodness has lost its essential reference to the social position of the individual in question.
\end{quote}
Equal liberty of conscience is the only principle that the persons in the original position can acknowledge. They cannot take chances with liberty by permitting the dominant religious or moral doctrine to persecute or suppress others if it wishes.\(^{32}\)

According to Rawls, the reason that we ought to respect an individual’s right of conscience is because, at bottom, there is wide disagreement in society in terms of ethical, moral and religious doctrines. If we were trying to agree amongst ourselves about the fundamental rules of justice without knowledge as to which person in society we were and which views we would have, we would agree to a right of freedom of conscience.\(^{33}\)

Let's consider all the ways in which conscientious action is of societal value. First, under the head of psychological benefit we saw conscientious action cultivates the ability of individuals to practically deliberate. This is a public good because a well-functioning society requires that individuals have this capability. Second, also under the head of psychological benefit, a well-functioning society requires that agents can distinguish those laws which it is more important to follow from those laws which it is less important to follow. A right of conscientious conviction cultivates the capacity for practical reasoning that facilitates making that distinction. A third psychological benefit of conscientious action is that the capacity of an individual to act on her own ends enhances the individual’s well-being and thereby increases the well-being of all individuals in society more broadly.

I then considered the Millian arguments in favour of the societal value of conscientious action. First, there was the argument that the conscientious action could be the right action and the action prescribed by the law could be wrong. This places the correct view before society so that, potentially, they can change their perspective. Furthermore, if the law is wrong, then the agent by acting contrary to that law is increasing the amount of right action going on. Second, if the law is partially incorrect, the agent acting on her conviction has the potential to rectify that and make the law completely correct. Third, if the law is correct, the agent, by acting on her conviction that is contrary to the law, acts as a reminder of the rational justification for that law and thereby decreases the chance that that law becomes ‘dead dogma’. Fourth, and related to the previous point, an agent's subversive but

\(^{32}\) Rawls (1971), p206. This argument in favour of a right of conscience has been defended recently by both Leiter (2013) and Swan & Vallier (2013).

\(^{33}\) Rawls also has very particular views about when exactly the individual is manifest that liberty in action through civil disobedience (1969).
conscientious action is a buffer against the moral principles that underlie the law, losing their meaning. If society does not regularly refer to and engage with a moral principle, then that moral principle ceases to mean what it once did. Finally I considered a contractualist argument: as we all possess such convictions and we cannot easily discard them, we would agree to a right of conscience in the original position.

To summarise: In this section I have considered three heads under which we can categorise the normative basis of the moral right of conscientious conviction: the humanistic head, psychological benefit and societal value. Together these different normative heads identify both the public and private interest that is enabled by that moral right. It is a private good insofar as acting on a conscientious conviction is an exercise of personal autonomy and preserves an agent's psychological integrity. It protects a public interest insofar as society functions better when agents (sometimes) act on their convictions. The three normative heads are not mutually exclusive; the psychological benefits to the agent are good for that agent as well as broader society. It is also worth mentioning that none of these considerations depend for their normative weight on whether an agent is always allowed to act in accordance with her conviction. It is implausible that it would always be good for society if agents always acted contrary to the law when it suited them. But it is enough to justify a right of conscientious action that sometimes those actions would be good for society. The fact that it would be good for society if agents generally acted as the law requires of them is simply a reason why the moral right of conscientious action is often defeated by other normative considerations.

4.4. The Normative Standing of the Communicative Principle

My initial purpose in discussing the normative heads of conscientiousness in this chapter was in order to explain how the positive element of the non-evasion requirement could be justified. The positive element required that the individual who honours her convictions not only fails to evade the consequences of her convictions but also does something positive in support of her convictions. In this section I will demonstrate how each requirement of the communicative principle can be justified by reference to the normative grounds discussed in the previous section.
I start with the non-evasion requirement. According to Brownlee, the reason why the positive element is an essential component of the non-evasion condition is not because there is something incoherent about an agent who satisfies only the negative element and not the positive element, but rather that the conviction of an agent who doesn't do anything positive in support of her action is not worth protecting. Do any of the three normative heads justify this positive element of the non-evasion requirement?

Let’s begin with the humanistic head or principle. Does the humanistic principle provide a justification for the positive element? No, it is perfectly consistent with acting on one’s own ends (whether ethical or otherwise), for that agent not to do anything positive in support of her conviction. What about the principle of psychological benefit? Again, I don’t think this justifies the positive element: An agent would experience psychological anguish if she can’t act in accordance with her convictions irrespective of whether she has been prevented from doing something positive in support of her conviction. What about the societal value of conscientiousness? This is the best contender for a rational justification for the positive element of non-evasion. It is positive acts on behalf of the agent, in support of her conviction, that place the agent’s moral views in the public sphere. If the agent is correct, this may serve as a prompt for society (and the law) to change its position from the incorrect view to the correct view. Even if the agent’s conviction is incorrect, her doing something positive in support of her conviction will serve as a reminder of the justification for the view incorporated in the law and reduce the chance that that law will become ‘dead dogma’.

It might be argued that this requirement that the individual do something positive in support of her action in order to count as conscientious is too stringent. This stringency can be mitigated, however, by making the non-evasion condition subject to a ceteris paribus clause just like the consistency requirement (as discussed in chapter three). In adding a ceteris paribus clause to the non-evasion requirement, Brownlee acknowledges that sometimes the individual does not have to do something positive in support of her conviction. Brownlee suggests that the agent is not required to satisfy either the negative or the positive elements of her conviction in order to count as conscientious when it would be overly burdensome for her to do so. Brownlee also mentions the possibility of an agent satisfying this condition when she does nothing positive in support of her conviction but it is a situation where the individual’s motivation is a morally reasonable one. For example, if one is a vegetarian and a

34 Brownlee (2012), p44.
guest for dinner, it may not be appropriate to try and convince one's hosts of the moral correctness of vegetarianism. More than that, it may even be appropriate for the agent to pretend that she is not a vegetarian (out of respect for one's hosts). What matters is that the motivation for the evasive conduct is a moral motivation.

One might argue that these cases, cases where acting non-evasively is too burdensome and where the agent's non-evasion is morally, rather than selfishly motivated, can meet the non-evasion requirement without any need to refer to the *ceteris paribus* clause. The requirement is only that the agent *sometimes* do something positive in support of her conviction. If, on any given occasion, it would be too burdensome to do so or moral considerations direct one not to do so, that is acceptable, as long as the agent *sometimes* does so. While the vegetarian may not be required to try and change people’s minds on meat eating while a guest at dinner with meat-eaters, sometimes she is obligated to try and change people's minds.

I think that it is, *prima facie*, necessary to attach a *ceteris paribus* clause to this requirement because we can think of cases where it would, arguably, be inappropriate for an agent to ever 'honour' her convictions in the way the non-evasion requirement would suggest. Consider that common case of putative conscientiousness; the clinician who refuses to provide or be involved in the provision of abortion. Arguably, there are good reasons for that clinician to *never* do anything positive in support of her conviction. There are reasonable moral reasons that a clinician might have *not* to reveal her convictions to a prospective patient. When a patient is deciding whether or not to have an abortion she is there to benefit from the clinician’s clinical expertise; she is not (I would suggest) there to receive a moral lecture. The clinician has no special expertise when it comes to what is right or wrong. And given that, I would suggest that it is disrespectful for a clinician to inform the patient in any detailed way about her moral beliefs that lead her to oppose abortion. It is not acceptable for a clinician to tell a vulnerable patient that, in the clinician's opinion, abortion is murder and give this as the reason for refusing to provide the abortion. It is sufficient, I would suggest, for the clinician to simply inform the patient that she is unable to assist her for non-medical reasons, and then refer that patient on to someone else that the clinician does not believe to have a moral objection to abortion.

The natural response to this case is the one just mentioned: The positive element required only that the agent *sometimes* do something positive in support of her conviction. The case of
dealing with a vulnerable patient is simply one of the many situations where the clinician shouldn't do something positive in support of the conviction. There are plenty of other opportunities for the agent to do so. I don't think this response works. There may be good moral reasons for the clinician never to air her views in public. Would it be any more appropriate for such a clinician to protest outside a family planning clinic in support of her conviction? Arguably, such behaviour would be unprofessional; conduct unbecoming a physician.

The claim that I have just made, that a health professional should not airing her controversial moral views in public, is a bold one, and needs a much stronger justification that what I have given. However, I would argue that as long as that argument has some merit, then the positive element seems to require a *ceteris paribus* clause. As long as it would be reasonable for a clinician to think that she should never do anything positive in support of her conviction then it should not be a *requirement* that the clinician do so. A *ceteris paribus* clause is required, I argue, in order to capture the idea that in some cases the agent does not need to satisfy either the negative or the positive elements of the requirement.

What can we conclude about the non-evasion condition? The non-evasion condition requires that an agent honour her convictions. Honouring one's convictions, in turn, requires two things of an agent's behaviour. One, the negative element of that condition, requires that the agent not seek to evade the negative consequences of her actions. Two, the positive element required that sometimes the agent do something positive to try and convince broader society of the correctness of her moral conviction. This positive element is the more controversial aspect of the non-evasion condition and I offered an argument as to why we might endorse the positive element. The correct analysis of conscientiousness requires appealing to both descriptive and the normative constraints on conscientiousness. It is an essential normative constraint on conscientiousness that it is behaviour worthy of being captured by a right of conscientious action. The justification for that right, the interests protected by it, consists in the humanistic principle, psychological benefit and societal value. Only through the positive element are an agent's actions valuable to society in the way that conscientious actions can be. The non-evasion condition seems to require a *ceteris paribus* clause just as the consistency condition did. In the last section of this chapter I shall suggest that we do away with *ceteris paribus* clauses altogether in our analysis of conscience. The purpose of a *ceteris paribus* clause, in the case of conscience, can be achieved by simply denying that the communicative principle is a necessary condition.
4.5. The Dialogue Requirement

Let’s now consider the fourth and final condition that Brownlee claims is constitutive of the communicative principle; the dialogue requirement. This condition requires that we be willing to communicate our conviction to others in order to engage them in reasoned deliberation about the merits of that conviction.\textsuperscript{35} It is by our willingness to do so that we attest that our beliefs are sufficiently credible to receive a reasoned defence and that those views are worth defending despite the potential risks to ourselves. What can we make of this condition? A plausible first thought is that this condition doesn’t really add anything that wasn’t covered by the positive element of the non-evasion condition. The positive element, recall, required that the agent do something positive to get others to share her conviction.

What the dialogue condition adds, however, is a prescription as to the particular way in which the agent is required to try and get others to share her conviction. The non-evasion condition is consistent with attempting to get others to share your conviction by force. The dialogue condition adds that the agent must attempt to get others to share her view through reasoned dialogue rather than some other method.

In the description of the condition I have provided above, we also see the putative justification for the condition. If we don’t even attempt to enter into reasoned dialogue about our condition we imply to others that there is no reasoned basis for the view that we advocate. This putative justification is dubious. Just because an agent doesn’t attempt to offer an argument in support of her view, it doesn’t follow that there is no reasoned argument in support of that view. As with the non-evasion condition, Brownlee tries to bolster the case for this condition by pointing out that it is a practical test of conscientiousness.\textsuperscript{36} If an agent claims to hold a particular conviction, and then engages in reasoned dialogue to try and get others to share that conviction, then this constitutes evidence that she genuinely holds that conviction. As I argued in the case of the non-evasion condition, we might accept that the proposition provides a good practical test for conscientiousness while denying that it offers a necessary condition.

\textsuperscript{35} Brownlee (2012), p42.

\textsuperscript{36} Brownlee (2012), p42.
A potential objection to this condition is that there are sometimes better ways of communicating our conviction than by talking about it. Returning to the Rosa Parks’ case; by merely refusing to give up her seat, Parks’ didn’t seem to be engaging in reasoned dialogue with her interlocutors, but doesn’t that action still clearly count as a conscientious? And perhaps, more to the point, it seems to count as a communicative act; in her act, Parks’ communicated successfully her opposition to segregation. Brownlee counters such objections by acknowledging this ‘standing up for our beliefs’ in a way that encourages others to reflect on their views and the reasons they have for their views is also sufficient to meet this condition:37 It is still an attempt to start a reasoned dialogue about segregation even if the act in question does not involve a speech act.

An important caveat on the dialogue condition is that it is not a requirement to willing engage with everyone in reasoned deliberation on the merits. This goes some way to accommodating the cases of the vegetarian and the physician mentioned in the last section, who seemed to have good reasons not to engage in dialogue about their views. As it was in the case of the non-evasion condition, these exceptional cases are captured by holding that the condition is only supposed to hold ceteris paribus. Brownlee offers a substantial list of the cases where 'all is not equal'. It is claimed that the condition is sensitive to the burdens of vulnerability, disadvantage, unpopularity, relative power and the relative cost of communicating the conviction.38 If the agent is sufficiently vulnerable to serious legal or professional consequences, she is exempt from having to engage in reasoned dialogue. If it would make her sufficiently unpopular, we might suppose, dialogue is not required. If she has relatively little power or it is in some other way particularly costly, she need not engage in such dialogue. The extensive list of exceptions here might leave pause for thinking that dialogue cannot be a necessary condition of the communicative principle at all. In section seven I suggest that we do away with the ceteris paribus clauses, replacing them with the idea that communicativeness is a weighty rather than necessary condition of conscientiousness.

It is worth noting that the dialogue requirement can be normatively justified in terms of the three heads of conscientiousness in just the same way that the non-evasion requirement could be. Engaging in reasoned argument with others about the merits of our convictions can be

37 Brownlee (2012), p43.
38 Brownlee (2012), p44.
justified on the same Millian grounds that the non-evasion requirement can be. In fact, so too can the other two requirements of the communicative principle; consistency and universality. However, I will argue shortly that those two requirements are better captured as components of two other, separate, principles of conscience; the morality and identity principles.

4.6. The Communicative Principle and Personal Disobedience

In chapter three I began my evaluation of the argument that civil disobedience is a more conscientious act than personal disobedience. Now that we have a good understanding of the communicative principle and its requirements we are in a position to assess personal disobedience in terms of this principle. We are in a good position to ask: Is personal disobedience just as conscientious as civil disobedience? Recall the difference between the two key practices: the difference is not between practices that we (the ‘folk’) refer to with the terms ‘civil disobedience’ and ‘personal disobedience’. The folk don’t use the term ‘personal disobedience’ at all. Also, it may well be that there are a huge range of practices that we (the folk’) refer to as civil disobedience, but here we are only interested in a subset of those cases; namely, those cases where the agents’ acts are communicative. These are the paradigm cases of civil disobedience. Consequently, if we are to assess cases of personal disobedience we can’t just ask whether any old case of personal disobedience counts as conscientious. What we need to ask is whether paradigm cases of personal disobedience are conscientious. So what is a ‘paradigm case’ of personal disobedience?

In identifying a paradigm case Brownlee claims that her aim is “not to identify a model or ideal case of civil disobedience, but rather to identify a class of actions that are clear examples of civil disobedience.” It is true that the examples focused on by Brownlee do seem to be clear cases of civil disobedience, in so far as any individual reasonably competent in the use of the term ‘civil disobedience’ would apply it to the behaviour of Rosa Parks and Sir Thomas More. In Brownlee's view, while it may be true that some people refer to the

---

39 This is, I think, a key methodological mistake in Brownlee's analysis. She makes it clear that she is only interested in paradigm cases of civil disobedience, which are necessarily communicative. And then compares them with all cases of personal disobedience, both paradigm and non-paradigm. It is no surprise then that, in the end, she concludes that civil disobedience is conscientious and personal disobedience is not. This is true almost by definition.

40 Brownlee (2004), p338.
practice of radical protest as 'civil disobedience', it is not a clear enough case of that concept being applied to count as a paradigm case. But one problem with this account of paradigm cases is that it makes comparing the conscientiousness of civil disobedience and personal disobedience problematic. Given that the cases of civil disobedience that Brownlee is using for assessing conscientiousness are paradigm cases, they must be assessed against paradigm cases of personal disobedience. But since, no cases are commonly recognised as 'personal disobedience' (they are cases commonly referred to by the public as 'conscientious objection'), what can we take to be the paradigm cases of personal disobedience?

I will treat as paradigm cases two of the examples of personal disobedience referred to by Brownlee: The case of the abortion-averse clinician and the reluctant civil registrar. But we need to fill in some details in these cases. Given that there are no (or few) clear cases of a practice that is referred to as 'personal disobedience', we need some explanation of which cases of personal disobedience we will take to be paradigm. My suggestion is the following: What was telling about the paradigm cases of civil disobedience, what made them quite distinct from the cases of radical protest, was the motives of the civil disobedients. The behaviour of Sir Thomas More and Rosa Parks was, in both cases, morally praiseworthy. I suggest, therefore, that in assessing the accounts of personal disobedience, in filling in the details, we posit that the personal disobedients have \textit{morally reasonable} motivations. This does not mean that their motivations have to be morally praiseworthy. If we didn't make this assumption then we would simply be begging the question against the conscientiousness of personal disobedience. If in assessing civil disobedience, the morally problematic cases can be ignored and treated as 'radical protest', then so too can the morally problematic cases of personal disobedience.

Now that we have the comparison class clear, let's assess personal disobedience against each requirement of the communicative principle. In accordance with the first requirement, does the practice of personal disobedience exhibit consistency just as adequately as civil disobedience? Yes. There is no reason why a personal disobedient would suffer from practical irrationality any more than a civil disobedient. The civil and personal disobedient have some different beliefs and desires associated with their respective convictions. The civil disobedient believes that she ought to engage the public in dialogue about her moral views and desires that others share her views. The personal disobedient doesn't have that belief or
that desire. But there is no difference between the civil and personal disobedient in terms of whether or not the agents act in accordance with their beliefs and desires. The clinician who, in an act of personal disobedience, believes that she ought not to perform abortions, is just as likely to act on that belief as the civil disobedient pacifist who believes she should burn her draft card.

What about the universality condition? Is there any reason to think that civil disobedience involves convictions with universal scope and that personal disobedience does not? One might argue that the essentially communicative nature of civil disobedience, the fact that a civil disobedient is necessarily trying to engage others in dialogue about her convictions, means that these convictions must have universal normative scope in a way that the convictions of a personal disobedient need not. It would make no sense, we might suppose, for an agent to try and convince others of the truth of her conviction if it only had very limited normative scope. Just because I have a conviction that hokey pokey ice cream is the best it doesn't follow that this conviction has any normative force on others and therefore that I ought to persuade others to change their views.

I suggested earlier that, while it is more obvious, perhaps, that the convictions of the civil disobedient must have universal normative scope, there is no reason why the convictions of the personal disobedient cannot be the same. I mentioned earlier, the plausible thought that the universal normative scope of a conviction depends largely on whether that conviction is moral or not. I argue in the next section that we ought to separate the communicative principle into two: One principle that is properly communicative and one that is moral. One consequence of this is that a conviction can be communicative without being a moral conviction. But this is no reason to think that one sort of behaviour involves convictions that are necessarily more ‘moral’ or ‘universal’ than the other. But this is no point of difference between Rosa Parks and the abortion-averse clinician. It is likely that paradigm cases of both civil and personal disobedience will involve the expression of universal moral convictions.

What about the third, non-evasion, condition? Recall that by way of illustration I am considering the cases of the abortion-averse clinician and the reluctant civil registrar. Do both these agents honour their convictions? If the agents acted surreptitiously, if they were

---

41 Or, perhaps, the personal disobedient has a pro tanto belief or desire to that effect but it is defeated by other beliefs or desires that she has.
motivated by self-interest to hide their convictions, then their behaviour would not cohere with this condition. But the paradigm personal disobedient does not have such motivations. The paradigm personal disobedient only has motivations that are morally reasonable. If the abortion-averse clinician believes that it would be unprofessional to publicise her views, then her motivation is morally reasonable. If the reluctant registrar would lose her job by publicising her view, then even if she is motivated by self-interest. It would be a morally reasonable regard for her own self-interest. Do those cases of paradigm personal disobedience satisfy the non-evasion condition?

I argue that they do. Recall that I divided that clause into negative and positive elements. The negative element requires that the agent risk the negative consequences of acting on her conviction. The personal disobedients that I mention are risking some negative consequences of their actions. And while it is true that they are not acting in as risky a way as they would otherwise be doing, there are two ways in which this can be justified. In the case of the reluctant registrar, her motivation is a modest level of self-regard. Given that Brownlee suggests that overly burdensome consequences need not be endured, the ceteris paribus clause seems to apply to the conduct. Of course, some account needs to be given as to why the ceteris paribus clause will permit this sort of conduct and not others. I discuss this issue in the last and final section of this chapter. It is true that the actions of the personal disobedient will not be consistent with the positive element. The agent is not doing anything to try and get others to share her view. So, even more crucially than in the case of the negative element it will have to be the case that the actions of these personal disobedients fall within the ambit of the ceteris paribus clause, if they are to count as conscientious.

Finally, let's consider the fourth requirement. The dialogue condition, recall, requires that the agent engage with others in reasoned dialogue about the merits of her conviction. Is the paradigm personal disobedient acting in accordance with this condition? This condition is the most relevant one in distinguishing the civil from the personal disobedient. By definition, the personal disobedient is not attempting to engage others in dialogue about her convictions. Nevertheless, I argue, the personal disobedient (paradigmatically) does act in accordance with this condition. There are several different ways in which the actions of the paradigm

---

42 It is often argued that morality cannot require that the agent completely abandon her own projects (e.g. Williams (1973) or Wolf (1982)). Even if this is wrong, it something a morally reasonable agent could believe.
personal disobedient can be seen as consistent with that condition: In the first case, the
requirement of dialogue was not that the agent literally engage others in discussion about her
conviction. It was thought to be sufficient if the agent stood up for her beliefs in such a way
that it encourages others to reflect on the merits of the disobedient's view. The abortion-
averse clinician, even if she does not engage others in deliberation about her conviction, if it
is well-known that she refuses to be involved with abortion-provision, then her behaviour can
be said to satisfy the dialogue requirement. Of course, it can't be the intention of the personal
disobedient to be communicative, as if she had that intention she would be a civil disobedient
rather than a personal disobedient. But if her action has the effect of encouraging those who
approve of abortion to reflect on the merits of their views, then her action satisfies the
dialogue condition. The other ways in which the personal disobedient may satisfy the
dialogue condition involve invoking the ceteris paribus clause. Recall that the dialogue
condition requires that “we appreciate the reasons for communicating and ceteris paribus
intend to do so, unless the reasons to communicate are outweighed by undue costs of
communicating”. As we saw, those costs included the “burdens of vulnerability, disad
vantage, unpopularity, relative power, and relative cost of communication”. What we
need to ask is whether the paradigm personal disobedient, the disobedient who has morally
reasonable motivations, is going to be covered by that clause. Any one of those listed costs
could apply in the case of the paradigm disobedient.

It is worth discussing at this point, where we are up to in this analysis of Brownlee on
conscientiousness over the last two chapters. The purpose of these two chapters is to
introduce a contemporary account of conscience which constitutes essential background for
my own analysis of conscience. I have articulated Brownlee’s communicative principle and
defended it in terms of three normative ‘heads’ of conscientiousness. I suggested that, contra
Brownlee, one species of putatively conscientious action, personal disobedience, is just as
conscientious as civil disobedience. However, both my defence of the four requirements of
the communicative principle and my argument that personal disobedience is just as
conscientious as civil disobedience, depend on the legitimacy of ceteris paribus clauses. This
is the subject of the next section.


4.7. *Ceteris Paribus* Clauses and the Principles of Conscience

The consistency, non-evasion and dialogue requirements all contained *ceteris paribus* clauses. We might even wonder whether the missing requirement, universality, should also contain a *ceteris paribus* clause. At the start of chapter three I mentioned that it is actually quite common to deny the universal normative scope of conscientious convictions. For, example, on Arendt’s analysis of conscience following one’s conscience is a matter of the individual working out if she could live with herself if she were to engage in some act. It does not matter whether any other person would be able to live with themselves. Given the possibility of such a case we might want to accommodate them by holding that the universality requirement itself has a *ceteris paribus* clause.

On the face of it, there is something paradoxical about these clauses. By adding such a clause to a putative necessary condition it is as if one is saying ‘Necessarily, x is y, except sometimes when x is not y’. One way of explaining such cases is to suggest that they are not quite as they appear. They don’t purport to give genuine exceptions to the putative necessary condition at all. Rather, the clause is a short-hand for some finite list of exceptions that could, (at least potentially) be explicitly spelled out. Furthermore, a substantive argument can be given as to why those exceptions are legitimate rather than merely stipulating that they are. For example, we might interpret the *ceteris paribus* clause in the consistency condition as referring to all instances of practical failure and then give a precise definition of practical failure.\(^{45}\) But the other *ceteris paribus* clauses do not look as though they can be as easily analysed in this way. Recall that in the case of the non-evasion requirement, Brownlee claimed that the agent must risk the potential negative consequences of her actions, except when those negative consequences will be particularly burdensome. No further explanation was given as to when the circumstances count as being particularly burdensome. I am

---

\(^{45}\) Michael Smith offers such a diagnosis in his defence of moral motivation internalism. His internalism thesis is the view that:

> It is conceptually necessary that if an agent judges that she morally ought to $\diamond$ in circumstances $C$, then either she is motivated to $\diamond$ in $\diamond C$ or she is practically irrational ((2007), p 211).

Smith argues that this practical irrationality clause which some (such as Jeanette Kennett & Cordelia Fine (2007)) refer to as a '*ceteris paribus*' clause can be given a substantive explanation as a failure of the individual to be sensitive to her reasons for wanting (2007), p215).
sceptical that any such explanation could be given. What about the list of exceptions built into the dialogue condition (vulnerability, disadvantage, unpopularity, relative power and the relative cost of communicating the conviction)? Again, it is hard to imagine what principled basis there could be to unite this seemingly disparate list. Is it possible to propose such a requirement as a necessary condition when there is some undetermined range of situations when it doesn’t apply?

Perhaps there is another way of analysing such clauses. In “From particularism to defeasibility in ethics” Mark Lance & Margaret Little offer a pragmatic analysis of *ceteris paribus* clauses where they argue that necessary conditions can be ‘porous’. That is, they contain exceptions that are, in some sense, ‘indeterminate’ and cannot be eliminated. We employ this type of condition when we wish to capture the notion not that some circumstances always hold, or even that they usually hold, but rather that there is a certain set of circumstances that are particularly telling of something’s nature. Lance & Little illustrate this point with the example of a scientific law pertaining to matches. How does one explain the fact that a match lights when struck? It can’t be explained simply by invoking a law ‘matches always light when struck’. Matches don’t always light when struck. Nor can it be explained by a law which holds ‘in x proportion of cases (e.g. 90% of the time) matches light when struck’ (a statistical generalisation). Imagine in the future that all humans inhabit colonies on a Martian moon and that due to a lack of oxygen, it is almost never the case that a match lights when struck. Even in that future, an agent will still, in some sense, be justified in claiming that matches light when struck. We make a claim about the nature of matches whenever we make the claim ‘matches light, when struck’. Another way of putting it is that we are taking as privileged in one way or another those cases where some feature is present as opposed to those cases where it is not. To grasp the concept <match> one needs to know the privileged conditions; those conditions under which the match lights, and some of the situations where the match does not light. It is to know for example, that there must be sufficient oxygen for them to light, and that the match must not be wet and that the temperature cannot be absolute zero.

46 (2008), p54.
47 Sometimes it seems as though the language that Lance & Little use conflates two separate questions: (a) what is it that makes the claim ‘matches light when struck’ true? And (b), under what circumstances can we justifiably assert ‘matches light when struck’? The first is a metaphysical question while the second is an epistemological one and the answer to each question could potentially be different. I assume that Lance & Little are answering
I don’t think that this sort of analysis will be useful in the case of the communicative principle. This pragmatic analysis of a *ceteris paribus* clause only works in cases like the matches case because we have independent information about which conditions are the privileged conditions; we have information about which natural properties of the matches mean that they light when struck. When dealing with a sufficiently vague moral concept like communicativeness we do not have this sort of independent information. How can I know that it is ‘engaging in reasoned dialogue’ which is the privileged condition of the dialogue condition without begging the question about what is constitutive of conscientiousness? One cannot simply appeal to one’s intuitions that it is the dialogue that is really important because this is (perhaps) exactly the intuition that defenders of personal disobedience do not have.

The status of personal disobedience as conscientious depended, I claimed, on the legitimacy of the *ceteris paribus* clauses contained in the requirements of conscientiousness. Given that I have suggested these clauses (in the case of the communicative principle) are illegitimate, where do we go from here? I suggest that we drop the *ceteris paribus* clauses and, in fact, drop the idea that we are dealing with necessary conditions at all in our analysis of conscientiousness. Instead, I suggest, we pay close attention to the fact that we are dealing with a principle of conscience, not a ‘condition’ *per se*. According to Ronald Dworkin’s influential analysis of principles, a principle can be distinguished from a rule in the way that it does not apply in an all-or-nothing fashion, but instead has a dimension of *weight*. If we interpret the communicative principle as a principle rather than a rule/necessary condition we can say that it *counts in favour* of some act being conscientious if it is non-evasive and dialogue-enhancing.

On my proposal, we can treat the non-evasion and dialogue requirements as necessary conditions of the communicative principle, but deny that they are necessary conditions of conscience. This allows us to drop the *ceteris paribus* clauses while (a) recognising what is special about the communicative principle and (b) preserving the conscientiousness of personal disobedience.

---

48 Dworkin, (1967).

49 I will argue shortly that the other two putative requirements of Brownlee’s communicative principle (consistency and universality) pertain to other principles of conscience altogether (i.e. the morality and identity principles).
In the first three sections of this chapter I argued, following Brownlee, that the ultimate justification of the communicative principle is not our descriptive intuitions about what conscientiousness consists in. Ultimately, any proposed condition of conscience/conscientiousness must be grounded in the three normative heads of conscience. I argued that the positive element of the non-evasion condition could be justified in this way. But, as I mentioned earlier, both of the core elements of the communicative principle can be justified in this very way. The non-evasion condition can be justified in terms of those three heads because it is individuals in honouring their convictions that bring those convictions into the public sphere and therefore allow for the possibility of those convictions influencing broader society. A similar story can be told in the case of the dialogue requirement. It is by entering into reasoned dialogue with others about the merits of one’s view, that one most effectively communicates to broader society what one’s views are. And the presence of these views in broader society has all the advantages referred to in the Millian arguments.

Framed in this way, personal disobedience is not consistent with the communicative principle. But personal disobedience can still count as conscientious. It is conscientious because while it counts in favour of some conviction or action being conscientious that that action is consistent with the communicative principle, it is not a necessary for it to do so.

4.9. The Cluster Concept of Conscience

My suggestion in the last section, that we interpret the communicative as a true principle rather than a necessary condition of conscience, is suggestive of a different way of thinking about conscience. In the introduction I claimed that conscience was a cluster concept. What do I mean by this? Philosophical treatment of cluster concepts is sometimes thought to originate in John Searle’s account of proper names. Searle defended a cluster theory of proper names. This theory held that a proper name refers to whatever it is that satisfies enough of the descriptions associated with it (it need not be all of those descriptions).

---

50 See Glenney (2013).
It is a necessary fact that Aristotle has the logical sum, inclusive disjunction, of properties commonly attributed to him: any individual not having at least some of these properties could not be Aristotle.\textsuperscript{51}

Hilary Putnam, applying such an approach in the case of concepts generally said:

Suppose one makes a list of the attributes $P_1$, $P_2$ ... that go to make up a normal man. One can raise successively the question 'Could there be a man without $P_1$?' 'Could there be a man without $P_2$?' and so on. The answer in each case might be 'Yes' and yet it seems absurd that the word 'man' has no meaning at all ...the meaning in such a case is given by a cluster of properties. To abandon a large number of these properties, or what is tantamount to the same thing, to radically change the extension of the term 'man', would be felt as an arbitrary change in its meaning. On the other hand, if most of the properties in the cluster are present in any single case, then under suitable circumstances we should be inclined to say that what we had to deal with was a man.\textsuperscript{52}

I think that conscience is a similar sort of concept to Putnam’s concept of <man>. A cluster concept is constituted by a series of considerations that count for some instance being an application of the concept in question. But while it may be true that a concept must satisfy any one of those considerations, it does not need to satisfy all of them.

In the case of conscience I suggest that there are three considerations that count for an action or conviction counting as conscientious. These are the principles of identity, morality and communicativeness. Since conscientiousness comes in degrees we can hold that the most conscientious of actions will implicate all three of the principles. The most conscientious of actions are those that implicate the individual’s identity, which the individual seeks to communicate with the broader public, and which contain moral content. Given that I have only discussed in detail the arguments for thinking that a given conviction or action must be communicative in order to count as conscientious, why are the other principles necessary?

First, let’s consider why the communicative principle ought to be supplemented with a principle of morality. I mentioned in the previous chapter that while the consistency and universality requirements did seem to have some conceptual connection to conscientiousness, there didn’t seem to be any conceptual connection between those conditions and the

\textsuperscript{51} Searle (1958), p172.

\textsuperscript{52} Putnam (1962).
communicative principle *per se*. In the case of the universality requirement, I suggested that the universality of those convictions was simply a consequence of them being *moral* convictions. Following Kant, the universal normative scope of the moral is often thought to be *the* distinctive feature of morality. Therefore, I suggest holding that morality is a separate principle of conscientiousness. In many cases an individual’s communicative convictions will be the same as her moral convictions so the two do not come apart. But it is possible for an individual to have a communicative conviction that is not moral. For example, perhaps I have a conviction that camouflage trousers look ghastly. And, in light of this I refuse to wear the trousers that I am required to wear as a soldier. Not only do I stand up for my belief in front of others, I try to *convince* them of the merits of my view. I try to convince them that they would look better in a plain khaki.53 This conviction and action would be communicative but not moral. Now, this would certainly be a relatively unconscientious action, but that is because it doesn’t satisfy the normative heads of conscientiousness in the same way that a communicative conviction that was also moral in content, would.

As I mentioned earlier, in this thesis I assume that the core requirement of the morality principle is the requirement that moral convictions are *categorical* convictions: Convictions that the individual believes to have normative force irrespective of her own desires. The fact that a categorical conviction does not depend on the desires or interests of the individual to have normative force means that it is *universal* in normative scope. We can see once we distinguish the communicative and moral principles that many of the arguments grounded in the three normative heads were primarily arguments in favour of a morality principle: It is not simply communicative convictions but *moral* convictions that, for example, we would agree to protect in the original position.

I mentioned a third principle of conscientiousness, an identity principle. Why the need for an extra principle? In chapter two I discussed the accounts of *conscience as identity* and suggested there that there is a close connection between conscience and identity. Using Arendt’s locution, claims about conscience are (often) claims about whether we could *live with ourselves* if we were to perform some act. This accounts for the tension I mentioned in chapter three between the universality requirement and the supposed first-personal nature of conscience. The reason that some conscientious actions are first-personal is that they make a

53 And in making this claim I am not making any claim that is necessarily categorical (and therefore, on my analysis, moral). My claim might depend entirely on a *hypothetical* notion of normativity.
reference to the individual’s *identity* whereas other conscientious convictions and actions are universal in scope because they implicate the morality principle, rather than identity. It is possible for a conviction to be both moral and identity-implicating; such a conviction or action would simply be *more* conscientious than the conviction which was only moral or only identity-implicating. Like the communicative and moral principles, the identity principle too can be justified in terms of the three normative heads of conscientiousness. But in order to give such a justification we need to be much clearer about what the identity principle involves. This is the subject of Part III of the thesis.

To summarise: The concept of conscience cannot simply be analysed in terms of the communicative principle. Nor can it simply be analysed in terms of the morality principle (as, say, Bishop Butler did). Rather, we ought to think of the concept of conscience as a cluster concept constituted by three core principles. While each principle ought to have some level of descriptive intuitiveness (be consistent with what we, the folk, would be inclined to name a ‘conscientious’ act or conviction), ultimately it must be justified in terms of the three normative heads of conscientiousness. Each principle has a certain weight and counts *for* some conviction or action being conscientious. In my view the identity and morality principles are the weightiest, while the communicative principle is relatively less weighty. The reason for this is that a right of conscience is primarily intended to protect the interests of the individual, rather than the broader interests of society; the identity and morality principles ally more closely with the individual’s own interests, whereas the communicative principle aligns more with the interests of society. The *more* conscientious of convictions and acts, the ones *more* worthy of protection in a right of conscience, are the ones that involve the agents’ moral and identity-implicating convictions. Given that conscientiousness comes on a sliding scale we can still say that the *most* conscientious act is the one that implicates all three of the principles.
Part III

Conscientious Identity
Chapter Five: *Les Misérables* and Conscientious Identity

I mentioned in the introduction that this thesis can be divided into three broad parts. In Part One I looked at historical accounts of conscience. In chapter one I considered various accounts of *conscience as moral agency*. This is the view that the individual’s conscience consistent in her capacity of moral agency (whether this is a capacity of moral judgement, moral feeling, moral knowledge, or some other capacity of moral agency). Then in chapter two I looked at what are perhaps the two most prominent philosophical treatments of conscience in the twentieth century. I described the accounts of Heidegger and Arendt on conscience and claimed that these were accounts of a quite different approach, the *conscience as identity* approach.

In Part B I moved on to considering in depth a contemporary account of conscientiousness: Brownlee’s communicative theory. Over chapters three and four I considered Brownlee’s argument that civil disobedience is a more conscientious act than ‘personal disobedience’ (what we might more commonly refer to as ‘conscientious objection’). In chapter three I introduced Brownlee’s communicative principle of conscientiousness and the four requirements that constitute that principle. I argued that the legitimacy of the communicative principle depends on what the normative grounds of conscientiousness are. In the fourth chapter I argue that there are three normative heads of conscientiousness: the humanistic principle, the psychological benefit principle and the societal value principle. Together, these ‘heads’ ground a moral right of conscience (or conscientious action). I concluded that the communicative principle could be justified in terms of these heads but that, on that basis, personal disobedience was just as conscientious as civil disobedience. However, personal disobedience only counted as conscientious, in my view, because of the presence of *ceteris paribus* clauses attached to the requirements of the communicative principle. I argued that these *ceteris paribus* clauses ought to be done away with. Their function is better fulfilled by treating the communicative principle itself as a *principle* of conscience, properly construed. I suggested that we treat the communicative principle along with the morality and identity principles as together constitutive of the cluster concept of conscience.

In this part, Part III, I focus just on the identity principle, or as I call it, ‘conscientious identity’. It is this principle which is absolutely crucial to a comprehensive account of conscience, but has been relatively neglected in recent years. I aim to show how recent
discussions of identity in moral philosophy can be used to shed light on the nature of such a principle. I begin in this chapter with what I take to be a paradigm case of conscientious identity, the literary case of Jean Valjean in Victor Hugo’s *Les Misérables*.

This chapter shall proceed in the following way: In section one I look more closely at the potential source materials for extracting the principles of conscience. Commonly in philosophy, when trying to work out the necessary and sufficient conditions of some notion (in this case a principle) we appeal to fictional cases constructed by the theorist herself. In this chapter, instead of using such examples, I appeal to two literary cases from Victor Hugo’s *Les Misérables*. In Section two I describe the first literary case that I wish to use from *Les Misérables*. This case involves an event of personal transformation or ‘transfiguration’ for the protagonist Valjean (henceforth the ‘transfiguration event’). In section three, I describe the second case; this case involves Valjean taking the identity that is first formed in the transfiguration case and employing it in his deliberation about what he ought to do.

In section four, I attempt to distil from those two cases the descriptive requirements of conscientious identity. I posit four requirements: First, I claim that there is a distinctness requirement. Not all genuinely normative practical considerations are conscientious. Conscientious considerations are to be considered alongside other practical considerations and it is not a given that conscientious considerations will always defeat other practical considerations. In order to illustrate the distinctness requirement I consider two prominent accounts of practical identity, those of Christine Korsgaard and Charles Taylor. I argue that neither account is capable of satisfying the distinctness requirement: Both accounts mandate that all normative considerations implicate identity. The second requirement of conscientious identity is the first-personal requirement. It is a peculiar aspect of conscientious thought, I suggest, and that it involves the agent’s own views about her identity. It implicates the way in which an agent sees herself. Third, I posit a practicality requirement. An agent’s conscientious identity is something that is implicated in her thought and actions when she is

1 Recall that while the three principles of conscience are not necessary conditions of conscience, we can legitimately talk of the necessary conditions that constitute each individual principle.

2 Or, at least, not all practical considerations are conscientious as a matter of conceptual necessity. It may well turn out that, as an empirical matter, all practical considerations turn out to be conscientious.
in the appropriate factual circumstances. Finally, I posit an *endurance* requirement. An agent’s conscientious identity is not something she possesses for a mere moment but is something that *endures* over time.

5.1. Paradigm Cases of Conscience

In chapter four, following Brownlee, I claimed that in order to work out the central conditions of conscience we ought to begin with *paradigm cases* of that concept being applied. In that chapter, in order to draw out the communicative condition, the focus was on paradigm cases of conscientious *action*; cases where we would judge that an agent's action was motivated by conscience. In this chapter, in order to draw out the identity condition, I instead focus on paradigm cases of conscientious *thought*; cases where we would judge that an agent’s thought processes implicate conscience.

Why this focus on paradigm cases? In the last chapter, I discussed the idea that a paradigm case is a *clear* case of the application of some concept. Some interpretation was required in order to come up with a ‘clear’ case of personal disobedience since that is not a term that people commonly use. What about in looking directly for paradigm cases of *conscientiousness*? Given that this is a term that is in the common parlance it seems as if the usual approach can be applied. By way of illustration, consider the concept of *<wrongness>*. Imagine that someone utters the following sentence: ‘You can't pair tassel loafers with a business suit! It's just wrong!’ Putting to one side consideration of whether the second sentence is true, is this a paradigm case of the application of the concept *<wrongness>*? I don’t think it is. We can't say that all those competent in the use of the concept *<wrongness>* would describe it in that way. According to many people, saying that an outfit is 'wrong' would just be a hyperbolic way of saying that an outfit is aesthetically displeasing or ugly. Having said that, perhaps there are some who would say this *is* a case of the concept of *<wrongness>* being applied. Whichever one's view, this is not a paradigm case of the concept

---

3 This last caveat is important as it is a contingent matter whether the agent is ever actually called to employ her conscientious identity in her thought or action.

4 In Brett Easton Ellis’s *American Psycho* (1991), Patrick Bateman claims that it is not. Black or cordovan tassel loafers can be properly matched with a business suit.
of wrongness. It is an example of the concept being applied in a way that many find objectionable.

So where do we go for a paradigm case? There are several different sources that we could use. We, the theorists, might construct our own fictional case of the concept being applied. But this is a problematic way of identifying a paradigm case. One problem with this approach is that, usually, such a case’s status as a paradigm case relies entirely on the theorist’s own judgement or intuitions about whether this is a case that a master of the concept in question would judge in that way. There is always the risk that her theoretical intuitions are corrupted. That in coming to that judgement the theorist is unduly influenced by her own theoretical commitments into thinking that the case before them would be accepted as such. Perhaps the widely held belief of theorists that Gettier cases do not count as knowledge is an instance of this. Let’s assume for the moment that we do not want to engage in time consuming empirical testing in order to work out whether that intuition is widely shared for every thought experiment we wish to appeal to.

Another possible source of paradigm cases are reports of actual historical cases. In chapter four we saw that Brownlee appealed to the actions of Rosa Parks sparking the Montgomery bus boycott as a paradigm case of the civil disobedience. Brownlee also appealed to the case of Sir Thomas More and his trial for refusing to act contrary to the dictates of his conscience. The advantage of using such a source as opposed to the theorist’s own imagination is that there is a better independent check on whether the case is a genuine paradigm case. It is not simply a matter of the individual theorist thinking that some situation is a good paradigm case. There is a long line of people treating it in that way.

Another possible source of paradigm cases, and the source I appeal to in this chapter, is literature. Literary cases, in my view, just like historical cases, involve an independent check as to their paradigm status. I suggest two (connected) ways in which a literary case can constitute an independent check. First, in order for a reader to understand a character in literature, the character needs to make sense. And a character does not make sense if that character appears to be deploying incoherent concepts. The second way in which literature can be revealing is that a specific understanding of a character, as employing certain

5 It has been argued that experimental data does not support the theorists’ view (e.g. Starmans and Friedman (2012), Weinberg, Nichols and Stich (2001)).
concepts, might be necessary in order to understand that character’s relation to the broader meaning of the literature in question. A particular understanding of the concept might be part of the best interpretation of what is going on in that literary text.

Starting with the first possibility, Marya Schechtman makes the observation that a ‘well-drawn’ fictional character requires that the various actions and thoughts of that character make sense.\(^\text{6}\) If the thoughts and actions of that character do not make sense then we would not consider that individual be a well-drawn character, there would be something ‘off’ about him or her. And it is the well-drawn characters that tend to resonate with an audience and, as a consequence, well-regarded literature tends to have well-drawn characters. This means that if some character is widely regarded as being well-drawn, and widely accepted by readers as being so, then that character pursues various thoughts and actions that make sense. Well-drawn characters possess coherent concepts.\(^\text{7}\) I think that the fact that Valjean comes across as a well-drawn character (and particularly so in the two incidents that I refer to) is because in those cases he is applying a coherent notion of conscience in his thought.

How is this suggestion, that we can determine the descriptive contours of some concept by reference to our intuitions about whether a character is well-drawn, any more reliable than the ordinary theoretical intuitions that I referred to earlier and suggested were dubious? The difference is that there is a modest independent check on that intuition here. In most cases, a novel will not be well-regarded unless it contains well-drawn characters who are, by definition, fictional individuals applying coherent concepts.

But it must be noted that the only conclusion that we can make from a character being well-drawn is a modest one. We can only conclude that some description of a concept is coherent. But when looking for a paradigm case we want something more than a merely coherent articulation of a concept. I am arguing in this thesis that conscience ought to be analysed in terms of three principles, communicativeness, morality and identity. But I don't think other views (such as Brownlee's) are incoherent. As we shall see in Hugo’s descriptions of conscience that follow, conscience is at times described as the internal word of God or the

\(^{\text{6}}\) (1996), p97.

\(^{\text{7}}\) Though this might not be true in every case. Perhaps in the world of George Orwell’s 1984, a character that engages in doublethink – that believes two contradictory statements at the same time – counts as well-drawn. Because part of what it is to be an ideal citizen in that society, a well-drawn fictional character in that society, is to be psychologically incoherent.
capacity to practically reason as well as the identity-implicating conception that I emphasise. So why do I pick out conscientious identity as the notion of conscience implicated in Valjean's thought?

I suggest that we can focus on the notion of conscientious identity implicated in his account because of the second way in which literary cases can constitute paradigm cases: Some articulation of a concept can contribute significantly to the meaning of the narrative in a novel, whereas other articulations of that concept might make no such contribution. Let’s consider this suggestion carefully: What precisely is the ‘narrative’ of a novel? It is a form of discourse which has as its principal property, *sequentiality*: a unique sequence of events involving humans as characters. In *Les Misérables*, these events, the elements of the narrative, include the goings-on in the lives of the principal characters (such as Valjean), but also extends to matters which do not bear directly on the lives of those characters (such as the extended discussion of the street urchins of Paris that opens volume III). In *Acts of Meaning*, Bruner points out that the meaning of a narrative is not given simply by looking at those individual elements alone but by looking at how those individual elements are related in terms of the plot or *fabula* of the narrative. And the reverse is also true: the reader can only understand the overall meaning of the narrative by understanding the individual components.

At the end of this chapter I argue that conscience understood in terms of the four requirements I list, plays a crucial role in understanding the overall meaning of the narrative of *Les Misérables*. A particular way of articulating conscience, the way that I describe, is essential to understanding that broader meaning.

How exactly does the relation of some constitutive element in a narrative to the broader meaning of an item of literature tell us *anything* about philosophical intuitions? The fact that a certain conception of conscience plays a crucial role in elucidating the meaning of a literary work is one way of showing a clear-cut instance of a concept being applied. Just like in historical cases, there is an independent check on whether that interpretation is a paradigm case. If some articulation of a concept is essential to an understanding of the overarching narrative of the novel and that novel is widely understood, then, I suggest, we have a paradigm case.

---

8 Bruner (1990), p43.

9 Ibid.
There is philosophical precedent for drawing on literary cases in order to explicate the conscience. In Jonathon Bennett's "The Conscience of Huckleberry Finn", Bennett argues that a description of conscience given in Mark Twain's *Huckleberry Finn* draws out some of the conceptual constraints on conscience.\(^{10}\) In particular, Bennett claims, the deliberation of Finn reveals the way in which the judgements of conscience come apart from the feeling of sympathy. In the case Bennett describes Finn is deciding whether or not he should continue to help an escaped slave (Jim) or whether to turn him in to the authorities. In Finn's own view, as described in the novel, it is his conscience that dictates that he turn in Jim. And because his conscience tells him to do this, he feels guilty for the help he has already given Jim. Nevertheless after Jim expresses his gratitude to Finn, Finn finds that he is unable to hand Jim over to the authorities.

According to Bennett's interpretation of Finn’s situation, what is revealed is the way in which the conscience can come apart from sympathy. In Bennett's view the conscience concerns itself with "principles, arguments, considerations, ways of looking at things".\(^{11}\) Because of this, conscience provides the agent with reasons for action. The capacity for sympathy, according to Bennett, is unlike conscience because even though it can move an agent to act one way rather than another, it is not reason-giving and often comes into conflict with the dictates of conscience.

Bennett claims that the Finn case "brilliantly illustrates how fiction can be instructive about real life".\(^{12}\) I have suggested thus far that fiction can be conceptually instructive– it can provide evidence as to how a particular folk concept is constructed. What is less obvious is that literature provides evidence of how our various psychological processes actually operate. For example, it seems as though Thomas Harris’s Hannibal Lecter captures a folk understanding of what a psychopath is, but it is doubtful that it reflects the actual psychiatric diagnosis.\(^{13}\) Twain might have accurately captured what the folk concept of conscience is,

\(^{10}\) Bennett (1974).


\(^{13}\) For example, see De Lisi et al. “The Hannibal Lecter Myth: Psychopathy and Verbal Intelligence in the MacArthur Violence Risk Assessment Study” (2009).
while it still being true that no such concept is ever instantiated in people’s minds in the real world.

Putting that matter to one side, does the Finn case accurately describe our concept of conscience, as Bennett suggests? There are a few ways in which we could push back against Bennett’s analysis. One could argue that, in Twain’s description, Finn is wrong about his own mental states. Finn may well refer to the principles that favour turning in Jim, using the term ‘conscience’, but be wrong to do so. It might be that the deeper sympathy that leads him not to hand in Jim that represents his bona fide conscience.\(^\text{14}\) Given that references to a concept in literature can be misguided or wrong, should we then doubt that these fictional literary descriptions are a good source of paradigm cases? I suggested two (connected) ways in which we can find literary descriptions helpful. In the first place, well-drawn characters must be described using coherent descriptions of concepts. Does the description of Finn, that his conscience is concerned with rational matters only and is not what in the end motivates his actions, contribute to our understanding of Finn as well-drawn? It doesn’t seem to. In fact, I think that the opposition between conscience and sympathy required by Bennett's analysis is contrary to some common intuitions about conscience. I think we often describe people as motivated by conscience when it is sympathy that moves them. It would be perfectly sensible, in terms of the novel, to say that Finn’s decision not to turn in Jim was in fact motivated by conscience. It could be argued that, in terms of this incident, the character is not particularly well-drawn.

The other way in which I suggested that one particular articulation of a concept, one formulation, might be supported by our analysis of conscience is that that articulation might be essential in contributing the meaning of the narrative of the novel. In the case of Finn, I don’t think that this peculiar articulation of conscience is essential to understanding the narrative of the novel. A concept of conscience that was emotionally grounded (where Finn’s

\(^{14}\) One might wonder whether the account of conscientious identity that I go on to extract from Les Misérables is consistent with the concept as articulated in Huckleberry Finn. I would suggest that it is. The account of conscientious identity that I articulate would allow for the judgements of conscience and feelings of sympathy to come apart in the way that they do in the Finn case. Furthermore, it might even be argued that the way in which Finn’s conscience is described invokes some notion of identity and is therefore not only consistent with but supportive of my own account of conscientious identity. My only claim in using cases from Les Misérables is that they bring out the identity condition more clearly than in the Finn case.
inclination to sympathy counted as an expression of conscience) would just as well serve to explain the meaning of the narrative in that novel.

Of course, as I continually emphasised, whether or not a putative analysis of conscience (or some other concept) is the correct one, in the end, depends on whether that articulation satisfies the normative constraints on the concept. In the case of conscience, the correct analysis of conscientious identity must capture the sense in which rights of conscientious action are valuable; the way in which rights of conscientious action capture the value of an agent’s autonomy, her psychological integrity and the social value of conscience.

As I mentioned earlier, I study two cases from *Les Misérables* that describe Valjean's practical thought. As in Bennett’s appeal to *Huckleberry Finn*, I suggest that the descriptions of conscience in *Les Misérables* are paradigm cases of the concept of conscience being applied. It is only by reading the novel in this way that we arrive at a good understanding of the character of Valjean and the role that character plays in the broader narrative of the novel. The first case, the transfiguration case, involves Valjean beginning to question what sort of person he is, and, perhaps, what sort of person he ought to be. The second case involves Valjean engaging this previous undertaking in practical deliberation. Valjean weighs his previous undertaking with other considerations before resolving to act on his undertaking from the first incident. On my interpretation, something distinctive about the exercise of conscience in both those events is an appeal to *identity*.

Before turning to consider the two Valjean cases in depth it is worth mentioning again that in this thesis I seek to draw out three principles of conscience, conscientious identity, being only one such of those principles. As well as the identity principle there is also communicativeness and morality. The *Les Misérables* cases implicate those other two principles as well. Valjean agonises over and ends up adopting an identity based on the moral features of that proposed identity. In the deliberation case, when Valjean determines that he will stop hiding and reveal himself to the court at Arras he confirms that his conviction is communicative: he intends to communicate (and in fact does so) to the wider world.


16 Valjean himself recognises these two cases as the two crises of conscience in the novel; "He felt he was on the second decisive crisis of his conscience and of his destiny; that the Bishop had marked the first phase of his new life, and that Champmathieu marked the second" ((1867), p803).
Nevertheless, I think that it is the identity principle that is the most prominent principle in both Valjean cases.

In this section I have suggested that in order to understand a concept, we need to look to paradigm cases of that concept being applied. And while we, the theorists can construct our cases, ‘thought experiments’, to draw out the central features of a concept, we can also use relevant actual cases or fictional literary cases. The advantage of those sorts of paradigm case is that there is a modest ‘check’ on whether or not some case is a paradigm case in that blatantly incorrect articulations of some concept would not resonate with readers. In the case of the concept of conscience, such a task has been carried out by Jonathon Bennett. Bennett uses a case of the deliberation of Finn as described in *The Adventures of Huckleberry Finn* in order to draw out the sense in which, supposedly, conscience involves an appeal to reason, and the capacity for sympathy does not. I suggest that we appeal to fictional cases in Victor Hugo’s *Les Misérables* in order to draw out the way in which conscience, unlike some other parts of our practical psychology, implicates an agent’s identity. For the remainder of the chapter I describe these two cases in detail in order to draw out the four descriptive requirements of conscientious identity.

5.2. The Transfiguration Case

Before proceeding to the first Valjean case it is useful to describe some of the core background of Valjean prior to the point in the novel where the transfiguration event occurs. This is useful because (a) it helps us understand what is actually going on in each of the cases (b) it helps us to see how both events fit into the broader narrative of the novel.17

At the point of the transfiguration incident, Valjean's views about the sort of person he is and how he ought to behave have been heavily influenced by his period of imprisonment in the Bagne of Toulon. Valjean was imprisoned at a young age for stealing a loaf of bread in order to feed his starving family. The five year sentence for that offence was eventually extended to 19 years in light of his repeated escape attempts. By the time he is finally released on parole, he has become embittered and misanthropic. Furthermore, a point that becomes crucial in my

---

17 I mentioned in the last section that this is particularly important as our judgement as to whether some case is a paradigm case depends to some extent on whether that description is crucial in explicating the meaning of the narrative in the novel.
interpretation, Valjean does not just act in a vicious way, he has also *internalised* that behaviour and 'identifies' with it. I devote part of this chapter as well as the next two chapters in attempting to explain what exactly it means for an agent to 'identify' with some view, but for now we can gloss the notion of identifying as 'the way Valjean sees himself'; Valjean sees himself as a brute.

It is the transfiguration event that leads Valjean to deeply question his previous behaviour and ‘identity’. Once released from prison, Valjean finds that his parolee status means he is turned away from all regular accommodation. His luck changes, though, when he is offered a bed and food by the charitable Bishop of Digne, Monseigneur Bienvenu (‘The Bishop’). As a result of this act of kindness, Valjean feels the first ‘pricking’ of conscience; he begins to ponder as to whether he should start again, to give up wrongdoing and to rescind his prior brutish identity. As it happens, Valjean is not yet moved to change and instead absconds in the night with the Bishop's silverware.

The next morning, law enforcement catch up with Valjean and present him to the Bishop, presuming Valjean to have stolen the silverware. The Bishop denies that the silverware was stolen, and insists instead that it was intended as a parting gift for Valjean. When alone with Valjean, the Bishop asks him to promise that he will use the money from the sale of those goods to start his life anew as a virtuous man. While Valjean is, again, perturbed by this act of the Bishop, he is still not ready to renounce his past. Soon after his encounter with the Bishop, Valjean steals a coin from a passing urchin, Little Gervais. It is only after this event, while reflecting on his ability to commit such a vile act, that Valjean experiences the identity crisis that culminates in his 'transfiguration'.

Valjean finds himself in a state of conflict between the way in which he has been acting up until that point and a new path for action that has been laid bare by the actions of the Bishop. Valjean comments:

> that which was certain, that which he did not doubt, was that he was no longer the same man, that everything was changed, that it was no longer in his power to make it as though the Bishop had not spoken to him and had not touched him.\(^\text{18}\)

---

\(^{18}\) (1867), p416.
This locution - 'that he was no longer the same man' - is an important one. It emphasises the sense in which the transfiguration case involves an identity crisis; it is not just a matter of Valjean wondering to himself what he should now do, but is a matter of Valjean choosing which sort of person he is and/or ought to become. In the remainder of this chapter and in the chapter that follows I will suggest that this talk of becoming another person is not merely a metaphorical way of Valjean talking about exchanging his old moral principles for new ones. According to my interpretation, there is a real sense in which Valjean undergoes a change in identity.

Valjean goes on to reason to himself as to why he might have stolen the coin from Little Gervais:

Let us say it simply, it was not he who stole it, it was not the man; it was the beast, who by habit and instinct, had simply placed his foot upon that money, while the intelligence was struggling amid so many novel and hitherto unheard of thoughts besetting it.¹⁹

Valjean here lays the blame for the stealing of the coin with his 'previous self'; his action was a result of deeply ingrained brutish inclinations rather than any deep reflection as to the sort of person he ought to be. In fact, by Valjean's lights, it is because of his internal struggle that he acted purely out of impulse.

Valjean then goes on to describe a peculiar feature of this contemplation. In that contemplation an image was revealed to Valjean which serves to influence the way in which Valjean will go on to act:

Thus he contemplated himself, so to speak, face to face, and at the same time athwart this hallucination, he perceived in a mysterious depth a sort of light which he at first took for a torch. On scrutinising this light which appeared to his conscience with more attention, he recognised the fact that it possessed a human form and that this torch was the Bishop.²⁰

Valjean's conscience reveals an image of the Bishop. What does this image of the Bishop mean for Valjean? I think that, here, the conscience is placing potential 'identities' before Valjean’s mind for assessment. The agent can then choose whether or not to endorse one of those potential identities. Admittedly, from this excerpt alone, it is unclear what exactly the

---

¹⁹ (1867), p417.  
²⁰ (1867), p420.
term 'conscience' refers to. It could simply refer to Valjean’s general capacity to practically deliberate. Nevertheless, I think that the fact that conscience is more usually referred to in the way that I suggest (it is referred to in the identity-implicating way) and the fact that (as I shall go on to argue) it is that interpretation that contributes most significantly to the meaning of the narrative, is instructive. Consider, for example, that as Valjean continues in his reflection, the term 'conscience' to refer to the process of weighing up different prospective identities for internalisation:

His conscience weighed in turn these two men thus placed before it, the Bishop and Jean Valjean. Nothing less than the first was required to soften the second. By one of those singular effects, which are peculiar to those sorts of ecstasies, in proportion as his reverie continued, as the Bishop grew resplendent in his eyes, so did Jean Valjean grow less and less and vanish.

In the passage just quoted, Valjean is contemplating two different personas: his earlier 'brutish' persona, and a new persona inspired by the Bishop (and the Bishop's moral direction to Valjean). Again, this description supports a peculiarly identity-implicating interpretation. Later in the novel, when summarising this crisis of conscience, this identity crisis, Hugo makes the following claim:

From that moment forth he was, as we have seen, a totally different man. What the Bishop had wished to make of him, that he carried out. It was more than a transformation, it was a transfiguration.

This incident is not interpreted by either Hugo, or the character of Valjean himself, as merely a point at which Valjean determines unequivocally to pursue certain values. It is a point where Valjean decides to become someone else. Hugo describes it as not just a transformation, but a transfiguration, because the new persona that he is to take on is one that is much more 'exalted' or praiseworthy than the prior one.

---

21 And of course, Hugo may well intend to use the terms in several different ways at different points in the novel. In any case, it is the account that I defend which is the best interpretation for our purposes.

22 (1867), p421.

23 (1867), p776.
In giving such a lengthy description of the transfiguration event I wish to emphasise two things. First, I want to draw out the repeated connections Valjean makes between how he ought to behave and the sort of person that he is. Second, I want to draw out the way in which that process, of appealing to the sort of person he is, is described by Hugo as an operation of conscience.

Valjean’s identity crisis in the transfiguration event can be summed up as follows: Valjean’s internal dialogue is interrupted by the Bishop’s kindness and his own repugnance at stealing from a child. Valjean questions whether to carry on as the brutish Valjean or whether to instead adopt a more virtuous persona inspired by the Bishop. His conscience operates by presenting various ‘characters’ as ideals (such as the character of the Bishop) before his mind, it is then up to Valjean to decide which of these, if any, he is to endorse and commit to acting in accordance with. This case does not involve a transition from the absence of a conscience to the possession of a conscience. More plausibly, it involves a fundamental transition from a previous identity to a burgeoning new identity that will now provide the content of his conscientious commitments. This provides a base from which the agent can then go on to consult his conscience in practical deliberation, and ask of himself ‘am I the sort of person who would do this?’.

We shouldn’t be too quick to judge the transfiguration event as the point where Valjean’s identity transitions from the relatively vicious to the relatively virtuous. If Valjean fails to ever engage this freshly endorsed identity in his practical deliberation, then we would not be inclined to say that there genuinely has been such a transfiguration, a switch to the new identity (this intuition motivates the ‘practicality’ and ‘endurance’ requirements, a point I shall come to shortly). That is why it is important to consider the second case of conscience being implicated in Valjean’s practical thought. It is in this case that the commitments referred to in the transfiguration event are confirmed as his conscientious commitments.

24 On the face of it, at least. It might be possible to think that the previous ‘identity’ of Valjean, as a brute, was not a bona fide identity as it was something that he had inculcated without reflection and acted on out of impulse. Perhaps possession of an identity requires active endorsement on behalf of the agent. This will depend on precisely what notion of identity is implicated in conscientious identity; that is the subject matter of the following two chapters.
Several years have passed since the transfiguration event. Valjean has become a wealthy industrialist and mayor, and is going *incognito* as Monseur Madeline. By this stage he is widely reputed as a benevolent and generous figure through his dealings with the townspeople and workers in his factory. Valjean finds himself confronted with a serious moral crisis, however, when he discovers that, in a case of mistaken identity, an innocent man, Champmathieu, is on trial for Valjean’s own past crimes. In disguising himself as Madeline, Valjean had been in breach of his parole conditions. And there is the still outstanding crime of robbing Little Gervais. Valjean agonises over what he ought to do.

I will describe what I take to be key steps in this deliberation in order to draw out the precise role that conscience plays in relation to Valjean's actions and the other practical considerations that Valjean attends to. Valjean admits that his first thought is one of instinct; to turn himself in and save Champathieu. But Valjean does not immediately act on that instinct. Instead, he begins to deliberate. He wonders whether this is all part of God's plan; a plan for Champmathieu to take the fall for Valjean's own crimes. Valjean suggests some reasons as to why God would will this:

> In order that I may continue what I have begun, that I may do good, that I may one day be a grand and encouraging example, that it may be said at last that a little happiness has been attached to the penance which I have undergone, and to that virtue which I have returned.

Valjean goes on to reject this as God's plan for him:

> Had he not another and a grand object, which was the true one - to save, not his person but his soul; to become honest and good once more; to be a just man? Was it not that above all, that alone, which he had always desired - which the Bishop had enjoined upon him - to shut the door on his past?

---

25 (1867), p780.

26 (1867), p793.

27 (1867), p797.
More plausibly, Valjean reasons, the purpose of his own actions is to act justly; this is what God and his intermediary, the Bishop, implore him to do. In the transfiguration event, we saw Valjean describe an image of the Bishop as entering into his mind. We see a similar description from Valjean when talking about his own practical deliberation:

He felt that the Bishop was there, that the bishop was present all the more because he was dead, that the Bishop was gazing fixedly at him, that henceforth Mayor Madeline with all his virtues, would be abominable to him, and that the convict Jean Valjean would be pure and admirable in his sight; that men beheld his mask, but that the Bishop saw his face; that men saw his life; but the Bishop beheld his conscience.28

The imagined presence of the Bishop leads Valjean to reason that it is better for him to be known as the criminal, and in doing so save an innocent man, than to carry on under the guise of ‘Monseur Madeline’ for selfish reasons. Again, the term 'conscience' is used in referring to this specific mental operation of a persona or identity being placed before the agent's mind for assessment. From Valjean's perspective, his conscience is subject to this idealised character of the Bishop and what the Bishop would want Valjean to do. At this stage, Valjean does not judge this consideration (the purported dictate of his conscience) to be conclusive.

Valjean goes on:

One moment he said to himself that he was perhaps taking the matter too keenly; that, after all, this Champmathieu was not interesting, and that he had actually been guilty of theft.29

Here Valjean is wondering whether it is really so unjust to subject an innocent man to such a punishment, given that Champmathieu is guilty of other crimes anyway. Valjean quickly dismisses this consideration; it would be obviously unfair to allow Champmathieu to suffer for Valjean's own offending simply because he has committed some other crimes.

Then, Valjean examines his own motives; perhaps the selfish action is to hand himself in. Plausibly, Valjean ponders, he is motivated to hand himself in by the fact that he will feel guilty if he does not do so. And this motivation, a desire to avoid feeling guilty, is a selfish

28 (1867), p799.

29 (1867), p804.
motivation. Compounding the selfishness of this act is the serious adverse consequences that Valjean’s imprisonment would occasion on the townspeople and workers who rely on him:

What if I denounce myself? I am arrested; this champmathieu is released; I am put back in the galleys; that is well - and what then? What is going on here? Ah! here is a country, a town, here are factories, an industry, workers, both men and women, aged gransires, children, poor people! All this I have created, all these I provide with their living; everywhere where there is a smoking chimney, it is I who have placed the brand on the hearth and the meat in the pot...I take myself off, everything dies.\(^{30}\)

Following this, Valjean hears a voice in his head reminding him that the only standard he will be judged by are not the standards set by other people, but the standards set by God:

This voice, feeble at first, and which had proceeded from the most obscure depths of his conscience, had gradually become startling and formidable, and he now heard it in his very ear. It seemed to him that it had detached itself from him, and that it was now speaking outside of him. He thought that he heard the last words so distinctly, that he glanced around the room in a sort of terror.\(^{31}\)

This voice is the last we hear of Valjean's deliberation before he goes to Arras and turns himself in. We can therefore suppose that it is this consideration that persuades Valjean that he must do so. Again the term 'conscience' is used to refer to a specific type of consideration. It is a 'personified' consideration, similar to the image of the Bishop. Except instead of an image the personage of the Bishop is revealed to Valjean's mind as a voice rather than an image. This is another way in which Valjean's Bishop-inspired identity is implicated in Valjean's deliberation.

From this lengthy description of Valjean’s practical thought, I suggest that we can extract the core requirements that any account of conscientious identity must meet. But before doing that, it would be worthwhile to summarise the rather bewildering number of steps in Valjean’s deliberation. Valjean's constant vacillation (though perhaps an accurate description of how we deliberate on occasions of moral crisis) makes it difficult to follow the exact progression of Valjean's deliberation and where conscience fits into it.

\(^{30}\) (1867), p811.

\(^{31}\) (1867), p824.
A useful way of keeping track of the deliberation is to organise the different considerations appealed to by Valjean by whether they counted for turning himself in, or counted against it. I will begin with the considerations that Valjean judged to count against: First, there is the thought that it is God's plan for Valjean that he not hand himself in so that he may serve as an example of how the virtuous life can be rewarding to agents. Second, there is the fact that it is established that Champmathieu is a thief and therefore letting him be punished for Valjean's crime is not a major injustice. Third, Valjean considers that his inclination to turn himself in is motivated by a selfish desire to salve a guilty conscience. Fourth, Valjean considers how turning himself in will greatly reduce the well-being of all those who rely on him.

The first consideration that Valjean judges to count in favour of turning himself in is the fact that his resolution (at the time of the transfiguration event) was to become a just and honest person. And that would seem to require that Valjean turn himself in. It is the image of the Bishop that leads Valjean to think this; an image that Valjean recognises as the output of his conscience. The second consideration is an internal voice, rather than an image, which by Valjean's reckoning, provides the standards that he will be held to. Both the image and the voice of the Bishop are referred to by Valjean as manifestations of conscience. And both of them, I argue, make reference to Valjean's identity. The 'gaze' and the 'voice' that Valjean refers to, compelling him to turn himself in are manifestations of the Bishop-inspired identity which was first formed in the transfiguration incident.

I acknowledge that my interpretation, that the term 'conscience' picks out the specifically identity-implicating considerations in Valjean's deliberation, is not the only interpretation of conscience supported by Valjean's language. For example, Valjean's invoking of the standards set by God in the final use of the term 'conscience' might support a religious conception of conscience where the conscience constitutes a conduit to God's will. Or, as I mentioned earlier, sometimes Valjean refers to the whole procedure of practical deliberation using the term 'conscience'. This could support an interpretation that, in fact, the term 'conscience' does not refer to specifically identity-implicating considerations but merely includes those considerations amongst others. But, for now, it is enough that this particular way of articulating conscience, as identity-implicating, is present in Valjean's work, whether or not it is the only conception present.

Now we are in a position to be more specific about the connection between the two cases. Both cases are crises of conscience. The first event, the transfiguration, involves a major shift
in the identity which is constitutive of Valjean's conscience. Certain events, the kindness of
the Bishop, his revulsion at his own actions towards Little Gervais, have led him to question
the identity upon which his conscience is based; have led him to question who he is. Then, in
the second case, we have a situation where this more mature conscience is employed in
practical deliberation. It is the conscience, as personified within his mind as ‘the Bishop’ that,
in the end, compels Valjean to hand himself in. It is not that there is an identity formed in the
first incident, and then in the second incident this identity is appealed to in Valjean's practical
deliberation. Valjean's identity is not settled by the fact that he has resolved to act in that way
at the conclusion of the transfiguration event. If Valjean had never acted on that resolve, if he
had continued to act in the same way that he had been, then we would have reason to say that
he had never truly taken on that new identity. What we should say is that there are two
important stages in the development of Valjean's identity (and thus his conscience). In the
first case, we have Valjean's own resolutions that he will act in a particular way, in the
second case we have action which confirms that identity as genuine and applies it to his
actions.

5.4. The Distinctness Requirement and Practical Identity in Korsgaard and Taylor

Let's assume, for now, that Valjean's conscience does pick out peculiarly 'identity-
implicating' considerations. It then remains to be explained what notion of 'identity' is being
referred to here. The first excerpt from Les Misérables that I referred to talked of Valjean 'no
longer being the same man'. And then, later, there was also a suggestion that he had 'become
a different man'. We also see the use of the terms 'transformation' and 'transfiguration'. Often,
in philosophical contexts, this sort of language is thought to implicate questions of personal
identity. Personal identity concerns what it is that allows for a person to persist through time
and what it is that separates one person from another.

If we interpreted the notion of identity in the Valjean cases as personal identity, then the
cases would involve Valjean literally stopping being one person and becoming another. The
claim would be that the individual that exists before and after the transfiguration event might
be referred to by Hugo using the name ‘Valjean’ but they are actually two distinct characters
in the novel just as the Bishop and Champmathieu are distinct characters; Valjean does not
persist over this event. It should be obvious that this interpretation of conscientious identity is
implausible: Valjean is still the same person before and after the transfiguration event. Our understanding of the character Valjean as triumphantly overcoming his past requires us to see these different psychological events as things that occur throughout the lifetime of one particular person.

In The Constitution of Selves Marya Schechtman argued that we need to distinguish two questions that pertain to the identity of agents. The first question is the re-identification question: what makes it the case that an agent at time $t_1$ is the same agent as the individual at time $t_2$. In other words, it is the question of persistence. The second question is the characterisation question. This question asks “which beliefs, values, desires and other psychological features make someone the person she is?” Clearly, the re-identification and characterisation questions are distinct as it may be that an individual counts as the same person for the purposes of the ‘re-identification’ question, but not the purposes of the characterisation question. Sometimes, if an individual does something completely out of character, we might say it wasn’t really ‘them’ doing it, in the characterisation sense, while still acknowledging that it was the very same person in the re-identification sense.

Would it help to interpret Valjean's transformation as a fundamental change in the way he is characterised? Was there a fundamental change in the 'beliefs, values, desires and other psychological features' that make him Valjean? Perhaps in taking on a new Bishop-inspired identity, and then engaging that identity in practical deliberation, Valjean changes certain psychological features that may have made him ‘who he is’. This would then justify Valjean in saying 'I'm not that person anymore'. However, interpreting these cases as involving a change in characterisation still doesn’t get us very far in explaining what is going on in the two cases. We still need to understand what it could mean for certain psychological features to make someone ‘who they are’. In chapters six and seven I will consider various different ways of answering the characterisation question and assess them as to whether they constitute good accounts of conscientious identity. But in order to do that we first need a criteria by

---


34 Though this does not preclude the possibility that there is an account of identity capable of answering both questions.
which we can assess those different accounts of identity: I claim that we can extract that criteria from the two Valjean cases.

The first requirement of conscientious identity is what I call the distinctness requirement. I have emphasised the way in which Hugo often uses the term ‘conscience’ to refer to the specifically identity-implicating considerations, a practice that is, perhaps, at odds with a common practice of using the term ‘conscience’ simply to refer to any practical or moral deliberation. This draws out something important about conscience (and something importantly wrong about our everyday talk of conscience). There is a distinction between two sorts of practical consideration: those that implicate the agent’s conscientious identity and those that do not. And both of these considerations can be genuinely normative and reason-providing for the agent. In this section I want to draw out the distinctness requirement by reference to accounts of ‘practical identity’. The accounts of Christine Korsgaard and Charles Taylor are both unable to capture the way in which only some of our genuinely reason-giving normative considerations implicate identity.

Perhaps the most influential recent account of the way in which identity is implicated in practical deliberation is the account offered by Christine Korsgaard in *Sources of Normativity.* Korsgaard’s aim in that book is to explain the normativity of our practical considerations; what makes it that practical considerations are aimed at what we ought to do and sometimes give us good reasons to act in one way rather than another. Korsgaard argues that the normativity of considerations arises from the reflective structure of the human mind. And once we analyse the reflective structure of the human mind, Korsgaard argues, we see that a notion of identity plays a crucial role.

Korsgaard points out that we often think of ourselves in terms of “roles and relationships, citizenship, members in ethnic of religious groups, causes, vocations, professions and offices”. For example, in my own case I can think of myself as a son, a citizen of New Zealand, a member of the Pakeha ethnic group, an enrolled barrister and solicitor, and a student of philosophy. Korsgaard holds that each of these ‘roles’ constitutes a ‘conception of my practical identity’. A ‘conception of practical identity’ is a “description under which you

---


value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking”. The set of all of an individual’s conceptions of practical identity together constitute that individual’s ‘practical identity’ (we might think of this set as the agent’s ‘all-things-considered’ practical identity). Korsgaard argues that the normativity of any practical judgement is necessarily tied to the agent's practical identity.

While it may well be true that, as an empirical matter, we often deliberate in ways that implicate a conception of practical identity, why must it be that all our normative judgements are structured in that way and what is it that makes them genuinely normative? After all, the conceptions of my practical identity I listed seem completely contingent; I don't need to have any of them. How, then, could such a notion be the grounding for our judgements about what we ought to do (rather than, say, judgements about what we would like to do)? In Korsgaard’s view, it is the reflective structure of practical deliberation that demonstrates how such conceptions could be genuinely normative.

To understand this, a good starting point is to think about what it would mean to never act on a normative reason. Plausibly, a wanton, someone who never reflects on what they ought to do, who simply acts on inclination, never has a normative reason to do anything. Perhaps some beings (such as many non-human animals) cannot help but always act on inclination. So, normative reasons require, at a minimum, that the agent is capable of reflecting on her inclinations and thereby deciding to act one way rather than another. Let's call this sort of reflection on our inclinations, first-order reflection. Could this ability, in itself, give us a normative reason to act in one way rather than another? It seems not - the agent could decide to endorse or reject one of their inclinations for whatever grounds they choose. Surely such grounds could not constitute a good normative reason. Korsgaard suggests another level of reflection. If I reflect on my capacity to reflect I see that I am a self-conscious being. I see that I am an individual that stands over and above my desires. This is second-order reflection. If we then reflect on that capacity, we see that to be able to self-consciously deliberate is to be capable of having a particular conception of my practical identity. We can call this third-order reflection. So Korsgaard would have it that, if we are sufficiently reflective, we realise that all of our practical reasons must implicate a conception of practical identity.

———

With these different levels of reflection are we any closer to an account of what will make any one of those conceptions genuinely normative? This leads us to the final step in the puzzle. If we introduce a fourth-order of reflection we can see that here is one conception of our practical identity that is not completely contingent. Our very ability to deliberate on a conception of our practical identity is what defines us: "a human being is a creature that needs a practical conception of her own identity, a conception of who she is that is normative for her." Remember, at the previous levels of reflection it was possible to endorse or fail to endorse any given conception, but at this four-order level it is no longer contingent.

Korsgaard claims:

You are an animal of the sort I have just described. And that is not a merely contingent conception of your identity, which you have constructed or chosen for yourself, or could conceivably reject. It is simply the truth.

This is one conception of her practical identity that the agent cannot fail to endorse as long as she is sufficiently reflective.

Charles Taylor is another contemporary moral philosopher who has argued that there is a necessary connection between normativity and identity (or, as Taylor often calls it, selfhood). In Sources of the Self: The Making of the Modern Identity Taylor's purpose is the reverse of Korsgaard's. While Korsgaard sought to explain the nature of normativity through reference to an agent's identity, Taylor seeks to establish the nature of the self or identity by reference to our normative judgements. Taylor claims that our identities are many-sided and multi-tiered in the same way in which Korsgaard thought an agent's all-things-considered practical identity is constituted by multiple conceptions of practical identity. Taylor replaces the term 'conception of practical identity' with the notion of a 'possible articulation' of one's identity. A 'possible articulation' can involve "universally valid commitments" (such as being a Catholic or an Anarchist) or a "particular identification" (such as being Armenian or Armenian or


40 And, as it happens, this fundamental identity is also a moral identity, as consistency requires that we value the humanity of others as we value our own. So, Korsgaard suggests, this can be the ground for moral normativity.


Quebecois). A universally valid commitment is an articulation that purports to have universal scope; to be a Catholic is to have views about how everyone, not just Catholics, ought to behave. Whereas a 'particular identification' is an articulation that does not have implications for how others ought to behave; perhaps identifying as Quebecois means that the agent ought to enjoy Poutine, but it does not mean that a British Columbian ought to enjoy Poutine.

For Taylor, 'possible articulations' and normative (including practical) judgements are interdependent. In order to be an agent we must be capable of making normative judgements, but just as it was for Korsgaard, in order to make a normative judgement, we must have an identity: An agent's identity consists in

the commitments and identifications which provide the frame or horizon within which I can try to determine from case to case what is good, or valuable, or what ought to be done, or what to endorse or oppose.

What is Taylor's argument as to why an identity (or possible articulation thereof) must underlie any even potentially normative judgements? Taylor argues that this claim follows from the way in which both identities and normative judgements both arise from the same socio-linguistic milieu and cannot be disentangled from each other. It is questions of identity, being able to be asked by others 'who are you?', and being able to ask it of oneself, that enables an individual to be 'placed' in her society of interlocutors. So in some sense, possessing an identity, is an essential part of inhabiting one's socio-linguistic world. And it is only as a result of being an inhabitant of that world and thus having relationships to others that the agent is able to make normative judgements. If one did not have such an identity, one would be 'at sea' and 'morally disoriented'.

Recall that, for Korsgaard, the absence of a conception of practical identity would mean that an agent had no reason to do anything and, in fact, to go on living at all. Lacking an identity is even more problematic according to

45 I say 'potentially normative' here, as it is not just the judgements which involve genuine normative reasons that Taylor claims must implicate identity, but any judgement which even purports to be normative.
Taylor's account. According to Taylor, an agent who lacks an identity is not a proper interlocutor within her social world.

It is not required, for our purposes, to determine whether Korsgaard or Taylor have given correct accounts of the connection between normativity and identity. The purpose of considering their accounts is simply to illustrate why there is a distinctness requirement for conscientious identity and how both accounts fail to satisfy that requirement. For both authors, identity is the normative ground for all practical judgements. On Korsgaard’s account of normativity, if an agent is to have any normative reason to act at all, then that reason must be grounded in a conception of her practical identity. Consequently, her account cannot allow for some of the agent’s practical considerations to be genuinely normative while not implicating that agent’s practical identity. Whereas an account of conscientious identity must allow for the fact that some of our practical judgements implicate conscientious identity and some do not.

For Taylor, an agent cannot make a normative judgement without implicating her identity as normative judgements do not make sense without an agent possessing an identity. On Taylor’s account of identity all of the practical considerations that Valjean attended to involve his identity, so his account also cannot distinguish the conscientious and non-conscientious considerations: it too fails the distinctness requirement.

One might respond on behalf of Korsgaard and Taylor that their accounts can, in fact, account for the distinctness requirement. Either account might do so in the following way: all the considerations that Valjean appeals to involve identity, it is just that it is in only some of those cases that the identity involved is made explicit before Valjean’s mind. In the case of Korsgaard’s account, one might argue that all of Valjean’s considerations implicate a conception of his practical identity. When he considers the adverse consequences on the townsfolk of handing himself in, he is considering his conception of himself as ‘the mayor’ or ‘the benevolent factory owner’. It is just that, in that case, the fact that that consideration implicates his identity is not something that he is immediately conscious of. Those considerations that I referred to as ‘identity-implicating’ (as presented through the image and voice of the Bishop) were simply those considerations where their ‘identity-involving’ aspect is on the surface for all to see. A similar argument might be made with respect to Taylor’s account of identity. All of the considerations implicate his identity, in virtue of the fact that they involve normative considerations, but it is only in those cases that make reference to the
Bishop that this identity-involving aspect is brought to the surface. When Valjean considers things like the consequences that his actions would have on the townsfolk, he is implicating his identity at some deeper level.

The reason why I think that this response does not work is methodological. Recall that the purpose of the discussion of both of the Valjean cases is to try and get clear on the descriptive constraints that are the starting point for our analysis of conscience (or, more specifically, our analysis of conscientious identity). We do not come to assess those requirements with reference to a prior held notion of conscience. To claim that the appearance of distinctness between conscientious and non-conscientious practical considerations is just an illusion and that, at bottom, both sorts of consideration implicate identity, is just to deny the distinctness requirement as I have stated it.

Having considered Korsgaard and Taylor’s accounts of practical identity in detail we are now in a good position to describe what precisely the distinctness requirement involves. The distinctness requirement holds that any account of conscientiousness must be able to explain how it is that only some of the potentially normative considerations that impact on our practical deliberation implicate our conscientious identity. Two influential recent accounts of identity in moral philosophy which may have otherwise been thought useful in analysing conscientious identity fail to satisfy this requirement. As we shall see in the next two chapters, other accounts of identity such as identification and neo-Aristotelian character also fail to satisfy this requirement.

5.5. The First-Personal, Practicality and Endurance Requirements

The second requirement of conscientious identity becomes clear when we think about the way in which Valjean’s identity-implicating considerations are revealed to the reader by Hugo. Recall that my original gloss on identity in Valjean's deliberation was that it referred to Valjean's view of himself. It is this element of conscientious identity that the second

I again refer to 'potentially' normative here as the distinction I am making here is conceptual rather than substantive. My claim is not that either of these considerations actually give the agent a reason for action. At this stage, I want to remain neutral on that issue.
requirement is intended to capture. Valjean's identity crises are not described as if being observed by an outsider. The account we are given is from the 'inside'. It is an account of Valjean's first-personal perspective. Valjean's own thoughts about the sort of person he is are one crucial aspect of his conscientious identity. I call this aspect of identity, the first-personal requirement.

I have mentioned a first-personal requirement several times in this thesis thus far. Earlier, I referred to the way in which authors like Adam Smith and Gilbert Ryle claimed that a judgement of conscience is a claim about what I morally ought to do, not what anyone else morally ought to do. Here, I strengthen this requirement such that conscience involves the way in which an individual sees herself.

It might be argued that such a requirement is too strong: if an agent must always have the thought that some conviction is implicated in the way in which she 'sees herself', agents would only very rarely have conscientious convictions. After all, it is not common in our deliberation that we explicitly appeal to this first-personal perspective as Valjean does. It is only the sheer depth of Valjean’s deliberation that brings out this aspect, but we would still (perhaps) be inclined to say that he possessed the conscientious conviction even if he did not do so. The correct response here, I think, is to alter the requirement so that it is a counterfactual. The requirement is not that an agent actually must actually think of herself as the sort of person who would do that at any given moment. The requirement is, rather, that if the agent were to pay sufficient attention to the action in question, she would not see herself as the sort of person who would do x. This is broadly analogous to the sense in which, on Korsgaard's account, an agent need not be aware that she possesses a conception of her practical identity to employ it in her deliberation. All that was needed was that the agent would be aware she operated on such a conception if she were to reflect on her actions sufficiently.

I considered two cases of conscientious thought. The reason that I described both cases, rather than simply one of the cases, I suggested, is that we can be sceptical as to whether some putative identity is actually an agent’s conscientious identity unless that putative

49 Or, at least, convictions that involve conscientious identity. I discussed in the previous chapter how an agent's conviction can be conscientious even if it does not involve any reference to the agent's identity (i.e. as long as that conviction is moral or communicative it can also count as conscientious).
identity actually affects that agent’s deliberation or action. This suggests a third requirement of conscientious identity, a practicality requirement. An agent does not possess a conscientious identity of the sort we are interested in unless she possesses something that affects the way both her practical deliberation and her action. Recall that in chapter three I discussed Brownlee’s consistency requirement. This is the requirement that an individual think and act consistently with her convictions if they are to be communicative. I suggested that while this was a plausible requirement of conscientiousness it didn’t seem to have anything to do with the communicative principle per se. It does, however, have something to do with the identity principle. In my view, Valjean’s Bishop-inspired conscientious identity is confirmed by the fact that it influences (and is indeed, determinative) in his thought about whether he should turn himself in to the authorities.

As I mentioned in chapter three, this requirement is closely related to the well-known thesis about moral judgements, moral motivation internalism; the claim that moral judgements motivate action (whether defeasibly or overridingly). But it is not the very same thesis. Whether or not one believes that moral judgements necessarily motivate, one ought to believe that conscientious convictions motivate. It is not absurd to say, I don’t think, that one can believe an action to be wrong, but not particularly care about one’s moral beliefs. But it does seem absurd to say that in the case of a conscientious conviction (or at the least, a conscientious conviction that implicates the individual’s identity), that one does not care for one’s conviction. To say that ‘I could not live with myself if I were to do that’ and to also say ‘but I don’t care about that’, sounds absurd. The sheer psychological power of an identity-implicating conviction is such that it must motivate the agent.

The final requirement of conscientious identity is the endurance requirement. Arguably, the import of Brownlee’s consistency requirement is not fully captured by the practicality requirement. It is not enough that the individual indulge her identity in her practical deliberation on just one occasion. What is required is that that conviction regularly crop up in her thought. A conscientious conviction is not one that lasts for a mere moment, but rather one that endures. Valjean’s conscientious identity is not just what he determined he would be at the time of the transfiguration event. Several years intervene between the transfiguration and deliberation events and during that period Valjean’s new identity affects his behaviour, influencing him to become a benevolent figure. The endurance requirement is clearly connected to the practicality requirement as the implication of one’s identity in one’s practical deliberation constitutes evidence that the identity in question has endured. This
requirement needs to be listed as a separate requirement from the practicality requirement because theoretically an agent could ‘form an identity’ and deploy it in her deliberation at a single point in time. Such an identity would not be sufficiently enduring to constitute the agent’s conscientious identity.

It is the endurance requirement that leads me to incorporate a narrative element into my account of conscientious identity. I argue that the best way for an account of identity to incorporate this requirement is to hold that the identity must play an integral role in the story or 'narrative' of that individual's life. I suggest that satisfying the endurance requirement is not simply a matter of that identity lasting a certain period of time, but, rather, having an enduring significance in that individual's life.

5.6. The Meaning of Conscience in Les Misérables

Where have we got to thus far? The aim of this chapter has been to try and set out more precisely the identity principle of conscience. In order to draw out the constraints on such a notion, I suggested, we should look at a literary case of conscience being applied just as Jonathon Bennett did in The Conscience of Huckleberry Finn. I suggested that we look at two cases of Valjean’s practical thought in Les Misérables as paradigm cases and in light of that extracted four requirements of conscientious identity. I mentioned that there are two tests of the paradigm status of some putative case that are available to us when we use literary cases. The first is simply that a character recognised as well-drawn will be one that applies coherent concepts. The second is that sometimes a particular articulation of some concept will contribute to the meaning of the narrative of the novel. In this way the interpretation that is involved in a paradigm case is not simply the one that the theorist finds intuitive. Now that we have explored those cases in detail and I have suggested some implications of those cases we are in a position to assess how exactly these cases, and this way of thinking about conscience, is implicated in the meaning of the narrative of the novel. I mentioned earlier that Hugo sees these two cases as worth singling out as the identity crises of the novel. The reason the Hugo mentions both cases, I submit, is that these two cases are both essential in explaining the narrative of the novel.
So what is the overall meaning of *Les Misérables*? Hugo offers the following description in the preface:

So long as there shall exist, by virtue of law and custom, decrees of damnation pronounced by society, artificially creating hells amid the civilization of earth, and adding the element of human fate to divine destiny; so long as the three great problems of the century—the degradation of man through pauperism, the corruption of women through hunger, the crippling of children though lack of light—are unsolved; so long as social asphyxia possible in any part of the world; in other words, and with a still wider significance, so long as ignorance and poverty exist on earth, books of the nature of *Les Misérables* cannot fail to be of use.50

This preface suggests to me that, in Hugo’s view, *Les Misérables* has a moral purpose. Its purpose is to draw the public’s (the reader’s) attention to the way in which, at least partially, society itself causes the misery of the worst off. Does this purpose have any discernible connection to Hugo’s in depth explorations into the nature of conscience? I think this becomes clearer when we consider what Hugo says towards the end of the novel:

The book which the reader has under his eye at this moment is, from one end to the other, as a whole and in detail, whatever may be its intermittences, exceptions and fault, the march from evil to good, from the unjust to the just, from night to day, from appetite to conscience, from rottenness to life; from hell to heaven, from nothingness to God. Point of departure: matter; point of arrival: the soul. The hydra at the beginning, the angel at the end. 51

There are several aspects of Hugo’s account here that are worth paying attention to. First and foremost is the recognition that this novel concerns the transition from appetite to conscience. In the beginning, the brutish Valjean merely acts on his appetites. His act towards Little Gervais is one of beastly instinct. But in the two cases I described, Valjean develops a conscience. There is also reference about the change from unjust to just; again something that captures accurately the transition for Valjean. And at the end there is a reference to a transformation; from a hydra to an angel. This captures the transition of Valjean from a vicious being to a virtuous one.

50 (1867), p32.
51 (1867), p4169.
This explanation of the meaning of *Les Misérables* complements the description of its moral purpose referred to earlier. *Les Misérables* draws attention not simply to the way in which we humans cause so much of the misery around us but also to the way in which individuals are capable of escaping those miserable circumstances. Despite their ill-treatment at the hands of others and the institutions of society, individuals can rise above those circumstances and become virtuous.

The notion of conscientious identity I have drawn out, one involving the four requirements, is crucial to explaining the way in which individuals can fundamentally change and transcend their circumstances. Conscience articulated in other ways, as an internal voice of God, or as the overall capacity to practically reason would not serve this overall meaning of the novel as well. And this makes the *Les Misérables* case a good paradigm case because the overall meaning of the narrative is better understood if conscience is interpreted in the way I suggest.

A prominent Hugo critic Katherine Grossman backs this interpretation:

> much of the book is devoted to the expansion of identity through the development of conscience which perpetuates his [Valjean’s] past selves as he creates new ones

As I mentioned earlier, I do not mean to suggest that my conception of conscience is the only one implicated by Hugo. But, I suggest, it is one of the conceptions implicated to explain broader themes in the novel which is evidence of its relevance as a paradigm case.

---

52 (1994), p141.
Chapter Six: Conscientious Identity and Identification

In chapters one and two I described a relatively neglected tradition of analysing conscience, *conscience as identity*. I suggested that elements of this approach were present in the conceptions of conscience defended by Adam Smith and Immanuel Kant. I argued that we can see *conscience as identity* in its mature form in the accounts of conscience of Heidegger and Arendt. In the final section of chapter four I argued that the *conscience as identity* approach should be incorporated into my analysis of the cluster concept of conscience in the form of a principle of conscientious identity. In articulating and defending such a principle over the next two chapters I ‘update’ the old *conscience as identity* approach to conscience through consideration of modern accounts of identity defended in moral philosophy. Just as Brownlee drew out four requirements that are constitutive of the communicative principle, in the last chapter I drew out four requirements that I claim are constitutive of the identity principle. Through a lengthy description and interpretation of conscience in Victor Hugo’s *Les Misérables*, I argued that there are four core requirements of conscientious identity; the first-personal, distinctness, practicality and endurance requirements. In articulating the distinctness requirement I addressed two influential accounts of practical identity, the accounts of Christine Korsgaard and Charles Taylor. I argued that neither account (at least, in the way the accounts were articulated by those authors), would satisfy the distinctness requirement.¹

In this chapter, following on from my discussion of Korsgaard and Taylor, I consider some other important accounts of identity from recent moral philosophy. My focus in this chapter is on what has been called accounts of *identification*. I assess several different accounts as to whether they satisfy the four requirements of conscientious identity. In the first section I describe Harry Frankfurt’s account of identification as second-order volition: Frankfurt argues that an individual identifies with those desires that are the object of a second-order volition; a second-order desire that some first-order desire be effective. In sections two and three I move on to a discussion of Michael Bratman’s account of identification. In section two I describe Bratman’s general planning theory of agency. This is not only essential

¹ In chapter seven I argue that a notion of self-conception, somewhat similar to Korsgaard’s notion of a conception of practical identity is a crucial element of conscientious identity.
background for understanding Bratman's analysis of identification, it is also useful for its articulation of a notion of a ‘quasi-policy’ (ideal) which I incorporate into my own analysis of conscientious identity in chapter seven. Then in section four I discuss Bratman’s account of identification which appears to come in two separate formulations; one formulation holds that identification is policy-based, the other holds that it is based on self-governing policies. In section four I evaluate Frankfurt and Bratman’s accounts as to whether they satisfy the four requirements. I argue that none of those accounts can, primarily because they do not satisfy the first-personal requirement.

6.1. Frankfurt and Identification

The type of identity I am going to consider is not so much an account of identity but a family of accounts; accounts of identification. I will describe three particular formulations of agential authority, the formulations articulated by Harry Frankfurt and Michael Bratman, but my conclusion, that identification does not satisfy the requirements of conscientious identity, is intended to apply to any account of identification. What is an account of ‘identification’? It is an account of what makes it the case that an individual has a certain authority over her thoughts and actions; an account of what makes an individual’s actions her own. To use a straightforward example: when an individual is struck below the knee with a reflex hammer, would we say that that was an action of the agent's or not? We might reasonably answer this question either way. On the one hand, it is the individual's action because it was her leg rather than some other person’s, that moved and it moved (we can assume) due to signals sent to her leg by her brain. On the other hand, it would be reasonable to say that it is not her action.

2 The only accounts of identification I consider are desire-based or ‘non-cognitive’ accounts of identification. Identification is defined in terms of an individual's desires or volitional states. But other authors defend cognitivist accounts of identification (for example, Watson (2003) and Raz (1999)) where identification is determined not by the agent's desires but her beliefs about what is valuable. For a recent summary of cognitivist and non-cognitivist approaches to identification see Wallace (forthcoming).

My two principal criticisms of identification; that identification cannot account for the first-personal requirement or the distinctness requirement are criticisms that apply equally to cognitivist and non-cognitivist accounts.

3 The term is Bratman’s (2001).
because she did not have sufficient control over it. At one point Frankfurt describes this as
the distinction between actions which are hers in a "gross and literal" sense and those that are
_ truly_ hers.⁴ Both Frankfurt and Bratman refer to this fact of ownership of one’s desires as
'identification' with one’s desires.

Is there any initial plausibility to thinking about an appeal to conscientious identity as an
appeal to identification? Returning to the _Les Misérables_ cases, consider the way that
Valjean's thought began in the transfiguration case: Valjean remarked that it was ‘not
Valjean', but ‘the beast’ who stole from Little Gervais. In further expanding on this thought
Valjean remarked that he stole from Little Gervais out of instinct, rather than any considered
view that this is an acceptable thing to do. A natural interpretation of these observations of
Valjean is that his inclination to steal (an inclination that ultimately moved him to act) was an
inclination that was _not truly his_. On such an interpretation those practical considerations of
Valjean's that I described as 'identity-implicating' are those considerations that he identifies
with. In the rest of this chapter I examine several different accounts of identification before
eventually evaluating such a notion employing the four requirements of conscientiousness.

Perhaps the most influential account of identification is the account developed by Harry
Frankfurt. For simplicity's sake, in this chapter I will concentrate on Frankfurt's position as
expressed in "Freedom of the Will and the Concept of a Person".⁵ As the title suggests, in that
paper Frankfurt has two aims; to give an account of freewill and personhood. In trying to
achieve both these aims, Frankfurt introduces the notion of identification. Frankfurt argues
that it is the ability of an agent to 'identify' with some of her desires, to make some of those
desires her own, that distinguishes persons from regular animals and explains how we are
capable of acting freely.

---


⁵ (1971). Frankfurt subsequently realised some of the most obvious problems with his account in this paper and
alters his account substantially in “Identification and Externality” (1977) and “Identification and
Wholeheartedness” (1987). In his later work Frankfurt emphasises the importance of decisiveness and
satisfaction in an account of identification. Given that Bratman takes those revisions as the starting point for his
own account and since he, to some extent, incorporates those revisions into his own account, I don't separately
consider Frankfurt’s later views. As I mentioned earlier, the core problems of these accounts of identification
apply to any account of identification so it is unnecessary to be comprehensive in my description of different
accounts of identification.
In chapter five I explained that Korsgaard and Taylor’s accounts of practical identity were not explicitly offered by those authors as accounts of conscience at all. In light of this, the failure of those accounts to satisfy the distinctness requirement was not intended to be a critique of those accounts per se. Either account may well provide an excellent explanation of practical normativity, but they are not accounts of conscience. A similar caveat must be acknowledged when evaluating the accounts of Frankfurt and Bratman. In this thesis I am not directly interested in the criteria for personhood or the pre-conditions for freewill. The purpose of an account of conscience, according to me, is not to explain either personhood or freewill. The purpose of an account of conscience is, ultimately, to explain why a certain subset of an individual's convictions are valuable and worthy of respect. Those convictions are worthy of such respect (I argued in chapter four) because they are an expression of the individual's autonomy or humanity, acting in accordance with them is psychologically beneficial and expression of those convictions has societal value. For now, I am simply interested in whether the accounts of Frankfurt and Bratman satisfy the four descriptive requirements of conscientious identity.

Frankfurt argues that an agent has authority over her desires, identifies with her desires, when the desires in question are second-order volitions. To understand this account it is necessary to take apart the two ideas used in constructing the account and consider them in turn: First, the idea that desires are ordered in hierarchy (and hence can be second-order), and second, the idea that desires can be effective or ineffective (and hence constitute volitions). In chapter six I described Korsgaard’s account of practical identity by introducing different levels or 'orders' of reflection. Korsgaard emphasises that not only can we reflect on our desires, we can also engage in meta-reflection by reflecting on the fact that we reflect on our desires. Corresponding to that hierarchy of reflection we can posit a hierarchy of desires. My first-order desires are my ordinary desires or inclinations for certain objects or states of affairs in the world. For example, as a heroin addict I would have a first-order desire to inject myself with heroin. Or, I might that I experience the pleasure that follows from that injection. A

---

6 In fact, I described Korsgaard’s distinct account of conscience as moral consciousness at the beginning of chapter one.

7 Or to put it more precisely, they are not accounts of conscientious identity. In chapter four I argued that while identity is the most important and weighty condition of conscience, it is not the only one. An individual can make a conscientious judgement or have a conscientious commitment as long as she satisfies either the identity the moral, or the communicative principles of some combination of them.
second-order desire is a desire that takes a desire itself as the object of the desire. In the heroin addict case, this may be expressed as a desire not to desire heroin.

The second tool that Frankfurt uses to construct his account of identification is the notion of a volition. Consider the sentence, 'A wants X'. There is an ambiguity in such a sentence. It might mean, merely, that A possesses a desire to x; a desire which may or may not be defeated by another desire of that agent. It can be accurate for me to say 'I want to get drunk tonight' even though I don't in fact get drunk tonight due to the fact that that desire is overwhelmed by my desire not to drink and drive. But the sentence could mean something different. It could express the fact that A has a desire that she wills to be effective; a desire she wishes to act on. It is this second type of desire that Frankfurt refers to as a ‘volition’.8

Employing both these ideas, Frankfurt attempts to analyse personhood and freewill. To be a person is to be an entity that is capable of possessing second-order volitions.9 That is, someone who has a desire that a particular first-order desire be effective. This sort of volition can be illustrated in the following way: We can imagine that the heroin addict mentioned earlier might possess two first-order desires which are in competition with each other. On the one hand, she has a first-order desire to inject heroin. On the other hand, because she knows that heroin is bad for her in the medium to long-term, she also possesses a first-order desire not to inject heroin. She is a recovering and unwilling heroin addict as her second-order desire is for her first-order desire not to take heroin to be effective.

Frankfurt’s claim here is not that an individual must actually possess a second-order volition in order to count as a person. Completely unreflective individuals (sometimes called ‘wantons’) who always act on their inclinations (a type of first-order desire) still count as persons. But, Frankfurt argues, it is the capacity, the potential, to have second-order volitions that distinguishes persons from other beings. Frankfurt argues that the capacity to form second-order volitions also explains the nature of free-will: To be unable to decide which of your inclinations will be effective is to be an automaton.

Frankfurt introduces the term ‘identification’ when describing the above distinction between the wanton and the unwilling addict. The wanton heroin addict simply follows her whims;

she is controlled by her first-order desire for heroin. There is no higher order desire that either endorses or rejects that first-order desire. By contrast, the unwilling addict has a first-order desire for heroin but does not desire that that desire be effective. Frankfurt describes the unwilling addict in the following way:

In either case he does something he himself wants to do, and does it not because of some external influence whose aim happens to coincide with his own but because of his desire to do it. The unwilling addict identifies himself however through the formation of a second-order volition, with one rather than the other of his conflicting first-order desires. He makes one of them more truly his own and, in so doing, he withdraws himself from the other.¹⁰

I began my analysis of the *Les Misérables* cases by suggesting that those cases raised the characterisation question. These cases moved us to question which mental states were distinctive of Valjean; which mental states make him, him. Frankfurt's account is one potential answer to this question. The mental states which make an individual herself are those that are the object of her second-order volitions. This explains how it is an account of what makes an agent’s mental states (or desires at least) her own;¹¹ it is the fact that they are endorsed by certain higher-order desires.¹²

Where have we got to so far? I have introduced Frankfurt’s account of identification as a contender account of conscientious identity. His account holds that an individual identifies with one of her desires just in case the individual desires that *that* first-order desire be effective. Does such an account of identification offer a good explanation of Valjean’s conscientious thought in both the *Les Misérables* cases? This account does seem to explain some aspects of Valjean’s thought. In the transfiguration case, it seems reasonable to interpret

---

¹⁰(1971), p13. It must be noted that the distinction between the wanton addict and the unwilling addict is not the distinction between those who actually act on their desire for heroin and those who do not. The unwilling addict might, due to practical irrationality, give in to her first-order desire for heroin. But in doing so that would not mean that she identifies with that desire. The second-order volition of the heroin addict is simply the desire that one of her desires be effective, not which desires she ends up acting on.

¹¹It seems that we can ask the question of identification about other mental states (such as beliefs and emotions) as well. Can't we ask the question, what makes an emotion truly mine? For example, perhaps I feel revulsion at hearing of incest between two consenting adults (even though I know they are using contraception). But I might not identify with this feeling at all. It seems as though we can extent Frankfurt's analysis to such cases. I can explain my lack of identification with that emotion by the fact that I desire not to have that emotion.

¹²Schechtman acknowledges this as a potential answer to the characterisation question ((1996), p82) and points out the way in which it captures the agent’s self-interested concern, in particular.
Valjean as grappling with which first-order desire to decide to be effective; he is trying to determine which desire, if any, he will desire at a higher level. On this interpretation, when Valjean determines that he wishes his desire to follow the example of the Bishop to be effective, he thereby identifies with that first-order desire. Then in the deliberation case we see this confirmed as Valjean’s second-order volition when he acts on that desire.\(^{13}\)

However, there are some obvious problems with the account as stated. Perhaps the most obvious problem is the apparent arbitrariness of claiming that it is an individual’s second-order volitions that are determinative.\(^{14}\) It seems as though there are more than just two orders of desire in the hierarchy of desires. We can imagine higher and higher orders of desire limited only by the finite psychological capacities of individuals. For example, again using the example of the Heroin addict, we can imagine a psychotherapist who desires to desire heroin (a second-order desire). But she doesn’t want the desire for heroin to be effective - she doesn’t actually want to end up a heroin addict. Rather she desires to be an unwilling addict in order that she can understand what it is like for the people she treats to fight an addiction. The psychotherapist in this case would possess a third-order volition; a desire that she possess the second-order volition of the unwilling addict. Isn’t it more plausible to say that this third-order desire is the one that the individual identifies with?

Frankfurt anticipates this problem and amends his account by adding a requirement of decisiveness; it is by being decisive about one’s second-order volitions that the individual can identify with them. By being decisive the individual’s commitment "resounds throughout the potentially endless array of higher orders".\(^{15}\) But what is it to be decisive? Frankfurt expands in his later work on the precise role that decisions play in our agency; for example, Frankfurt claims that “a function of decision is to integrate the person both dynamically and

\(^{13}\) This is not to suggest that the mere fact of acting on a desire is proof that it was the object of a volition. I have already mentioned how the effectiveness of a desire, the existence of a volition, is not a matter of whether or not the agent actually acts on that desire. An individual can end up acting on a desire that that individual had not previously desired be effective. Nevertheless, seeing an agent acting on desire is defeasible evidence that she identifies with that desire.

\(^{14}\) As we shall see, Frankfurt does recognise this as a problem in his (1971) and attempts to rectify it. This objection was made quite forcefully by Watson ((1975), p218).

\(^{15}\) (1971), p16.
But nowhere does Frankfurt give an account of what precisely these decisions are about or the criteria that the individual is required to use in making that decision. In the next three sections of this chapter I will describe Bratman’s account of identification which combines Frankfurt’s appeal to decisiveness with as sophisticated account of what those decisions are supposed to be about.

6.2. Bratman and the Planning Theory of Agency

The problem referred to in the previous section, the arbitrariness of stopping at second-order volitions in attempting to give an account of identification is suggestive of two possible responses. One, develop an account of identification that does not define identification in terms of non-cognitive states. Two, develop a non-cognitivist account that can explain what it is to be decisive vis-à-vis one’s desires. It is the second approach that is pursued by Michael Bratman in his accounts of identification. I will discuss these in sections three and four. Like Frankfurt, Bratman’s views on identification evolved over time and in light of this I consider two distinct but related accounts of identification that he defends. The first defines identification in terms of the individual’s policies, the second in terms of her self-governing policies only.

Frankfurt’s account of identification is inextricable from his general theory of agency. To be an agent is for that individual to possess the capacity to identify with her desires. Although Bratman, like Frankfurt, is interested in both agency and identification, he does not argue that the capacity for agency is the very same thing as the capacity for identification. Bratman makes it clear that the capacity of an agent to identify with her desires is only one aspect of agency. While my eventual goal is to describe Bratman’s accounts of identification and determine whether it can constitute a satisfactory account of conscientious identity we must

---

16 (1982), p175.

17 I mentioned earlier the possibility of cognitivist accounts like Watson's which defines identification in terms of the individual’s beliefs about what is worth doing ((1975), p209).


19 For example, Bratman claims simply that “we have seen reason to include in our model of intentional agency such phenomena as deciding to treat, and of treating, certain of one’s desires as reason-giving” ((1999), p206). He does not claim that an account of identification would exhaust a model of intentional agency.
begin by setting out the various different elements of Bratman’s general theory of agency. This is a necessary first step for two reasons: First, in order to understand Bratman’s accounts of identification we must understand his general theory of agency: His account of identification is built on various technical notions drawn from his general theory of agency. Furthermore, Bratman’s account of identification, like Frankfurt’s, is supposed to be a theoretical notion. It is not intended simply to capture the intuitions of the folk about which mental states they identify with, but also to explain certain features of our agency, such as the role that ‘plans’ play in our practical reasoning and how it is that we can be temporally extended. Second, while I go on to reject accounts of identification as a viable basis for conscientious identity, in the next chapter I go on to defend an account of conscientious identity that depends on the notion of a quasi-policy articulated in Bratman’s account of identification.

Bratman focuses on three core features of human agency.20 First, to be an agent is to be reflective. Second, to be an agent is to make plans. Third, to be an agent is to possess a conception of one’s agency as temporally extended. Let’s consider each feature in turn. Reflectiveness is that feature of agency which was emphasised in the work of Korsgaard and Frankfurt. It refers to the individual’s capacity to stand back from her inclinations and either endorse or reject that inclination. The second feature, planfulness, is a feature of human deliberation that Bratman puts particular emphasis on. Humans are capable of, and often adept at, forming plans which have a rational influence on their thought and actions. Recall that for Frankfurt, desires were the primary mental state involved in identification. Desires also play an important role in Bratman’s account of agency but he suggests that Frankfurt’s account neglects the importance of plans; mental states which also involve motivational states but are not reducible to them.21 A plan moves the individual to act in a certain way but also involves an individual’s commitment to acting on that desire.

We do not simply act from moment to moment. Instead, we settle on complex-and, typically, partial and hierarchically structured-future-directed plans of action, and these play basic roles in support of the organization and coordination of our activities over time. In settling on a

20 (2000), p35. Bratman suggests that these three features may even be distinctive of human agency – but he does not come to a determination on that point.

21 Plans may even necessarily involve desires, depending on one’s own theory of desire, but there is no need to come to any conclusion on that matter.
prior plan of action one commits oneself to the plan—though of course one's commitment is normally not irrevocable, and new information can make it imperative to reconsider and abandon a prior plan.\textsuperscript{22}

In the above passage from Bratman we see some of the core features of plans. First, plans are partial—while they exert rational pressure to carry out a particular action, plans do not \textit{necessitate} a particular action. For example, I can make a plan to write another chapter of my thesis over the next month. This does not necessitate that on any particular occasion over the next month that I write that chapter, but the possession of that plan does exert rational pressure on me to write that chapter at times. In this respect, plans are like (non-effective) desires. Second, plans play a role in organising and co-ordinating activity over time. Once I make the plan to write a chapter of my thesis over the next month I am rationally required to arrange my activities over the next month in terms of this plan. At least some time over the next month I will be required to forgo some other activity in order to write the chapter. Third, sometimes an agent will be permitted or required to abandon that plan. For example, if I subsequently discover that a close family member has a terminal illness I may be required to abandon my plan to work on my thesis and spend time with him or her. The content of a plan can be either specific or general. For example, if I intend submitting a paper to \textit{Australasian Journal of Philosophy} tomorrow then I have a specific plan. Bratman refers to these as schedules. Plans that are not specific, that are “general with respect to their occasions of execution”\textsuperscript{23}, are referred to by Bratman as \textit{policies}. Bratman’s notion of a policy is central to the account of conscientious identity that I go on to defend in chapter eight.

So far I have described two of what Bratman claims are the three features of human agency. What of the third? Bratman argues that in order to be an agent the individual must have a conception of her agency as temporally extended.\textsuperscript{24} This capacity of an individual to have a

\begin{footnotesize}
\begin{enumerate}
  \item \textsuperscript{22} (2000), p40.
  \item \textsuperscript{23} (1989), p444.
  \item \textsuperscript{24} What is the connection between a conception of oneself as temporally extended and a \textit{self-conception} (discussed extensively in chapter eight)? The former can be seen as a particular formulation of the latter. The precise nature of a self-conception depends on the “self” that is being investigated. The account of self-conception that I defend concerns the individual’s conception \textit{vis-à-vis} what makes her life meaningful. The notion defended by Bratman concerns the individual’s conception \textit{vis-à-vis} some aspect of her personal identity; her belief that she is a being that extends over time.
\end{enumerate}
\end{footnotesize}
conception of herself as temporally extended is supposed by Bratman to be conceptually entailed by the fact that an individual possesses plans (and therefore policies),

Why is it that individuals who possess plans must see themselves as temporally extended? Bratman explains it in the following way: When I am in the middle of a project I see myself as the same person who began it. Similarly when I finally complete a project I see myself as the same individual who began and continued on with that project. Consequently, for an individual to possess plans it is not possible for her to see herself as a mere time-slice; she must see herself as temporally extended. The capacity of an individual to see herself in this way is not only a ‘central’ feature of human agency but a distinct human feature as other animals are incapable of seeing their projects as cross-temporal.

Bratman argues that this capacity, indeed this necessity, of individuals to see themselves in this way tells us something important about personal identity. In seeing myself in this way, Bratman argues, I arrive at an important truth about myself. I realise that I am one and the same individual who engages in all these processes. According to Lockean and neo-Lockean accounts of personal identity, it is these overlapping psychological connections that constitute an individual’s personal identity. How exactly do these plans/policies play that important co-ordinating role? Bratman suggests the following:

Such plans and policies have as their function the support of cross-temporal organization and coordination of action in part by inducing cross-temporal connections (for example, between prior plans or policies and later action, and between present intentional action and later, planned activity) and continuities (for example, stable plans and policies). A point of having plans and policies is to induce organization and coordination by way of such continuities and connections.

In that passage, Bratman suggests that it is not only by possessing plans that the individual sees herself in a certain way; possessing plans actually means that an individual has the sorts of mental states that are constitutive of persons.

Before moving on from Bratman’s account of agency to his thoughts on identification specifically it is necessary to mention another important concept introduced by Bratman; the notion of a *self-governing* policy. A policy need not only be a plan for a general pattern of future action – that is a first-order policy. We can also have higher-order policies that involve a general plan that pertains not to the individual’s future actions but to the individual’s future motivations:

One might have, say, a policy of developing and supporting a strong concern with honesty in writing, of trying to be more willing to be playful or less inclined to be impatient with others, of trying not to be so attracted to chocolates or to other temptations, or of never acting on or treating as providing a legitimate consideration in one’s deliberation a desire for revenge or a desire to demean.\(^{28}\)

This type of policy is referred to as a *self-governing* policy. In the next section I will consider Bratman’s argument that these policies play an essential role in an individual’s temporally-extended agency. Before leaving this discussion of identification it is necessary to consider another notion that also plays an important role in the individual’s temporal agency; the *quasi-policy*. An example of a quasi-policy is an ideal. An individual’s ideal may well sustain an individual’s temporally extended agency in a similar fashion to a self-governing policy. However, an individual who fails to act in accordance with her ideal does not thereby act *inconsistently* with that ideal.\(^{29}\) This notion plays an important role in the account of narrative self-conception I defend in chapter seven.

### 6.3. Bratman and Policy-Based Identification

Earlier, I discussed the notion of identification defended by Frankfurt. In this section I begin discussing Bratman’s accounts of identification. I argue that two distinct formulations of identification can be discerned in Bratman’s work. The first formulation defines identification in terms of the decisions that an individual makes about her *policies*. The second formulation defines identification in terms of just a sub-set of the individual’s policies; *self-governing policies*. Given that the individual’s policies include her self-


\(^{29}\) (2007). p43.
governing policies, while my critique of Bratman focuses on the first formulation, my criticisms apply equally to both accounts. Bratman’s first formulation of identification is not supposed to have direct theoretical pay-offs like Frankfurt’s account. It is not an attempt to define personhood or freewill. Bratman’s focus in this initial formulation is to analyse the folk concept of identification. Bratman leaves open the possibility that this account plays an important role in a compatibilist theory of moral responsibility, but that is not what guides his analysis.

I mentioned that the starting point for Bratman's account is that initial problem with Frankfurt’s analysis; it doesn’t seem as though second-order volitions (or perhaps, even, any sort of volition) can explain identification. Frankfurt, recognizing this problem, suggested that perhaps second-order volitions need to be supplemented by a notion of decisiveness which ‘resounds’ up through potentially higher-order desires. In “Identification and Externality” Frankfurt claimed:

> It is characteristically by a decision...that a sequence of desires or preferences of increasingly higher orders is terminated. When the decision is made without reservation, the commitment it entails is decisive.

Here, Frankfurt is claiming that it is the individual’s decision, in combination with her second-order volition that provides the basis for an individual to identify with a desire. But what is this decision supposed to be about? The decision cannot be about which of the individuals desire to act on. For example, we could imagine the unwilling heroin addict deciding that her desire not to take drugs will be effective. Nevertheless, couldn’t that very same addict give in on just one occasion and decide on that occasion that her desire for heroin be effective? Such a situation seems, at the least, a conceptual possibility. And in such a situation we don’t want to be committed to the view that the individual must identify with that desire for heroin that she had on that particular occasion. In light of this possibility, Bratman suggests an addition to Frankfurt’s analysis: identification involves an agent

---

30 The one discussed in “Identification, Decision, and Treating as a Reason” (1999).

31 “My primary concern here will simply be to provide a coherent characterization of the phenomena that are the target of such talk of identification” (1999), p187.

deciding to treat a desire as constituting a *practical reason* for that individual.\(^{33}\) The reason that the unwilling heroin addict who gives in and acts on her desire for heroin does not count as identifying with that desire is (we can assume) that she has not decided to treat that desire as reason-giving. What exactly does Bratman mean by deciding to treat a desire as reason-giving? Bratman suggests the following:

> I treat my desire as reason-giving in the relevant sense when I treat it as end-setting - where to treat it as end-setting is, in part, to treat it as potentially justifying, at least to some extent, my performance of relevant means and/or relevant preliminary steps.\(^{34}\)

So as an unwilling heroin addict, I would treat my desire not to take heroin as end-setting if I set this as the standard by which I will organise my various first-order desires and actions. For example, I might desire to check myself into a rehabilitation clinic in light of this end. At one stage Bratman describes this decision to treat a desire as end-setting as treating that desire as a policy.\(^{35}\) Is deciding to treat a desire as end-setting the same as setting a policy? On might argue that while the above example of end-setting *does* involve a policy, a plan which is general with regard to its occasions of application, the very fact of setting an end does not entail setting a policy. Perhaps one could set a *schedule* as an end. Most schedules are not ends as they are instruments towards achieving one’s policies (e.g. I might set a schedule of how much of my thesis to complete over the next month, not as an end in itself, but as a means towards satisfying my policy of completing my thesis). As little seems to rely on this distinction I will simply assume from here on that to set an end just is to establish a policy.

Bratman recognises several objections that can be made against the account so-described.\(^{36}\) First, just because an individual has made the relevant decision, has set a policy, that does not mean that the individual has a *conclusive* reason for action. I mentioned earlier that it is an essential feature of a plan that the reasons that a plan provides can be defeated by other reasons. Specifically, it seems as though an individual’s decision to treat some desire as end-

---

\(^{33\text{(1999), p192.}}\)

\(^{34\text{(1999), p198. The claim that it is only ‘end-setting’ reasons that are relevant means that instrumental reasons are not identification-constitutive.}}\)

\(^{35\text{(1999), p192.}}\)

\(^{36\text{(1999), p200.}}\)
setting can come into conflict with other desires of the individual that she has decided to treat as end-setting. A heroin addict may decide to treat her desire for heroin as end-setting just one time even though this decision is contrary to a long standing policy of that individual not to indulge in heroin. The individual has simply given in to her addiction and decided to treat it as end-setting, this time. It seems as though the very action of setting an end does not itself exclude the possibility that the individual has other ends which are inconsistent with that end. Working on the assumption that such decisions are possible, Bratman suggests an addition to his analysis of identification; the agent must be *satisfied* with her decision to treat some desire as end-setting. An individual is satisfied with her decision as long as her will is not divided; as long as that decision is not contrary to another decision of the agent not to treat that desire as end-setting. The addict described above does not identify with her desire to take heroin as it is in direct conflict with a long-standing decision of the individual not to take that desire to be end-setting.

Bratman suggests another objection to his account as originally stated. Bratman points out that there is a difference between *deciding* to treat some desire as end-setting and *actually* treating it as end-setting. This difference is important because it seems as though an individual should only count as identifying with one of her desires in the latter case. Bratman suggests that a success condition needs to be added to the analysis: an agent must not only decide to treat a desire as end-setting, the agent must *succeed* in doing so if she is to identify with that desire. An individual identifies with her desire to refrain from injecting heroin where she is satisfied with her decision to treat that desire as end-setting, and she follows through and actually treats that desire as end-setting. This is not to say that the individual must actually *act* on that desire that she has decided to treat as end-setting. In the first place, it does not follow from being satisfied with a decision to treat a desire as a policy, and actually treating it as a policy, that the individual possesses a reason that necessarily defeats all other reasons. Second, even if the desire that the individual identifies with does override or defeat in some way all other normative reasons of the individual, weakness of will is still

37 It might be wondered whether this inconsistency really is conceptually possible. It might be argued that if the initial goal is truly an end, then it rules out the operation of other ends in the individual’s reasoning.

38 Frankfurt emphasises the satisfaction element of identification in his (1987).

39 Bratman adds the caveat that an individual might not actually be faced with a situation that calls for the decision to be applied, and therefore the condition might be more correctly stated as a requirement that the agent succeed in treating that decision as end-setting if faced with a situation that calls for doing so ((1999), p202).
possible. Plausibly, an individual can act in a way that she has all-things-considered normative reason not to.

6.4. Bratman and Self-Governing Policy-Based Identification

I claimed that Bratman offers two different formulations of identification. In the second formulation, Bratman suggests another addition to the earlier account of identification. It is not enough, Bratman suggests, for an individual to make a decision as to her policies, for identification it is required that an individual make a decision about her self-governing policies. I think two arguments can be discerned in Bratman’s work for altering the account. The first is Bratman’s claim that self-governing policies play a special role in our agency that other policies or plans do not:

Self-governing policies are embedded in a planning framework whose organizing roles involve the constitution and support of Lockean continuities and connections characteristic of temporally extended agency. Further, such policies—unlike intentions and plans that concern only particular occasions—are explicitly concerned with the functioning of relevant desires generally in one's temporally extended life.

I mentioned in the previous section the way in which Bratman’s account of agency fits with neo-Lockean theories of personal identity. Bratman distinguishes two sorts of psychological tie. Desires and beliefs are often continuous over time and, therefore, it makes sense to refer to these sorts of psychological tie as ‘continuities’. But plans and policies don’t work in the same way. The intentional execution of an earlier plan is not ‘continuous’ with that plan, though it is still connected to it in some way. In light of this Bratman refers to the latter sort of psychological tie as a ‘connection’ rather than a ‘continuity’.

---

40 To further complicate matters, at one point Bratman suggests that it is not only self-governing policies but also ‘quasi-policies’ that are identification-constitutive. A point made in (2007) footnote 37, p83.

41 Bratman does not explicitly acknowledge the divergence in the two formulations.

Why is it these policies that play a “central role” in the constitution of the individual over time? While it is true that only self-governing policies concern the individual’s mental states rather than actions and therefore only self-governing policies have psychological ties as their objects, it does not follow that only self-governing policies that have a special role in constituting the individual over time. It seems as though all plans (whether policies or schedules) can sustain connections and continuities. It might be the case that, in general, self-governing policies have a more substantial presence in an individual’s agency than first-order policies. For example, if I form a self-governing policy to be honest, then this commitment to honesty will affect my practical deliberation regularly because honesty will be relevant to many instances of practical deliberation. On the other hand, my policy to complete a chapter of my thesis over the next month will affect far fewer instances of my future practical deliberation. And the relevance of this is that, as I explained earlier, self-governing policies are, in some sense, constitutive of an individual’s personal identity. They (as well as other policies) provide the temporal connections required by psychological accounts of personal identity. However, it seems to me as though all plans can have this effect. While it may be that higher-order policies are often more stable than ordinary policies this is a contingent feature of those policies. There is no reason to rule out, by definition, ordinary policies from being identification-constitutive.

In summary, Bratman offers an account of agency which is more sophisticated than Frankfurt’s; a place is reserved for plans and the special role that they provide in an individual’s practical psychology. I described two different formulations of identification offered by Bratman. The first formulation concerned the individual’s decision to adopt certain policies. The second formulation concerned the individual’s adoption of self-governing policies. I suggested that one possible reason for this change is Bratman’s thought that it is self-governing policies especially that sustain the psychological ties which are constitutive of personal identity. I suggested though that ordinary policies seem equally capable of sustaining those ties.

In the next section I shall consider whether Frankfurt’s volition-based account or Bratman’s policy-based approach satisfy the four requirements of conscientious identity. In chapter seven I defend an account of conscientious identity that employs Bratman’s account of policy.

---

43 (2002), p82.
6.5. Identification and the Four Requirements of Conscientious Identity

A viable account of conscientious identity must satisfy the four requirements: it must explain the distinct normative import of conscientious convictions, how those convictions are first-personal, their connection to practical deliberation and their enduring nature. In this section I argue that neither Frankfurt’s nor Bratman’s account satisfy all these requirements. Frankfurt’s account satisfies the practicality requirement. Bratman’s account can satisfy this as well as the endurance requirement. Neither account can satisfy the first-personal requirement.

How exactly do both accounts satisfy the practicality requirement? Recall that in Frankfurt’s account the capacity to identify with one or other of the individual's desires is a pre-requisite to having free will. One way in which such an account can be connected to practical reasons is via the well-known Kantian argument that possessing free-will is a pre-requisite to being bound by normative reasons. If that Kantian argument is sound then an individual’s being able to identify with some of her desires, possessing a second-order volition, would be a pre-requisite to engaging in practical deliberation. Frankfurt’s own view of the connection between identification and practical reasons is that the capacity to reflect on one’s desires and decide to act on one or some of them rather than others just is the ability to rationally deliberate. In “Identification and Wholeheartedness” Frankfurt argues that rationality is not a different capacity from the will. In order to reason an agent must be able to reflect on one’s desires at a higher level (and this is why a wanton lacks the capacity to reason).

What is the connection between identification and practical reasons for Bratman? Bratman’s first formulation required deciding to treat some desire as end-setting, and actually treating that desire as end-setting. To treat a desire as end-setting is to have a certain policy. And to have a certain policy is to possess a type of practical reason. According to the second formulation of identification offered by Bratman self-governing policies provide reasons for the individual by providing reasons to have certain motivations (say, ‘honest’ motivations) in

---

44 (2012), 5:95.
45 (1982), p146.
the future. So this formulation also implicates the individual’s practical reasons, though a smaller sub-class of those reasons than the first formulation.

We can describe this connection between identification and practical reasons in terms of the *Les Misérables* cases: In the first case Valjean reflects on his desires and determines that the virtuous ways of the Bishop will constitute his second-order volition. In the second case Valjean acts on that second-order volition. According to Bratman’s account, the first case involves Valjean making a decision about a self-governing policy to be virtuous like the Bishop. In the deliberation case this policy is then employed in his practical deliberation so that he can work out what he ought to do. In the deliberation case Valjean is actually taking that policy to determine his all-things-considered reason.

Does either account of identification satisfy the endurance requirement? Frankfurt’s account cannot do so. An individual can possess a second-order volition for a mere moment. While as a contingent matter, it might be thought that once the individual has arrived at that volition by reflection she now has a reason to act that would affect a range of future actions, volitions do not do so as a matter of conceptual necessity. What about Bratman’s account of identification, does that involve reasons which endure? Recall that a policy, as a type of plan, is something that supports and co-ordinates an individual’s activities over time: “Prior plans have, in this sense, a certain stability: there is, normally, rational pressure not to reconsider and/or abandon a prior plan”. Plans necessarily endure. That is not to say that they always act as conclusive reasons for the individual or that they aren’t subject to being revised or abandoned when circumstances call for it. Given that in Bratman’s second formulation of identification it is only self-governing policies that are identification-constitutive there is an even stronger case to be made that the policies endure. Recall that self-governing policies and quasi-policies are supposed to be peculiarly important for maintaining the cross-temporal Lockean connections which are necessary for personal identity. These policies seem even less likely to be amended or revised than an individual’s ordinary, first-order, policies. While it is possible for an individual to decide on a policy and then immediately abandon that policy, it is of the nature of a policy that it *purports* to endure.

Does either account satisfy the first-personal requirement? In the case of Frankfurt’s account, this is the question of whether the possession of a second-order volition entails the agent *believing* that she possesses that volition. It seems not. The unwilling heroin addict does not necessarily believe that she desires her first-order desire not to take heroin be her effective
What about Bratman’s account? Does it follow from deciding to adopt a policy that an individual has particular beliefs about herself? For the same reason discussed in the case of Frankfurt’s account, it does not follow from the adoption of a policy that an individual believes herself to have adopted that policy. So it does not follow from the mere fact of identification that an individual possesses any beliefs about her identification.

I mentioned in the previous section that, according to Bratman, it is the possession of plans that permits an agent to have a conception of herself as temporally-extended. The cross-temporal organisation that plans facilitate are what allow for an individual to have a conception of herself as an agent, as something more than a time-slice.

Is this sufficient for Bratman’s account of identification to satisfy the first-personal requirement? I don’t think it is. Even if identification is what permits the individual to see herself as a temporally extended agent, identification does not itself involve the agent having any particular belief about the sort of person that she is (the sort of belief that is pertinent for the first-personal requirement).

Finally, do either Frankfurt or Bratman’s accounts satisfy the distinctness requirement? Recall that distinctness requires that an account of conscientious identity be able to capture the distinction between those normative considerations that implicate the agent's conscientious identity and those that do not. Frankfurt’s account held that practical reasons were constituted by the desires that are a consequence of higher-order reflection. This means that Frankfurt’s account is plagued by the same problem as the accounts of Korsgaard and Taylor. I referred to Frankfurt’s claim that the capacity to identify is a pre-requisite to any practical deliberation. This means it can’t account for the distinctness requirement. But this doesn’t seem to be a necessary element of his account of identification. Perhaps the possession of second-order volitions is only a pre-requisite to a possessing a certain sort of

---

46 Though there is still the possibility that possession of that volition entails that if the agent were sufficiently reflective she would believe herself to possess that volition. Just as in Korsgaard’s account, the fact that an agent endorses a desire entails that she endorses a particular conception of herself, as on reflection, she would possess such a conception of herself.
normative reason and there are considerations that can be genuinely normative while not being the object of such desires.\footnote{Frankfurt considers this possibility in the case of instrumental reasons before ultimately rejecting it ((1971) p174).}

I argue that even if Frankfurt's account can allow for a distinction between conscientious and non-conscientious considerations, it is clear that the correct distinction, for our purposes, cannot be made. The distinction between practical considerations that are the agent's 'own' and those considerations that she is alienated from is not the distinction between the conscientious and the non-conscientious. Recall that we need an account that will explain why it is that, in the \textit{Les Misérables} cases, only some of the considerations Valjean appeals to implicate his conscience. It is true that Valjean does initially make reference to his brutish inclinations as not being 'his own' - which might suggest he is talking of an issue of identification. Nevertheless, I think there is good reason to think that the distinction being appeal to here is \textit{not} the distinction between identification and alienation. Consider, in particular, the deliberation case. There, Valjean is deliberating over whether he ought to follow his Bishop-inspired ideals or some other practical considerations. He mentions, in particular, the fact that turning himself in would have adverse consequences on the townspeople. This normative consideration, this \textit{pro tanto} reason, not to hand himself in, does not eventually move Valjean to action. But we would not say that that consideration, that reason, is not 'his own'. It is nothing like the unwilling addict's desire for heroin. The distinction between the conscientious and non-conscientious is simply not the distinction between practical considerations/desires that are the agent's own and those that are not.

What about Bratman's account (according to the second formulation), can it satisfy the distinctness requirement? I argue that it can. The first \textit{Les Misérables} case, the transfiguration case, might reasonably be interpreted by Valjean as deciding which of his desires to endorse in his self-governing policy. Plausibly, he determines that, as a self-governing policy or a quasi-policy, he plans to live the virtuous life as inspired by the Bishop and in future will deliberate in terms of that policy. How would such an analysis explain the deliberation case?

Valjean has formed a self-governing policy inspired by the Bishop. But it is of the nature of policies that they do not determine a particular outcome. What this self-governing policy does is then influence his future practical deliberation in a certain way. This policy needs to
then be weighed up against other practical considerations such as the adverse consequences of handing himself in. But as a matter of fact, there is nothing to overwhelm that self-governing policy, so he ends up acting in accordance with it. The distinctness requirement is the requirement that any account of conscientious identity be capable of distinguishing those considerations that Valjean appeals to in his deliberation that implicate conscience and those considerations that do not. As Bratman’s account can make this distinction it satisfies the distinctness requirement.

In this chapter, I began my evaluation of accounts of identity from contemporary moral philosophy in terms of the four requirements of conscientious identity. I began with Harry Frankfurt’s account of identification. Frankfurt analysed identification in terms of an individual’s second-order volitions. I then described Bratman’s account of agency and his two different formulations of identification. I suggested that neither of Bratman’s formulations of identification can satisfy the four requirements of conscientious identity. Both accounts can satisfy the practicality requirement, and Bratman’s account can satisfy the endurance and distinctness requirements. Neither account can satisfy the first-person requirement.
Chapter Seven: Conscientious Identity as Narrative Self-Conception

7.1. Neo-Aristotelian Character

In chapter six I considered the possibility that conscientious identity could be analysed in terms of identification. I argued that conscientious identity could not be analysed in that way as identification does not satisfy the four descriptive requirements of conscientious identity. In this chapter, the last in Part III and the last of the thesis, I consider some other contender accounts of identity from moral philosophy. I consider neo-Aristotelian character, Bernard Williams’ ‘deep’ character and accounts of narrative self-conception. I claim that the first two accounts also fail to satisfy the four requirements. But I develop an account based on narrative self-conception which I argue can satisfy the four requirements. I argue that a conscientious conviction is one that captures the individual’s beliefs about what would make her life meaningful (the self-conception part) where those beliefs also cohere with her life as a whole (the narrative part).

In chapters five and six I suggested the starting point for our analysis of conscientious identity should be Marya Schechtman’s characterisation question. We want to know, which features of an individual are the ones that make her ‘who she is’? The fact that this is a question of ‘characterisation’ means that we would be remiss if we were not to consider whether accounts of character can satisfy the four requirements. In recent decades, by far the most prominent analyses of character have been virtue theory accounts of neo-Aristotelian character.¹ These accounts of character concern an agent's dispositions to act in certain ways (to possess certain character traits). Often the mix of traits is taken to be particularly distinctive of that individual in question; hence why we often describe individuals who act contrary to the ways that we expect as acting ‘out of character’. An individual of good character, on these accounts, is one that possesses the appropriate balance of virtuous and

¹I refer to these as 'neo-aristotelian' accounts in order to distinguish them from Williams’ account of character which I discuss in the section that follows. These accounts are neo-aristotelian because they are based on Aristotle’s discussion of character in Nicomachean Ethics (Book II, in particular). I call Williams’ account of character ‘deep’ character as it isolates a particularly ‘deep’ subset of our mental states and claims that it is this sub-set that constitutes our character.
vicious dispositions. As a representative of this approach, I consider a recent account of character described by Christian Miller. According to Miller, an agent’s character consists in an agent’s "character traits and the relationships between them" and a character trait, in turn, is defined as "a disposition to form beliefs and/or desires of a certain sort and (in many cases) to act in a certain way (when in conditions relevant to that disposition)".²

In order to assess such an account for its consistency with the four requirement each element needs to be explained in reasonably precise detail. Why, we might ask, is a character trait a disposition? Consider, for example, the character trait of shyness. We would not judge it sufficient for an agent to possess this character trait that she acted shy on just one occasion: a shy person is not someone who, on one particular occasion, finds it difficult to make small talk to a stranger. We expect a shy person to regularly engage in shy behaviour or possess shyness-relevant mental states. Not only is the possession of an occurrent mental state insufficient for possessing a character trait, it isn’t necessary either. Recall that according to Bratman’s account of identification, it was not required that the agent actually treat a desire as end-setting for that desire to be one that she identifies with. What was required, rather, was that the agent would treat that desire as end-setting if she found herself in the relevant circumstances. A similar counterfactual is necessary in articulating character traits. We could imagine a shy person who never encountered a situation where her shyness actually came into play (perhaps she is the only person left on earth). Consequently, we say that being shy is not a matter of actually thinking or acting in any particular way, but being disposed to do so when in the relevant circumstances.

I mentioned that being shy seems to involve not just being disposed to act in certain ways, but being disposed to have certain thoughts as well. Why do we include the disposition to think as well as the disposition to act? Consider the shy person who finds it difficult to make small-talk with strangers: She might manage to work through that difficulty and make small talk. That agent can still count as shy, perhaps, because she is disposed to have a desire not to make that small talk, even if she can successfully overcome that desire. So plausibly a character trait is constituted by a mix of dispositions to think and act in certain ways with the precise mix being dependent on the particular trait in question. It is worth noting from this example that there is no requirement that the character traits be in any sense ‘moral’. Being shy or logical is not, prima facie, a moral attribute of a person.

Returning to Miller's initial description of character, it is not simply the set of traits that constitutes an individual's character but also the way in which those traits are woven together. For example, an individual could possess the traits of being shy, being logical and being anxious in one particular mix (they are shy, logical and anxious on certain occasions) whereas another individual could have those same traits in a completely different mix (they could be shy, logical and anxious on a completely different set of occasions).

Is neo-Aristotelian character a plausible account of conscientious identity? Could we see the transfiguration event as a point in which a prior character is discarded by Valjean and replaced with a new character? And could we interpret the deliberation event as one where a putative or half-formed character is confirmed and re-inforced in Valjean's actions? Consider Valjean's claim in the transfiguration case that he would 'no longer be the animal'; this could be interpreted as a point at which Valjean makes a decisive break with his previous traits and decides to adopt new traits, a new character.

Following the method used when evaluating accounts of identification, neo-Aristotelian character must be assessed in terms of the four requirements. In the first place, neo-Aristotelian character seems to satisfy the practicality requirement. An individual’s character is constituted by the ways in which she is disposed to think and act. Some character traits are directly practical, such as the trait of honesty; to be honest is to be disposed not only to act in a certain way but to possess a capacity to think about how one ought to act, in a certain way. The fact that honesty is one of an individual’s fundamental character traits will mean that being honest weighs heavily in her practical deliberation. Possessing a certain character trait does not mean that an agent will always think or act in accordance with reference to that character trait: after all, sometimes people think and act 'out of character'. Nevertheless there is still a direct conceptual connection between having a certain character trait and how one acts and practically deliberates. There can also be a connection between non-practical character traits (such as being logical) and practical deliberation, in the sense that being a logical person will affect the way in which one reasons about practical issues. Just as in the previous chapter I mentioned how the possession of a plan, policy or quasi-policy does not entail a particular way of thinking or acting, the same applies in the case of neo-Aristotelian character. It is not essential that an individual always act in a certain way for that way of acting to be a character trait.
Does neo-Aristotelian character satisfy the endurance requirement? It seems natural to think that if an individual is disposed to do something that that disposition will have some lasting significance. And certainly, as an epistemic matter, it would be hard to know whether someone possesses a character trait if it wasn’t an enduring feature of that person.

Nevertheless, the generally lasting nature of a character trait, just like the generally lasting nature of a second-order volition on Frankfurt’s account seems to be a contingent matter. Since it is possible that an individual could possess a character trait for a mere moment, traits do not endure. What about the fact that, for the neo-Aristotelian account, character isn’t determined simply by the set of character traits that the individual possesses but the fact that those traits are *configured in a particular way*? Can an individual possess a configuration of character traits for a mere moment? Consider the individual who was shy when it came to parties, but is bold when it comes to conferences. It sounds somewhat tortured to say that such an individual could possess that peculiar (and possibly unique) mix of traits for a mere moment. Nevertheless, that is a conceptual possibility. Therefore, neo-Aristotelian character does not satisfy the endurance requirement.

Neo-Aristotelian character fails to satisfy the first-personal requirement just as accounts of identification failed. It does not follow from the fact that an individual possess a certain character that that agent see herself as an individual possessing that particular character. In fact, I think, this disconnect is quite common. For example, a particularly modest person might not believe herself to be as generous as her character reveals. Or, a particularly immodest person might believe themselves to be far more generous than seems to be evidenced by their character.

In contrast to identification, neo-Aristotelian character is incapable of satisfying the distinctness requirement. How would an account of neo-Aristotelian character attempt to explain Valjean’s thought in the *Les Misérables* cases? Such an account would hold that, in the transfiguration case, Valjean is deciding what his character shall be. Then, in the deliberation case, he is considering what he ought to do in light of that character. Is such an interpretation of Valjean’s thought plausible? It doesn’t seem to capture (very well, at least) how the non-conscientious considerations that Valjean appeals to in the deliberation case can count as normative. If Valjean had decided that he ought not to turn himself in order to protect the well-being of the townspeople, would we say that he had acted contrary to his true character? I don’t think that we would. If we are to interpret the deliberation event as involving deliberation about character it is more plausible to interpret it as his deliberation
(still) about what his character truly is, rather than deliberation about how character should be considered alongside other considerations.

Neo-Aristotelian character does not seem to be a good candidate for conscientious identity. While it can satisfy the practicality requirement, it does not satisfy the endurance, first-personal and distinctness requirements. Nevertheless, we shall see in the account of conscientious identity that I defend; conscientious identity as narrative self-conception, something like neo-Aristotelian character plays an important part. The coherence of an individual’s character over time is the way in which narrative self-conception satisfies the endurance requirement.

7.2. Bernard Williams and ‘Deep’ Character

The question of what conscientious identity consists in has been framed around Schechtman’s characterisation question. I suggested that the characterisation question is not really one question, but a series of different questions. So far I have considered the possibility that it is a question about identification or neo-Aristotelian character. I have suggested that the latter account of character does not explain what is going on in conscientious identity. In this section I consider another prominent account of character that is substantially different than the neo-Aristotelian account. I now turn to consider Bernard Williams’s account as defended primarily in *Utilitarianism: For and Against* and in the essay “Persons, character and morality” in *Moral Luck: Philosophical Papers 1973-1980*. In this section I will describe and evaluate Williams' account of character against the four requirements. As Williams’ account of character incorporates only the projects of the agent that are, in some sense, ‘deep’, I refer to his account as ‘deep character’ in order to distinguish it from ‘neo-Aristotelian character’. I argue that like Bratman’s account of policy-based identification, Williams’ account of deep character satisfies the distinctness requirement. Whereas policy-based identification satisfied the distinctness requirement by formal features of the practical

---

3 Even though, in other ways, Williams advances a neo-aristotelian approach to ethics (e.g. (2013)).


reasons involved (conscientious convictions involved reasons which had a special organisational relationship to other practical reasons), on Williams’ account the distinctness of conscientious considerations comes from the fact that they are of a distinct normative type; they pertain to what that individual judges would make her life meaningful. In this respect conscientious considerations can be distinguished from other practical reasons of the individual which purport to involve other domains of normativity such as self-interest or morality. I argue that, like accounts of identification and neo-Aristotelian character, Williams’ account is also unable to satisfy the first-personal requirement.

The foundational element of Williams’ account of character is a ‘project’. In the preface to *Morality: An Introduction to Ethics*, Williams claims that:

> the considerations of the moral kind make sense only if they are related to other reasons for action that human beings use, and generally, to their desires, needs and projects.

Here, Williams establishes his view that the normativity of morality must be grounded in what agents care about. In this respect his account differs from the most common approach to the normativity of practical reasons which holds that normativity must be something ‘categorical’ or external to the individual’s interests. Williams does not explicitly define the term ‘project’, but it is clear from his discussion that a project refers to what we might generally call the personal pursuits or commitments of the individual (though it is worth noting that for Williams the notions of ‘commitment’ and ‘project’ are distinct). An individual will generally have several projects going on at any one time: One of her projects might be to provide for her family, another project might be to be a successful musician, another might be too contribute to the welfare of the worst off. In *Utilitarianism*, Williams specifies a certain subset of these projects as the individual’s ‘commitments’. These are “those [projects] with which we are most extensively involved and identified”.

---

6 Recall that the distinctness requirement pertains to the way in which both conscientious and non-conscientious considerations can be distinct while both being potentially normative; that is, both sorts of considerations are at least potentially capable of providing the agent with normative reasons for action.

7 (1972), xiii.

8 Perhaps most prominently defended in “Internal and External Reasons” in *Moral Luck* (1982).

9 Commentators who also interpret Williams’ notion of a project in this way include Wolf (1997) and Buss (2006).

What does Williams mean by ‘identifying’ with a project? Does it bear any relation to the notion of identification employed by Frankfurt and Bratman? Just as with 'projects', Williams gives no explicit definition of identification but the several different uses of the term at various points gives a good indication of what the notion refers to. At one stage, Williams claims that identification involves those "projects and attitudes that in some cases, he [the agent] takes seriously at the deepest level and are what his life is about". At another point Williams says that identification involves "a project or attitude round which he has built his life".

The purpose of Williams' appeal to 'projects' or 'commitments' in *Utilitarianism* is to motivate the idea that utilitarianism, as a moral theory, demands too much of the agent. It demands that the agent treat his projects as of the same worth as the projects of other agents, and in the case of commitments, at least, this is too much to ask of the agent. As is clear from the nature of a commitment, it would be psychologically distressing, and perhaps even psychologically impossible, for the agent to ignore such commitments. The integrity of the agent requires that she live up to her own commitments.

In “Persons, Character and Morality”, Williams elaborates on the notions of project and identification introduced in *Utilitarianism*. In this paper, Williams generalises on the earlier point that utilitarianism has insufficient regard for the personal pursuits of the agent and claims that this is true of all 'impartial' moral theories (for example, it applies equally in the case of Kantianism). Williams introduces the term ‘ground project’. This is described as the project or set of projects which are “closely related to the individual’s existence and to a significant degree gives meaning to his or her life”.

At another point, Williams describes these projects as the sort of pursuits that, if frustrated, the agent would judge that she 'might have well as died'. The descriptions Williams uses employing the term 'ground project' suggest that the notion is broadly equivalent to what Williams referred to as a 'commitment'.

---

in *Utilitarianism*. Williams introduces the specific term ‘character’ to apply to the unity of the individual's ground projects/commitments. To possess a character is for one to possess “projects and categorical desires with which one is identified”.

One implication of the fact that individuals possess ground projects or commitments, Williams contends, is that it explains why it is that impartial theories of morality, at least sometimes, give the wrong answers to moral questions. First, it would require that the agent, in order to achieve some supposedly moral purpose, give up what might be a condition of her having any interest in being in the world at all (such as when the project involves a categorical desire). Second, it would mean giving the wrong sort of reasons as to why the agent should act as she ought to. For example, the reason that a husband ought to save his drowning wife is because of his commitment to his wife. An impartial theory of morality would hold that the reason why a husband ought to save his wife is because some abstract moral rule (such as the maximisation of utility, or a condition of the social contract) holds that the action is morally required. But that explanation relies on ‘one thought too many’. It is the husband’s love for his wife, his commitment to her, which means that he ought to save her.

I introduced Williams’ account as an account of character that is distinct in type from the neo-Aristotelian account. How exactly do the accounts differ? There are two crucial differences. One, Williams’ account concerns the agent’s projects rather than her dispositions; the notions are quite different. Possessing a character trait means that the agent is disposed to do or think certain things. I mentioned Miller’s example of a shy person. A shy person is disposed to do certain things, such as avoid large gatherings where she is surrounded by strangers. A shy person is also disposed to have certain desires. For example, a shy person might always force themselves to attend such functions, but nevertheless possess a desire not to do so. It might involve affective states; such as being disposed to be anxious at those functions. It might even...

16 (1981), p111. A categorical desire is a desire that the agent has that is not conditional on her being alive ((1993), p77). Categorical desires are closely related to our ground projects. Often ground projects will be ends that the agent desires to achieve, unconditional on that agent being alive. For example, my interest in eating dinner tonight is conditional on my being alive, but my interest in decreasing the amount of poverty in my community (say) is not. It remains possible, though, that an agent might have a ground project which does not result in/ is not constituted by a categorical desire.

involve beliefs, such as the belief that one will feel anxious at those functions. But the possession of any of these dispositions does not mean that an individual has shyness as one of her projects. Remember that possessing a project requires that the individual actually cares about that activity. This implies some level of endorsement of an inclination, rather than the mere possession of an inclination.\(^{18}\)

The second crucial distinction between the two notions of character is that Williams’ is only concerned with that sub-set of projects which is, in some sense, deep. We have seen this idea of depth articulated in various different ways by Williams: Character involves only those projects that the agent takes most seriously, give meaning to an individual’s life, that she builds her life around, that relate closely to her existence, that if frustrated she would judge she might as well have died. What are we to make of all these five different descriptions of ‘ground project’? I think that we can best interpret these descriptions as different ways of articulating one general criterion for ground projects. My proposal is that ground projects are those projects that make the agent’s life meaningful (putting to one side, for the moment, what exactly it is for a life to be meaningful). This notion of a project that contributes to a meaningful life is then to be defined subjectively in terms of what the individual cares most about, takes seriously at the deepest level. They relate to her existence and are part of the organisational structure of her mind because of the role that the cares of an agent play in her psychology. Williams’ suggestion that such projects, if frustrated, would have the consequence that the agent would judge she might as well have died, is hyperbolic. It refers to the way in which a life will generally feel worthless to the agent if she doesn’t achieve anything meaningful, rather than referring to a literal criterion for deep character.

What does the notion of meaningfulness mean here? According to Williams, meaningfulness is defined subjectively in terms of what the individual cares most deeply about. Richard Taylor and Harry Frankfurt have defended similar accounts.\(^{19}\) But we could imagine an account that was not subjective, for example Susan Wolf’s account requires that, in order for any project to count as meaningful, the project must be objectively worthwhile.\(^{20}\) For now, I

---

18 Williams’ claim earlier about the internal nature of practical reasons is just the claim that all reasons must be grounded in the individual’s subjective motivational set; it is not the claim all subjective motivation constitutes a practical reason.


will assume that, for the purposes of an account of conscience, meaningfulness might be defined either subjectively, objectively or according to some hybrid account.\(^{21}\) It is also worth mentioning here that the only projects that are relevant for an individual’s deep character are not just the meaningful projects but the ones that make her life meaningful.\(^{22}\)

Is Williams' notion of deep character, interpreted in the way I have suggested, a viable account of conscientious identity? Again, let’s consider this account against the requirements of conscientious identity: Deep character appears to capture both the practicality and endurance requirements. An individual’s ground projects figure prominently in her practical deliberation. That is because consideration of what makes an individual’s life meaningful figures heavily in an individual’s practical deliberation. For example, an artist might reject a utilitarian solution to a practical problem she faces (what she should pursue as a career) on the basis that it is contrary to her ground project.\(^{23}\)

Does deep character capture the endurance requirement? Initially one might think that it doesn’t in the same way that second-order volitions failed to satisfy this requirement. While it is probably true that, as a contingent matter, an agent’s ground projects endure, it is possible for an individual to possess those projects for a mere moment. However, remember that the projects which constitute deep character are not merely meaningful projects but ones that make the individual’s \textit{life} meaningful. And consider that one of the ways in which Williams describes a ground project is as a project that the individual has ‘built his life around’. Perhaps a project that does not endure cannot possibly be part of what makes a life meaningful or be a project that an individual’s life is built around.

Is deep character consistent with the distinctness requirement? It seems that if we think of conscientious considerations as those that implicate an agent's deep character and non-conscientious considerations as those that \textit{do not} implicate an agent’s deep character, we have an account that fits rather nicely with the distinction drawn in Valjean's thoughts in the \textit{Les}

\(^{21}\) Wolf's account is a hybrid account combining the objective worthwhileness of an activity with the subjective attraction to that activity as necessary conditions of meaningfulness ((1982), (2010)).

\(^{22}\) Potentially, one could engage in a meaningful activity without that activity contributing to the overall meaningfulness of her life. For a discussion of this issue see Kaupinnen (2011).

\(^{23}\) Williams himself argues that Gauguin was justified in pursuing the life he did despite the hardships that it might have caused others (1981). For now I make no assumptions as to whether acting on such a project would be normatively or morally justified. My claim here is simply that if we possess a deep character then this affects our practical deliberation.
Misérables cases. As Williams’ account would have it, when Valjean is appealing to conscientious considerations he is thinking about those considerations that give meaning to his life. In the transfiguration event, Valjean is pondering over what it is that it is that would make his life particularly meaningful, what would make it worth living. Whereas in the deliberation case Valjean is weighing those prior-formed judgements about what is meaningful against other normative considerations. It also offers an explanation as to why, incidentally, conscientious considerations might trump other normative considerations: Practical considerations involving meaningfulness are usually thought to be of particularly weighty normative significance. Recall that the accounts of practical identity offered by Korsgaard and Taylor discussed in the previous chapter could not explain how only some of Valjean’s practical considerations implicated conscientious identity and some of those considerations did not? Because, on Williams account, considerations that implicate meaningfulness are only a subset of our practical considerations, it is perfectly consistent with Williams’ account for other practical considerations to be genuinely normative. Non-conscientious considerations, on Williams’ account, are simply those considerations which capture the agent’s ‘ordinary’ projects or interests. For example, in the deliberation case, the adverse consequences on the townsfolk can (on Williams’ account) still constitute a normative reason for Valjean, as he has an interest in the welfare of the townsfolk and, plausibly, helping them is one of his projects. What happens in the deliberation event is that that very reason is defeated by the conscientious consideration (and reason) that counts for turning himself in.

Finally, does Williams’ account of deep character satisfy the first-personal requirement? Recall that initially I phrased this requirement as pertaining to the agent’s beliefs about what ‘sort of person she is’, and then suggested that this can be interpreted through the lens of each of the different accounts of identity. For example, in the case of identification this can be interpreted as the agent’s beliefs about which desires are ‘her own’. In the case of neo-Aristotelian character, this was interpreted as the agent’s beliefs about her dispositions and the relationships between them. What would the first-personal beliefs pertain to in the case of deep character? They would have to be beliefs about whether or not the projects in question

---

On Williams account of normativity (which I find plausible) considerations of meaningfulness are one sort of normative practical consideration to be considered alongside other normative practical considerations. Wolf also defends a similar account of the connection between considerations of meaningfulness and other normative considerations (2011).
give her life meaning. Deep character doesn’t seem to require that the agent is consciously aware that she is a certain sort of person. An agent can have as, one of her ground projects, the well-being of her husband, and save him when drowning on that basis, without (as she is swimming out to him) having the conscious thought ‘this is the sort of person I am’. But recall that the first-personal requirement did not mandate that the agent have any particular occurrent mental states; it required that the agent would have such a mental state, if she was in the appropriate situation.

The question is, is deep character committed to that counterfactual? It may well be committed to certain related counterfactuals. For example, when describing the criterion for a ground project I talked of the counterfactual that one would judge a project to be what ‘makes life worth living’. This was supposed to be another way of stating that some project or set of projects are what make the individual’s life meaningful. But it does not seem to follow from the mere possession of ground projects (which are determined by one’s subjective motivational states) that the individual has any particular beliefs about those projects, i.e. the belief that they are what make her life meaningful.

In summary, Williams’ account of deep character can explain the practicality and distinctness of conscientious convictions. Our ground projects are heavily involved in our practical deliberation and are of a distinct normative type from some of other normative considerations. It may be that an account of deep character can also satisfy the endurance requirement, depending on whether ground projects are necessarily defined in terms of the individual’s life as a whole. However, in common with the other accounts of identity that I have considered, this account is completely unable to satisfy the first-personal requirement. It is an attempt to understand this requirement and how it can be incorporated into an account of identity that is the subject of the next section.

7.3. The First-Personal Perspective and Self-Conception

All the accounts of identity that I have evaluated thus far have had difficulty in satisfying the first-personal requirement. This suggests, perhaps, a different starting point. Remember my initial gloss on the transfiguration event is that it pertained to the way in which Valjean saw himself. This suggests that, perhaps, instead of beginning with established accounts of identity from moral philosophy and then assessing them against the four requirements one by one, we could start with accounts of self-conception which, by definition, satisfy the first-
personal requirement and then check whether such an account can satisfy the other three requirements. In this section I suggest that we start with an account of self-conception. Then, in order for this account to satisfy the other three requirements of identity I introduce a narrative element. This allows for me to describe an account of conscientious identity that draws on insights from my discussion of identification, neo-Aristotelian character and deep character.

The starting point in considering accounts of self-conception is to recognise that an individual’s identity and her conception or beliefs about her identity can come apart. Identity and self-conception can come apart in two different ways. An agent can believe herself to be something that is inconsistent with her identity or an agent could possess an identity that she does not believe herself to have. For example, consider traditional accounts of personal identity or what I called ‘human persistence’. An individual might believe herself to have biological or psychological continuity with some later time-slice, but that individual might not actually possess the relevant continuity. For example, we can imagine a delusional person who believed himself to be the King of France even thought that could not possibly be true. Likewise we could imagine Louis XVI in a case of amnesia not believing himself to be the King of France when in fact he was. I mentioned some of the ways in which self and self-conception can come apart when considering the ways in which the prior accounts (identification, neo-Aristotelian character and deep character) failed to satisfy the first-personal requirement. For example, in the case of neo-Aristotelian character we could imagine a modest individual who does not see herself as a generous person, but her actions reveal otherwise. Likewise we could imagine a particularly mean individual, who sees herself as generous, but her actions reveal that she is not. Or, if we are referring to Williams’ notion of deep character, an agent might believe that her commitment to philanthropy is one of her ground projects, but the fact that she never performs any philanthropic acts and the fact that she doesn’t seem to appeal to such a notion in her deliberation would be evidence that that is not in fact one of her ground projects.\(^\text{25}\)

\(^{25}\) It is not clear that this disconnect applies for all of the accounts of identity I have considered. For example, consider Korsgaard’s account of practical identity: It seems that an individual could have a conception of her practical identity (a description under which she values herself) without believing that she values herself under that description. For example, ‘being a father’ could be a description under which an individual values himself, yet he may not believe that to be the case as he is not sufficiently reflective. But is the reverse case possible? Is it possible for an individual to believe herself to assent to a description under which she values herself, while
So far I have suggested that self-conception pertains to whatever way the agent sees herself. But just as there are many different philosophical accounts of identity there can be many different types of self-conception depending on the type of ‘self’ the conception is directed at. For example, a self-conception vis-à-vis identification is constituted by the beliefs that an agent has about the desires that are ‘truly hers’. A self-conception vis-à-vis neo-Aristotelian character is constituted by the agent’s beliefs about her dispositions and the connections between her dispositions. So now that I am beginning with an account of self-conception, what notion of self, should the conception we are investigating pertain to? The success of deep character in capturing several of the requirements of conscientious identity means that this is a good starting point for accounts of self-conception. That is, rather than focusing on the individual’s beliefs about, say, human persistence, we are interested in her beliefs about what makes her life meaningful. Marina Oshana defends such an account of self-conception in *The Importance of How We See Ourselves: Self-Identity and Responsible Agency.*

A person’s self-conception consists in a set of beliefs about himself, where these are beliefs that guide practical deliberation and enable a person to make sense of his life. These can be beliefs about one’s cognitive states, one’s emotional disposition, one’s character, one’s body, one’s life plans, one’s commitments, values, social roles, and so on.

Let’s consider each element of this analysis in turn. The first important element of Oshana's analysis is the claim that self-conception necessarily implicates an individual’s beliefs. It is this feature which means that we are actually dealing with an account of self-conception at all and means that we are dealing with an account which, by definition, satisfies the first-personal requirement of conscientious identity. Later, however, I will suggest that this leads to a ‘subjectivity problem’ for such an account of conscientious identity.

The second element of Oshana’s account worth drawing attention to is the objects of the relevant beliefs. They could be beliefs about the individual’s own cognitive states. For example, an individual’s belief that she is a theist would often guide her deliberation and help

---

26 (2010).

27 (2010), p35.
her make sense of her life. The belief could be about her emotional disposition. So an agent’s belief that she is a calm person could be a belief that guides her practical deliberation and helps her make sense of the world. It could be about her character (in the neo-Aristotelian sense); that is, it could be her beliefs about her various behaviour dispositions, dispositions to think and act. The beliefs could be about her body; perhaps an agent with body integrity identity disorder cannot make sense of having a particular limb. That individual’s belief about her unwelcome limb would be constitutive of her identity. Then there are the agent’s values. Depending on one’s definition of ‘value’ these could be many different things but they seem closely related to Williams’ notion of a ground project or commitment.

Oshana does not intend this list to be exhaustive; it simply sets out some of the phenomena that might be relevant to an individual’s self-conception. It is worth noting that a far broader range of phenomena are potentially relevant for Oshana’s account of self-conception than for the accounts of identity considered thus far. It is not simply the individual’s beliefs about her second-order volitions (say) or her, beliefs about her ground projects. The beliefs range over a whole range of mental phenomena.

The third important element of Oshana’s account worth pointing out is the criteria for which of those phenomena are relevant. It is those phenomena that help guide practical deliberation and enable her to make sense of her life. I will assume that the notion of phenomena which ‘make sense’ of an individual’s life is the same notion of projects that ‘make that individual’s life meaningful’ in Williams’ account of deep character.

In evaluating each account of identity, I looked at the broader theoretical purposes of the author in giving that account of identity. For example, when assessing Bratman’s account of policy-based identification I didn’t simply consider whether or not that account cohered with folk intuitions about identification, but also the role that that account played in Bratman’s broader theory of agency. It is important to consider the role that the account in question plays in a broader philosophical theory for two reasons. First, it helps us to understand what the value of that notion might be, where it is initially counter-intuitive. Second, the degree to which the theoretical purpose of an account of identity correlates with the theoretical purpose of an account of conscientious identity increases the chance that we are dealing with a plausible account of conscientious identity. So what is the theoretical purpose of Oshana’s account of self-conception? Oshana argues that the possession of a self-conception articulated in the way that she suggests is a necessary pre-condition to an individual being morally
accountable. Oshana argues in support of this by suggesting that the sort of reflection that is involved in having a self-conception is the sort of reflection that is necessary for an individual to be responsive to reasons and to see those reasons as issuing from herself. To be capable of being morally responsible we need to be aware why we do the things that we do and that we do them for our own reasons.28

Does Oshana’s account of self-conception satisfy the four requirements of conscientiousness? By definition, of course, it satisfies the first-personal requirement: a self-conception incorporates an individual’s first-personal perspective because an individual’s self-conception is constituted by her beliefs about the sort of person she is. Oshana’s self-conception also satisfies the practicality requirement. First, it satisfies it by definition, as Oshana has defined practicality as a condition of some belief being constitutive of self-conception. Second, plausibly, those matters that help an individual make sense of her life (matters the agent thinks makes her life meaningful) necessarily constitute practical reasons for the individual.

Does Oshana’s self-conception satisfy the distinctness requirement? On the face of it, it would seem to satisfy the distinctness requirement in a similar way that deep character satisfied the requirement. The criteria for identity-relevance in both Williams’ account of deep character and Oshana’s account of self-conception seem very similar. Both hold that it is the extent to which some matter contributes to the meaningfulness of an individual’s life that makes that matter identity-constitutive. However, there is a potential problem for Oshana’s account, here, that is similar to the problem that beset Korsgaard’s and Taylor’s accounts of practical identity in their capacity to satisfy the distinctness requirement. The problem for those accounts is that they posited that all of an individual’s practical reasons (insofar as those reasons were genuinely normative) were grounded in the individual’s practical identity. Oshana’s claim about the connection between self-conception and practical reasons is similar to Korsgaard’s claim about the connection between a conception of practical identity and practical reasons. Oshana claims that the possession of a self-conception is a necessary pre-requisite to an individual being responsive to reasons. We need to know whether this connection between self-conception and responsiveness to reasons can be severed.

28 (2010), p144. There is a recurrent theme of accounts of identity being used to explicate moral responsibility. This was particularly prominent in Frankfurt’s account of identification discussed in chapter six.
What is Oshana’s argument to the effect that the possession of a self-conception is a pre-requisite to that individual being morally responsible? Oshana claims that moral responsibility requires self-awareness and uses the terms ‘self-awareness’ and ‘self-conception’ interchangeably. It is the latter move that is particularly problematic. Oshana draws on Daniel Dennett’s argument that self-awareness is a pre-requisite to normative responsiveness. Dennett claimed:

Because only if I was aware of the action can I say what I was about, and participate from a privileged position in the question-and-answer game of giving reasons for actions.²⁹

While it may well be true that such self-awareness is a pre-requisite to reason-responsiveness, self-awareness is not the same thing as self-conception. On Oshana’s account self-conception was not simply a matter of an individual being conscious in some way, but having a set of beliefs about which mental phenomena of the individual help enable her to make sense of her life. We need an argument as to why this notion, not simply self-awareness is a pre-requisite to being reasons-responsive.

Could we construct an argument to the effect that reasons-responsiveness requires self-conception? In chapter five I discussed Christine Korsgaard’s argument that the agent’s possession of a conception of her practical identity is a necessary pre-requisite to an individual possessing any reasons whatsoever. What is distinctive of humans, Korsgaard claimed, what gives us our rationality, is our capacity to stand back from our inclinations, to reflect and ask ‘ought I to do that’? And this capacity, Korsgaard claims, necessarily involves a conception of practical identity. While most conceptions of our practical identity are optional, the fact that we must have a conception of our practical identity is mandatory.

Might there be a similar argument for Oshana, that self-conception is a pre-requisite to an individual possessing any practical reasons whatsoever? Just as an individual must have a description under which she values herself and judges her life to be worth living, perhaps an individual must have beliefs about which of her mental states give her life meaning, in order to be responsive to reasons.

As it was with Korsgaard’s argument, the claim that the possession of a self-conception is a pre-requisite to possessing, being responsive to reasons, is dubious. It seems more plausible

²⁹ Dennett (1976), p191.
that, as Bratman and Williams argued, while the possession of a self-conception or something like it is of great normative value, it need not ground all normative reasons. Conceivably, we can sever a notion of self-conception from the claims of Korsgaard and Oshana about the connection between self-conception and normative reasons. Let’s consider whether such a notion of self-conception, one severed from its theoretical base as the grounding for an account of normativity, can satisfy the distinctness requirement.

Deep character accounted for the distinctness requirement because it allowed for both conscientious and non-conscientious considerations to be distinguished by normative type. Conscientious considerations incorporated those considerations that pertained to the meaningfulness of the individual’s life. Self-conception, suitably refined, seems to be able to make the same claim. It would account for the normativity of conscientious considerations by holding that it is only those considerations that implicate the individual’s beliefs about what makes her life meaningful. Non-conscientious considerations are those normative considerations that don’t involve such a self-conception. How would such a notion be applied in interpreting the Les Misérables cases? The deliberation event could be interpreted as Valjean acting on what he believes would make his life meaningful after weighing that consideration up against other normative considerations. As he believes that acting in a Bishop-inspired way is what would make his life meaningful rather than doing what would have the least negative consequences on the townspeople, it is by acting on the former consideration that he acts in accordance with his conscience.

Earlier I made reference to a potential problem for self-conception which I call the subjectivity problem. This is a problem that seems to follow from the radically subjective nature of Oshana’s self-conception. Recall my initial claim about self-conception: It is entailed by the very notion of a self-conception that an individual’s beliefs about her identity and her actual identity can come apart. While an individual may well believe, on some level, that one or more of her convictions implicates her identity, that belief can be wrong. In the case of Oshana’s account of self-conception this becomes the claim that an agent’s beliefs about those mental phenomena that make her life meaningful can come apart from the phenomena that actually make her life meaningful. It was this subjective feature of self-conception which meant that it satisfied (by definition) the first-personal requirement. But it is this requirement, I would suggest, that is in a very real tension with the distinctness requirement.
My view is that our intuitions about conscientious identity mean that such an account cannot be satisfied by an individual’s beliefs about her identity. Depending on whether we define conscientiousness in terms of beliefs, or in terms of actual identity we end up with radically different accounts of the distinction is between the conscientious and the non-conscientious. For example, let’s consider an individual who believes that it is her philanthropic tendencies that make her life meaningful; that it is her philanthropic tendencies that she cares about at the deepest level. Just as an individual’s beliefs about her neo-Aristotelian character can be overly rosy, so too can her beliefs about which mental phenomena she cares about at the deepest level. That individual’s constant stingy behaviour would constitute evidence that she did not, in fact, care sufficiently about being philanthropic in the way she believed she did. In such a case, it seems counter-intuitive to say that that belief of the agent’s is a conscientious conviction of hers. The distinctness requirement, insofar as it is intuitive at all, captures the idea that the agent's actual identity is captured in her conscientious convictions, not just her beliefs about her identity no matter how unreasonable or deluded those beliefs turn out to be. There needs to be some external constraint on the agent's beliefs about her identity to make them more closely linked to that agent's bona fide identity in our account of conscientiousness. It seems that self-conception simpliciter will not satisfy the distinctness requirement.

One possible solution is to adopt a non-subjective account of meaningfulness. I mentioned earlier the possibility of adopting an account of meaningfulness with an objective element such as Susan Wolf’s account discussed earlier. On such an account an individual’s beliefs about her identity could easily come apart from her identity. Someone could hold a completely objective account of meaningfulness that did not involve an appeal to any subjective component whatsoever. This would solve the problem. But, I think, it would be an ad hoc solution. It would make an individual’s conscientious identity completely removed from matters that she values and therefore difficult to reconcile with the essential first-personal requirement of conscientious identity. In the next section I look at an alternative solution to the subjectivity problem. It offers another way of placing an external constraint on self-conception while still retaining what is importantly subjective about conscientious identity.

---

30 For example William Lane Craig has recently defended the view that the meaning of life is given by God’s purpose for us (1994).
7.4. The Narrative Constraint on Self-Conception

In chapter five I introduced the notion of narrative in order to provide a particular interpretation of Valjean’s behaviour in *Les Misérables*. I supported my interpretation of Valjean’s thought processes, and its status as a paradigm case of conscientiousness, by arguing that such an interpretation best satisfied Hugo’s broader intentions in the novel. In this section I look at accounts of narrative identity. Accounts of narrative identity hold that the notion of narrativity employed in novels (and other literature) is also present (or is ideally present) in the lives of real-life individuals. And, according to narrative theorists, this narrative structures the lives of those individuals in a way that provides an answer to Schechtman’s characterisation question. My particular interest in an account of narrative identity is in looking at those accounts that build on something like Oshana’s self-conception. It is an account of self-conception, with this added external constraint of narrativity which means, I suggest, that such an account can overcome the subjectivity problem and satisfy the distinctness requirement.

Narrative identity, in broad terms, holds that what makes an individual’s desires, beliefs etc. her own is that they play a part in the 'story' of that individual’s life. Most accounts of narrative identity emphasise the way in which narrative identity is constituted by the story that the agent *herself* tells about her life; it is this aspect of narrative identity that explicitly incorporates the individual’s first-personal perspective. I suggested in the last chapter that we can think of the question about conscientious identity as similar to Schechtman’s characterisation question. However, without further specification as to what an answer to the characterisation question should look like, there is no way to specify that, say, an account of identification or an account of neo-Aristotelian identity constitutes a better answer to the question than some other account. In light of this, Schechtman posits four requirements that an answer to the characterisation question needs to satisfy.

---

31 However, I did point out there that this does not mean that my interpretation is the only interpretation of conscience that enables us to make sense of the narrative of the novel. In fact, I suggested there that my interpretation left out an important part of the account of conscience described by Hugo. It left out the idea that the conscience was, in some sense, an external voice speaking to the agent.
First, an answer to the characterisation question must explain how an individual can be morally responsible for her actions. In general, we say that an individual is only responsible for her own actions, rather than the actions of anyone else. We have already seen how accounts of identification and self-conception purport to satisfy this requirement. Second, Schechtman posits, an individual is interested in her own future. While agents are concerned about the futures of other people (such as their children) an agent has a special concern for her own future. An answer to the characterisation question must give an account of identity that explains this fact. Third, and related to the previous feature, is the agent’s interest in compensation. Usually, I can only be compensated for my sacrifices by benefits that accrue to me specifically. Fourth, we care about our own survival. We care about whether, if we get Alzheimer’s disease, we will be the same person afterwards. Schechtman argues that just as any viable account of conscientious identity needs to satisfy the four requirements I set out, a viable answer to the characterisation question must satisfy the four requirements that she specified.

Schechtman defines her account of narrative identity as follows:

According to the narrative self-constitution view, the difference between persons and other individuals (I use the word “individual” to refer to any sentient creature), lies in how they organise their experience and hence their lives. At the core of this view is the assertion that individuals constitute themselves as persons by coming to think of themselves as persisting subjects who have had experience in the past and will continue to have experience in the future, taking certain experiences as theirs. Some, but not all, individuals weave stories of their lives, and it is their doing so which makes them persons.32

In the above excerpt Schechtman emphasises that her account of narrative identity is an account of narrative self-constitution. It is an account of how the individual constitutes herself. Furthermore, the individual constitutes herself by thinking about herself in a certain way, by having a self-conception. There are similarities here between Schechtman’s view and Bratman’s view discussed earlier that it is the individual having a conception of herself as temporally extended that, at least partially, means that that individual is temporally extended.33

33 This is also similar to Williams’s argument earlier that it is by possessing ground projects and categorical desires that an individual persists across different time-slices.
It is when describing the self-conception contained in her account of narrative identity that Schechtman introduces the notion of narrative:

A person’s self-conception is a narrative self-conception, then, insofar as the experiences and incidents that make it up are not viewed in isolation but as part of the ongoing story that gives them their significance.\(^{34}\)

Here Schechtman is claiming that in order for an individual to possess an identity that individual’s self-conceptions over time must have ‘the form and the logic of a story’. What exactly is it for an individual’s self-conception to be viewed as part of an ongoing story that gives that self-conception its significance? According to Schechtman it is for that story to be in some sense intelligible or coherent. Intelligibility or coherence, it is claimed can come in degrees. Some lives are more or less coherent than others. It is worth noting how on Schechtman’s account, narrative identity has both subjective and objective elements. The identity is constituted both by the individual’s possession of a self-conception and also by that self-conception meeting certain external, objective criteria. It is this aspect of narrative identity, I shall suggest, which means such account can overcome the subjectivity problem.

It is common for accounts of narrative identity to combine subjective and objective elements. Kim Atkins defends such an account where she says:

The narrative model is not simply a first-person report of an individual’s subjective experiences and point of view, it is a complex model that interweaves the first-person subjective perspective with the second-person communicative situation of social existence, along with a generalisable or third-person perspective presupposed by a shared world of meanings with public standards of objectivity.\(^{35}\)

For Atkins there is not just a purportedly third personal ‘objective’ constraint on self-conception but also a second-personal constraint in the sense that a self-conception must be directed at another person. In Atkins view, an individual’s narrative identity is constituted by the unity of these three perspectives.\(^{36}\)

---

\(^{34}\) (1996), p97.

\(^{35}\) (2004), p343.

\(^{36}\) Other theorists who incorporate external constraints on the first-personal perspective into their accounts of narrative identity include Catriona Mackenzie and Marina Oshana. Mackenzie defines narrative identity as
Can we supplement Oshana’s notion of self-conception discussed in the previous section with this external constraint of narrative and arrive at an account of self-conception that satisfies the distinctness requirement? Supplementing the account of self-conception would result in the following account of conscientious identity:

An individual’s conscientious identity consists in that individual’s beliefs about the various mental phenomena that:

(a) Make her life meaningful

(b) Are coherent with her life taken as a whole.

Interpreting conscientious identity in this way avoids the subjectivity problem because the beliefs of the individual have to cohere with her timeline of thoughts and actions in general. Conscientious identity is about the way in which only some of our genuinely normative practical considerations implicate identity and some do not. But that distinctness is not captured by holding that some of those considerations capture the individual’s beliefs about her identity and some do not: if it is to constitute an individual’s conscientious identity a self-conception cannot be radically subjective. The narrative constraint means that only those beliefs of the agent which cohere with her history of action and thought that are constitutive of conscientious identity.

Earlier when referring to Oshana’s account of self-conception I described her extensive list of mental phenomena which could be the objects of belief for the purposes of self-conception. While that list was not supposed to be exhaustive, it included phenomena such as that individual’s cognitive states, her emotional dispositions, states of the body and values. There was no mention by Oshana as to what, if anything, unites those phenomena in being the objects of self-conception. In this section I suggest that we should restrict these objects so that it is only the individual’s projects that can be the object of her self-conception.

patterns of coherence and psychological intelligibility within our lives, connecting our first-personal perspectives to our history, actions, emotions, desires, beliefs, character traits, and so on ((2008), p12).

In the case of Oshana’s ‘hybrid account’ of identity (not to be confused with her account of self-conception discussed in the previous section) she makes the following claim:

The self and the distinctive identity it bears, owes its genesis and sustenance in part to the active role persons take in directing their lives, and in part to the constructive and interpretive contributions imposed by the person’s environmental position ((2010), p46).
There is clearly a sense in which considerations like body shape and emotional disposition are relevant to an individual’s identity. But I don’t think that they are relevant to an individual’s conscientious identity. For example, I mentioned earlier the possibility of an individual believing that her left leg is not ‘part of her’. But that belief does not seem to be conscience-relevant. If an individual decided to get surgery to amputate that limb we would not say that was an act of conscience. My suggestion is that we restrict the objects of self-conception simply to the values of that individual – one of the items on the list. An individual’s ‘values’ in turn can be defined as those things/actions that the agent believes to be desirable.\(^{37}\) On plausible accounts of the connection between valuing something and desiring it, there is a rational connection between possessing this belief and actually desiring that object.\(^{38}\) Thought about in this way, an individual’s values are similar to what Bratman called a quasi-policy, the ideals of an agent that can be incorporated into her plans and general practical deliberation. The advantage of thinking about not just the individual’s beliefs and desires but her quasi-policies, is that these ideals play a role in the individual’s cross-temporal organisation that mere beliefs and desires do not.

7.5. Narrative Self-Conception and the Practicality Requirement

Thus far we have established that an account of self-conception (suitably refined as narrative self-conception) can satisfy the first-personal and distinctness requirements. Can it satisfy the practicality and endurance requirements? According to my interpretation, conscientious considerations, if an individual has any, necessarily affect the individual’s practical thought and action. That is because conscientious convictions implicate the ideals of the agent (quasi-policies) that she believes make her life meaningful. Or, to put it in simpler terms, conscience concerns the agent beliefs about the ideals that give meaning to her life. These beliefs are conceptually connected to the individual’s thought and action because if an individual is committed to some ideal, insofar as she is practically rational, she is motivated towards that thing. Furthermore, the narrative constraint on conscientious identity means that narrative self-conception satisfies the practicality requirement in another way. If an individual’s

\(^{37}\) David Lewis defends such an account of value in (1989).

\(^{38}\) See, for example Smith (1989) and (1994).
putative conviction does not cohere with her actions then that putative conviction is not a \textit{bona fide} conscientious conviction.

Perhaps the connection is even stronger than my presentation would suggest. Recall that Schechtman described her account as one of \textit{self-constitution}. It not just an account of who the individual is but also an account of how the individual creates herself. It seems plausible that there is a causal connection between an agent’s actions and the self-conception that she ends up having. If an individual only has an embryonic self-conception it may well be through continual action that that embryonic conception is endorsed as her fully-fledged narrative self-conception.

John Davenport makes this point when he argues that narrativity develops cumulatively through “compounding associations” between elements of that narrative.\textsuperscript{39} It is through constantly acting (and thinking about acting) in a certain way, that an individual creates a self-conception or identity.\textsuperscript{40}

Finally, does narrative self-conception satisfy the endurance requirement? I argue that it does satisfy this requirement because of the narrativity constraint. Consider the ways in which the accounts of identity considered thus far failed to satisfy the endurance requirement: Accounts of identification failed to do so because there was only a contingent connection between possessing a second-order volition and that volition enduring. Similarly with policies and quasi-policies. While it may well be the very purpose of such policies and quasi-policies that they endure, the mere existence of a policy does not guarantee that it will endure. Neo-Aristotelian character, too, can exist for a mere moment. While we may well take it to be evidence that an individual did not possess a second-order volition, a policy, or a character if it did not endure, nothing in principle prevents such a notion from doing so.

In order for a self-conception to be constitutive of an individual’s conscientious identity that self-conception must be coherent with that individual’s general history of thought and action. The narrative component does not simply require that some commitment purport to endure, like in the other accounts of identity I have considered. The requirement is that that ideal has

\begin{itemize}
\item \textsuperscript{39} (2012), pp82-82.
\item \textsuperscript{40} Christine Korsgaard also defends an account of identity (though not narrative identity) whereby, in action, the agent is constantly constituting that identity. She claims that “we have a form of identity that is constituted by our chosen actions” ((2009), p130).
\end{itemize}
some place in the story of that individual’s life. One might wonder though, would it not be possible for an individual to possess a self-conception for a mere moment as part of her narrative? Perhaps we could imagine an individual with dissociative identity disorder who possessed a personality for a mere moment but that conception still plays an important role in that individual’s narrative. That self-conception could have a high degree of coherence with the rest of the events in her life. But recall that the endurance requirement is not the requirement that that self-conception exist for a set amount of time. Rather, it is the requirement that that self-conception be of enduring significance. And it is this requirement of conscientious identity that is best satisfied by narrative self-conception.

7.6. Valjean and Narrative Self-Conception

In the previous section I argued for an account of narrative self-conception that satisfies the four requirements of conscientious identity. Unlike accounts of identification, neo-aristotelian character and deep character, this account unites the first-personal perspective with the actual practical thought and actions of the agent. This account also satisfies the distinctness requirement as it is not committed to the view that all of our purportedly normative practical considerations are, at base, an appeal to narrative identity. While it is common for accounts of self-conception to claim a strict conceptual connection between matters of self-conception and all purportedly normative reasons, this is not an essential element of self-conception. Consequently, a suitably refined account of self-conception can satisfy the distinctness requirement. It can explain how it is possible that only some of our potentially normative practical considerations involve an appeal to conscientious identity. It satisfies the practicality requirement in several different ways. First, self-conception is an appeal to meaningfulness which is itself a type of practical normative consideration. Second, it is plausible that by acting on a practical consideration that an individual goes someway to constituting her very self-conception. Narrative self-conception satisfies the endurance requirement because the narrative element requires that the self-conception have a lasting significance in the individual’s life.

At this point it is worth explaining just how this account of narrative self-conception, as I have articulated it, connects with my interpretation of both the Les Misérables cases in chapter five. In the beginning of my description of the transfiguration case, I initially glossed
the event of transfiguration as Valjean arriving at a certain way of seeing himself. In this chapter this notion has been re-interpreted as the individual’s beliefs about herself. Specifically, I offered the following analysis of conscientious identity:

An individual’s judgement implicates her conscientious identity just in case that judgement is about an ideal or ideals of the individual that she believes makes her life meaningful and those beliefs satisfy her life narrative; they cohere to a high degree with her general history of thought and action.

How would an interpretation of the Les Misérables cases involving this notion of conscientious identity go? I suggest the following: In the transfiguration event, Valjean comes to a belief that carrying on in his usual brutish manner would not be the route to a meaningful life. In light of this Valjean decides that he shall use the ideal of the Bishop as the basis for some more virtuous commitments. This initial self-conception is confirmed as his conscientious identity in the deliberation event because of the role that that self-conception plays in the individual’s life. This self-conception coheres to a significant degree with his history of thought and action as revealed in the deliberation event where that self-conception moves Valjean to act.

The analysis of conscientious identity that I have just offered is the culmination of a long line of argument traversing the last three chapters. It is worth, at this stage, summarising the argument that has got us to this point. At the end of chapter four I had suggested that, strictly speaking, there are no necessary conditions of conscience, but rather, three principles. As chapters three and four had focused on the communicative and morality principles, and since those principles have been reasonably heavily discussed in moral philosophy already, I suggested the remainder of the thesis be devoted to conscientious identity. In chapter five, I offered an interpretation of conscience in Victor Hugo's Les Misérables, which, I claimed, constituted a paradigm case of conscience emphasising conscientious identity. From those two paradigm cases, I extracted four requirements of conscientious identity.

Then in chapter five I used those four requirements to assess accounts of identification. After determining that those accounts failed, I moved on, in this chapter, to considering accounts character and self-conception. I constructed an account of conscientious identity built on Marina Oshana’s notion of a self-conception. An agent's self-conception is constituted by her
beliefs about those of her ideals which make her life meaningful. These beliefs, while a necessary condition of an agent possessing a conscientious identity, I argued, were not a sufficient condition. In order to overcome the subjectivity problem (that my identity can’t simply be constituted by subjective mental states), I suggested an external constraint on those beliefs. This was the constraint of *narrative*. It is those beliefs of the agent about what makes her life meaningful that also *cohere* with her general run of thoughts and actions which are constitutive of her conscientious identity. This is an account of conscientious identity that satisfies the descriptive constraints on such a notion.

There is a crucial element of my defence that is still missing. At the end of chapter four I argued that each principle of conscience had to be justified in terms of the three normative heads of conscience. I suggested at that point that we could not assess the principle of conscientious identity as to whether it satisfied any of those normative heads without having a more precise idea of what conscientious identity involved. Now that, we have a precise notion of conscientious identity in sight, we are in a position to assess that notion in terms of those three normative heads.

7.7. Narrative Self-Conception and the Normative Heads of Conscience

I have argued throughout this thesis that conscience should be defined in terms of three principles: the communicative principle, the morality principle and the principle of conscientious identity. A set of descriptive requirements applies to each of these principles (such as the four requirements of communicativeness or the four requirements of conscientious identity). For any conviction or action to count as conscientious it must satisfy *one* of these principles and the *most* conscientious actions will satisfy all three. What we want, ultimately, is the conception of conscience that best justifies a *right of conscience*, if there is such a thing. In that chapter I argued that the morality and communicative principles can be justified in terms of the three normative ‘heads’ of conscience (the humanistic, psychological benefit and societal value heads) that underlie the right of conscience. An individual has a right to act on her *moral* convictions as they are justified by all three normative heads. A moral conviction is the expression of an individual’s autonomy. It is psychologically beneficial for an individual to act in accordance with her moral convictions.
because of the guilt that can result from failing to do so. It is societally beneficial because of (a) Millian epistemic arguments as to the value of free speech/action and (b) a Rawlsian argument that in the original position we would all agree to a right to act on such convictions.

An individual’s right to act on her communicative convictions was also a legitimate exercise of her autonomy. There wasn’t any obvious psychological cost in not being able to communicate one’s convictions. There was societal value in allowing individuals to act on communicative convictions. As in the case of moral convictions, there were good Millian arguments as to the societal benefits of acting on a conscientious conviction.

Turning to the identity principle (‘conscientious identity’) now, how do the normative heads of conscience ground narrative self-conception? Beginning with the humanistic head; narrative self-conception is an expression of one’s autonomous choice of some ideals over other ones. Because it involves ideals, and in particular, ones that make the individual’s life meaningful then possessing a narrative self-conception is an exercise of one’s right to pursue a particular conception of the ethical. But it is more fundamental than this, I think. Because conscience concerns the individual’s ideals, and because conscience involves the matters that the individual considers the most meaningful, it has an organisational role in her practical psychology. It doesn’t simply provide the individual with one practical reason, on a par with any other. Rather, it concerns an overarching quasi-policy in terms of which many practical reasons have to be arranged. It is not open for an individual to ignore the dictates of her conscience in the way that, perhaps, it might be open to her to ignore some of her moral and ethical views. In many respects convictions of conscientious identity play a foundational role in the individual’s practical psychology, upon which other convictions of the individual can be based.

What about the principle/head of psychological benefit? We have already seen several descriptions of the psychological consequences of failing to act in accordance with one’s identity. Taylor referred to this state as a state of disorientation; being ‘at sea’. Korsgaard referred to it as ‘life not being worth living’. For Arendt, it was the feeling of not being able to ‘live with oneself’. These locutions suggest a definite sense in which it is particularly psychologically troubling to act contrary to one’s identity. When thinking about narrative self-conception, I think, this psychological effect is particularly obvious. In asking an individual to act contrary to her narrative self-conception we ask her to act in a way that is
contrary to what she finds meaningful in her life, and understandably, a life without acting in accordance with one’s own views about what is meaningful will be difficult for the agent.

What about the third head of normativity, societal value? Society, and indeed the state, have certain views about what makes a meaningful life. Perhaps the state funds art galleries and symphony orchestras not simply because they maximise the welfare of citizens, but because they contribute to the meaningfulness of those lives. Perhaps the state introduces compulsory military service not simply so the country can carry out the campaigns that it is believes are morally appropriate but also because of a belief that individuals ought to stand up for their country; it will contribute to the meaningfulness of their lives. These views can be wrong, of course. An individual, in acting on the beliefs contained in her conscientious identity, can act as a corrective of those views. And even if those views are correct, the individual in acting on her conscientious identity can serve as a reminder of the justification for those views and keep the meaning of those views alive. Furthermore, as per Mill’s fourth argument, an agent acting on her narrative self-conception helps prevent society from forgetting the underlying meaning of its views.

Finally, a Rawlsian contractualist argument would also support the right to act on one’s narrative self-conception. Our beliefs about what is meaningful are ones that most of us have and we cannot easily do away with. Therefore, if we were deciding on what the just rules for our society would be, we would decide on a rule that allowed us to act in accordance with such conscientious views.
Bibliography


Eweida and Others v. The United Kingdom [2013] ECHR 37


