University of Adelaide

Department of Philosophy, School of Humanities

Self-Deception and Wilful Ignorance:

‘Self-induced, purported ignorance?’

By:
Margaret Leila Penhall-Jones
BA (Hons) MA

Thesis submitted in fulfilment of the requirements of the degree of Master of Arts by Research, MPhil. March, 2022.

Acknowledgements

It has been an honour and a pleasure to work with my supervisor and advisor in philosophy, Emeritus Professor Greg O'Hair. I feel I am fortunate to have had the benefit of Greg's depth of philosophical knowledge, and his wisdom in his approach to supervision. I greatly appreciate Greg's ability to intuit what I might be trying to explore and offer suggestions without prescription. Under his influence, this thesis has transformed, as has my own understanding of the puzzling and unwieldly central concepts. I also thank Greg for many enjoyable discussions across a wide range of philosophical issues, which provided background, and expanded my understanding of areas of philosophy which were new to me.

My thanks and admiration go to my primary supervisor, Professor Natalie Edwards, who provided excellent advice, support and solutions, particularly to those issues which sometimes arise when 'life gets in the way' of study. Natalie kept this project on track and navigated the administrative issues with a calming professionalism. Along with other postgraduate students, I appreciated Natalie's efforts in providing a very human contact through special online presentations and discussions during the coronavirus lockdown in Adelaide.

My appreciation also to co-supervisor Professor Philip Gerrans who stepped up and stayed the course as co-supervisor, providing ongoing essential oversight and insight.

Also from the Department of Philosophy, I acknowledge Dr Jonathan Opie, who as Head of Department offers workable solutions and is open, approachable, and sensitive to students' individual situations.

Looking back to the semester before I enrolled in this program, my appreciation goes once again to Philip Gerrans and also to Professor Gerard O'Brien. I thank them for encouraging me to enrol in postgraduate work. I had believed that I had closed the door on philosophy many years ago. Although I had since pursued careers in government policy and psychotherapy, I was uncertain that I could return to the study of philosophy. Their generosity in suggesting I audit their courses and attend professional staff lectures allowed me to again experience my enthusiasm for philosophy, and also made me realise I should make the attempt.

One of the positives of returning to study philosophy has been to re-connect with people I had known or worked with previously. These include Emeritus Professor Max Deutscher, an inspiring influence for me in my previous studies at Macquarie University. Now, I thank Max for his generosity in sharing his thoughts and some of his more recent writing.

My research on this thesis was undertaken entirely during the coronavirus pandemic. I attended campus only once before the University of Adelaide, and Adelaide itself, was locked down. As a result, it has been disappointing for me that I missed out on much of what makes philosophy exciting. That is, sharing diverse ideas from different schools of philosophical thought with other postgraduate students. However, when I did manage to attend discussions and presentations, whether in-person or by Zoom, I found my my postgraduate colleagues to be positive, open, and generous with time, ideas and suggestions. I wish them all well in their future careers.

Finally, I would like to thank my husband, Tom Benjamin. During the coronavirus lockdowns, we were each working at home on our separate research projects, and we'd share mid-morning coffee breaks. I have clear images of us having coffee on our porch near the rose garden, Tom patiently listening to a stream of more or less familiar names and unfamiliar theories: Sarch, Sartre, Davidson, Mele, Moody-Adams, Wieland…whoever fascinated or worried me at the time. I thank Tom for listening, offering his feedback and for some significant insights.

Margaret Penhall-Jones.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the Australian Government Research Training Program.

Signature:

Date: 7 March 2022

Contents

Acknowledgements **2**

Declaration **3**

Abstract **7**

Part One **8**

Introduction and Overview **8**

The concepts **9**

Definitions **11**

Outline **11**

Overview of approaches to self-deception and wilful ignorance **13**

Existentialism **13**

Self-deception in Analytic Philosophy **13**

Intentionalism **14**

Motivationalist Theories **14**

Other approaches to self-deception. **15**

Theories consequent on Intentionalism and Motivationalism **15**

Self-Deception as Failure of Self-Knowledge **17**

Self-Deception and Tension: **18**

Normative Aspects of self-deception **18**

Evolutionary Biology **18**

Focus of this thesis in relation to self-deception **19**

Wilful Ignorance **19**

Part Two **21**

Self-deception **21**

Introduction **21**

Paradigm Examples **21**

The key features of self-deception **24**

The puzzle of self-deception **24**

The paradoxes of self-deception **24**

The Static Paradox **24**

The nature of belief **25**

The Dynamic Paradox **27**

First Person Authority **28**

Selectivity issue **31**

Tension ... **31**
Further Question: What is the purpose of self-deception? ... **31**
Approaches to Self Deception ... 33
Approach to Theory ... **33**
The structure of arguments explaining self-deception. ... **33**
Additional Criteria for Assessing Theories of Self-Deception ... **34**
Comparison of major theories of self-deception ... 35
Existentialism - Jean-Paul Sartre ... **35**
Sartre's paradigm examples ... **36**
The phenomenology of 'belief' and self-deception ... **37**
Translucent consciousness and self-deception ... **37**
'The Other' and first-person authority ... **38**
Remarks on Sartre's theory of self-deception ... **39**
Intentionalism - Donald Davidson ... **40**
Donald Davidson ... **40**
Rationalism and Naturalism ... **40**
Holism and the web of belief ... **41**
Reasons as Causes and Rationalising ... **42**
Radical Interpretation and the Principle of Charity ... **42**
Triangulation ... **42**
The Puzzle of Irrationality ... **43**
Motivationalism ... **47**
Alfred R Mele ... **47**
The challenge of twisted self-deception ... **51**
The explanatory power of the three major theories of self-deception ... **56**
Table Showing Theories of Self-Deception, Explanatory Desiderata and application to Paradigm Examples. ... **57**
Conclusion to Part Two – Self-Deception ... **61**
Conclusion Additional Issues ... **62**
Part Three ... **64**
Wilful Ignorance ... **64**
Paradigm Examples ... **64**
Albert Speer and wilful ignorance ... **64**
Hannah ... **66**
Drug transporting cases ... **67**

Definition **68**
The 'knowledge' issue in wilful ignorance **69**
Access to Knowledge **70**
Motivation **70**
Proposed definitions **70**
Negligence: when ignorance is not a defence **71**
Wilful Ignorance and the duty to inform oneself **73**
Wilfully Ignorant for Good Reasons? **73**
Kevin Lynch **74**
Michelle Moody-Adams **77**
Jan Willem Wieland **78**
Normativity and the Counter-factual **81**
Conclusion for Wilful Ignorance **85**
Conclusion **86**
Appendix A Table of Paradigm Examples **90**
Self-Deception and Wilful Ignorance **90**
Appendix B Table showing necessary or sufficient conditions for self-deception **92**
BIBLIOGRAPHY **94**
Cases cited: **98**

Abstract

Self-deception and wilful ignorance each involve professed attitudes involving ignorance which, in the circumstances, are difficult for others to believe. Some recent philosophical discourse considers whether they are the same or similar phenomena, with conflicting results. A clarification of issues, including identifying some conditions conducive to comparison of these phenomena, such as I present here, is aimed at contributing to this debate.

Comparison of two phenomena requires a clear concept of each. A review of the literature showed that this is not available for self-deception. Various incompatible theories have been proposed to explain self-deception. In these theories, argument often seeks to derive sufficient conditions from paradigm examples. Theory has reached an impasse between the two leading theories in Analytical Philosophy: Intentionalism and Motivationalism. The first task, then, is to identify the theory of self-deception to be used for comparison.

Paradigm examples can also be useful as tools of analysis. This is what I have done here, with the aim of discovering the currently established theory of self-deception with the most explanatory power. Analysing some standard paradigm examples for distinctive features of self-deception, I have derived several 'desiderata' that a theory must address. I have added two questions, regarding the reasons for and purpose of self-deception.

I then review representative philosophers from three schools of philosophical thought on self-deception. They are:

- Existentialism –Jean-Paul Sartre
- Intentionalism – Donald Davidson
- Motivationalism – Alfred R Mele

While each of these philosophers deepens our understanding of self-deception, there are reasons to prefer the theory that Davidson provides, as the one with the most explanatory power. This theory will provide the comparator to wilful ignorance.

As a legal concept used to establish *mens rea* in criminal cases, wilful ignorance is a more clearly defined phenomenon than self-deception. Recent philosophical discussion concerns wilful ignorance in relation to our moral and social responsibilities, locally and as global citizens. Thus its boundaries become less well-defined. As well, normative considerations, whether concerning justice or ethics, seem to be inextricable from the concept of wilful ignorance.

In relation to the question of whether self-deception and wilful ignorance are similar phenomena, I conclude that provided a Davidsonian theory of self-deception is accepted, they share many structural and attitudinal features. The reach of each phenomenon is apparently different: self-deception is more personal while wilful ignorance has broader social ramifications. Yet neither self-deceivers nor the wilfully ignorant are seeking insight or change. The purpose of these attitudes is arguably to maintain one's self-concept or lifestyle when that is challenged.

In relation to the second question in this thesis, this is an area of similarity. The state of ignorance in both is self-induced and incomplete. In both, there is an understanding of the matter over which there is a claim of ignorance. Such ignorance is 'purported'. Wilful ignorance and self-deception both involve self-induced, purported ignorance.

Part One

Introduction and Overview

An emerging debate in philosophy is whether wilful[1] ignorance and self-deception are the same phenomenon. My aim is to contribute to that debate by attempting to explore the relationship, if any, between self-deception and wilful ignorance. In particular, I will be considering the proposition that both wilful ignorance and self-deception involve 'self-induced, purported, ignorance'. That is, I am arguing that the 'ignorance' is modified, as these terms suggest and as outlined below.

The first term, 'self-induced' means 'caused by oneself'. It refers to agency and strongly implies intentionality. I am considering whether, and if so to what extent, wilful ignorance and self-deception might be intentional.

The second adjectival term, 'purported' means something that is said to be true but is not definitely true.[2] It refers to something that has been stated to be true, or to have happened, although this may not be the case. Alternatively, something that is claimed but not proved to be true.[3] This term leaves open the question of the status of the third term, the noun 'ignorance'. In this thesis, at the outset, I am neither assuming nor rejecting the claim of 'ignorance' in relation to self-deception or wilful ignorance.

The term 'ignorance', in its colloquial understanding, is a state of not knowing or not being aware of something. There are negative connotations to ignorance, partly deriving from its association with lack of education, and therefore with uninformed or biased views. However, this is not the sense in which I am using it here. I am concerned with how an individual can claim not to know a particular piece of information.

Amongst a developing interest in the concept of 'ignorance' in a number of disciplines, including philosophy and epistemology, there is yet to be consensus on a definition. A useful attempt at identifying acceptable premises which can set boundaries to the concept of ignorance is proposed by Selene Arfini. I reproduce two of her premises which are useful for this project, as follows[4]:

---

[1] Amongst a number of possible ways of spelling the term 'wilful' I have chosen the Australian spelling as is used in Australian legislation for my text. When quoting other texts I reproduce the spelling used by the original author.

[2] Merriam-Webster.com Dictionary, s.v. 'purported' https://www.merriam-webster.com/dictionary/purported.

[3] https://dictionary.cambridge.org/dictionary/english/purported

[4] Arfini, Selene. *Ignorant Cognition. A Philosophical Investigation into the Cognitive Features of Not Knowing.* Springer. 2019. 12

1. ignorance can be defined as a cognitive condition that can be either passively (and unconsciously) born by an agent or actively nurtured by her;
2. the cognitive state of ignorance entails epistemic limitations (which can be any lack of knowledge, belief, information or data) that affects the behaviour, the belief system, and the inferential capacity of the agent. [5]

Of the two phenomena considered here, self-deception is the more familiar. Self-deception is an attitude which has been acknowledged for centuries. Self-deception is exhibited by our neighbours and ourselves. We recognise it and think we know it. Yet, when we turn a critical eye to it and attempt to understand it, we are confronted with a puzzle; how can one really deceive oneself?

We are less familiar with the term wilful ignorance, which is also called wilful blindness,[6] but we do acknowledge it. It is useful in criminal law to prevent a misuse of the defence of ignorance, when ignorance is deliberate. This narrow focus has expanded to encompass situations in which individuals deliberately maintain ignorance on issues, particularly those with serious social, economic, or environmental impact.

Each of these two phenomena apparently involves ignorance in circumstances where it is difficult to believe the ignorance is complete and genuine. This, in turn has opened up interest in the relationship between wilful ignorance and self-deception: are they the same, similar or diverse phenomena?

The concepts

To begin, consider these paradigm examples: one of self-deception and the other of wilful ignorance.

Emily represents the paradigm example of self-deception.

Emily

Three years into her marriage to Daniel, Emily finds herself more often alone at home. Daniel recently works late and is frequently away on work-sponsored conferences with his personal assistant, Lily. When she meets their friends for coffee, alone, Emily knows they think Daniel is having an affair. Emily believes Daniel is faithful to her and tells her friends so.

This is a standard, paradigm example of self-deception. Putting oneself in the position of one of Emily's friends, consideration of this example allows us to tease out some unusual features of this situation. Her friends are likely to say about Emily:

'She really must know what is going on.'

This suggests the following issues about self-deception:

Although Emily has perfectly good reasons for believing that her husband is still faithful, the weight of evidence suggests that he is not. As friends, we cannot believe Emily doesn't know Daniel is having an affair. Yet, we do not accuse Emily of cynically lying about her belief.

Emily's beliefs seem to be irrational in relation to her husband's possible affair. This is the source of the puzzles of self-deception. Emily's stated belief seems to be genuine enough, but

[5] There is a third premise, which concerns Arfini's particular project and is beyond the scope of this thesis. Arfini S. *Ignorant Cognition.* 12. The premise is: '3. describing how the human agent is in a condition of ignorance means to recognize those cognitive traits that define ignorance as a particular cognitive state and investigate how they specifically affect her cognitive capacities.'

[6] There appears to be some connotative difference, but these terms are in practice used inter-changeably.

it is a genuine belief held and maintained in defiance of the evidence. Even though she believes against the evidence, Emily does not believe with no evidence. There are the three years of marriage.

We do not believe Emily's statements about her own beliefs. That is, her first-person authority is challenged.

Yet, Emily is an independent, highly-functioning person, with a job, a marriage, friends and considerable insight. She would have difficulty maintaining these if she had pervasive problems such as magical thinking, delusions or other psychopathology which impacted on her grasp of reality.

The fact that Emily is mostly understandable raises the question of why is it this issue, of her husband's faithfulness, about which Emily deceives herself? This is the 'selectivity issue' in self-deception.

As well, there is a normative aspect to self-deception. We tend to hold Emily responsible for her self-deception.

Now, consider Hannah, as a paradigm example of wilful ignorance.

Hannah

Hannah is socially aware. With significant effort, she recently eliminated single-use plastic from her life. She has a good job, but is 'time poor' She buys new 'fast fashion' clothes for work every season, reasoning that now, everyone can afford clothes and offshore manufacturing means work for people in low-income nations.

When the tea-room chatter at work turns to the impact of 'fast fashion' in socio-economic terms and on the environment, Hannah quickly leaves the room. Hannah can't feel guilty about something if she does not know about it.

Suppose we are among Hannah's work colleagues. Some time after the conversation about fast fashion, someone remarks on Hannah's wardrobe. Hannah, we acknowledge, is not oblivious to social and environmental issues. After all, she went to great lengths to eliminate single-use plastic from her life. How is it she continues to wear fast fashion? Doesn't she know about the environmental impact of polyester production? Doesn't she know about the possibility of poor conditions, including modern slavery, of overseas workers in this industry? Or the impact of discarded fast fashion on the environment?

Finally, someone, perhaps rather judgementally, says: 'The information is all there on the internet. If she doesn't know, it's because she doesn't want to know.'

People who know Hannah do not accept her apparent ignorance. They identify some intention in her maintenance of ignorance.

Nor is it the case that Hannah is unthinking or careless about moral or social issues. She is a responsible and competent person capable of taking significant action on social or environmental issues.

There is a 'selection' issue in Hannah's case. Why is Hannah wilfully ignorant about fast fashion, when she has demonstrated she is capable of investigating social issues? Further, if Hannah's colleagues are right about her, that she doesn't know because she doesn't want to know, then how is it that Hannah continues in ignorance? How does she know where *not* to look?

There are some parallels in the situations of Emily and Hannah, and apparent differences. This situation allows us to pursue the question of the relationship, if any, between these phenomena.

Definitions

When considering phenomena such as I am considering here, the first recourse would normally be to established definitions.

However, although self-deception is the more familiar term, it has no accepted definition in the philosophical literature. In recent philosophical debates, there have been attempts to outline necessary and sufficient conditions for self-deception. However, no agreement has been reached on these. A few have argued it is a matter of synonyms, mere 'folk psychology', or misconceptions due failure to properly examine the concept.

This attitude will not answer the dilemma. We recognise self-deception. Self-deception is an attitude which has been acknowledged for centuries, including in discourses originating in theology or moralism, (where congregations are exhorted to avoid self-deception) and more recently, in psychology and philosophy.

By contrast, the concept of wilful ignorance is one which has been explicitly recognised in Common Law jurisdictions and in United States courts since the 19th century, for the purpose of determining knowledge, or *mens rea*, in relation to a crime.[7] This means that there are attempts at definition in the law which can form a basis for our understanding of the issue.

The advantage of relying on legislative definitions which may be used in a criminal trial, and which may in turn involve serious consequences for a defendant, is that the concept is applied in practice, and usually with considerable care. The difficulty is that the concept is applied on a case-by-case basis, and interpreted for juries by the sitting judge. Judges' rulings are not always consistent. The result is that while the core concept of wilful ignorance is relatively stable, it is also a little 'soft around the edges'.

This, along with some infamous examples of wilful ignorance, has brought attention to the relevance of this concept for the individual's responsibility as a social and global citizen. We do not prosecute people for social failures, for their thoughts, nor for apparent gaps in knowledge of social issues. Consequently, while a legal definition may guide our consideration, it will not necessarily place a boundary on which instances might be identified as wilful ignorance.

Thus, in this project, we are contrasting a concept which is familiar, but has no accepted or acceptable definition, with a concept that seems to be well defined, but the definitional limits may be too narrow.

Outline

The above considerations mean that asking a general question regarding the relationship between self-deception and wilful ignorance is unlikely to be a productive way to proceed. Therefore, I will first attempt to delineate the most familiar of the concepts: self-deception.

Following introduction to the topic in Part One, in Part Two, I will consider the major theories of self-deception and attempt to identify the theory with the most explanatory power. Above I have suggested some recognised characteristics of self-deception, and these will inform the consideration of the relevant theories.

[7] Alexander Sarch. *Criminally Ignorant: Why the Law Pretends We Know What We Don't.* Oxford University Press. 2019. 13. DOI:10.1093/oso/9780190056575.001.0001

I will start with paradigm examples. I will not be using these as a basis for a theory of self-deception. That has been done by many philosophers already. Rather, I will use the paradigm examples to identify indicators or desiderata that a theory of self-deception must explain.

I will then outline and analyse three major approaches to self-deception in the philosophical literature. The aim will be to use the indicators ascertained from the paradigm examples to identify the theory of self-deception with the most explanatory power. That is, the theory which best accounts for the desiderata, either by explaining each or obviating the need for explanation of that indicator.

I will consider representative theorists in each major approach. These are:

- Existentialism – represented by Jean-Paul Sartre
- Intentionalism – represented by Donald Davidson
- Motivationalism – represented by Alfred R Mele

While each of these theories advances our understanding of self-deception, I will be arguing that the Intentionalist approach as articulated by Donald Davidson provides the most explanatory power. Problems remain, due to the perceived necessity of a theory partitioning the mind. I further argue that this partitioning is not necessarily the strong division suggested by theories positing an unconscious. Self-deception is a puzzling phenomenon occurring in the conscious attitudes of individuals.

In Part Three, I will consider the concept of wilful ignorance. I will first focus on the concept as derived from legislation and court proceedings. I will then consider the understanding of wilful ignorance as it moves outside of the courtroom to encompass our responsibilities as social and global citizens.

The issue of whether wilful ignorance and self-deception are the same phenomenon has already been considered by some philosophers. It is appropriate to assess two of these. Kevin Lynch considers doxastic issues and argues that the two phenomena are different psychological kinds. Jan Willem Wieland, focussing on relevant attitudes, develops an approach which does not make a distinction between classic cases of wilful ignorance and those which others have called self-deception. Neither Wieland nor Lynch allow us to definitively answer this question, however much they contribute to our understanding.

If the current project can clarify some aspects of this debate, including the explanation of some perplexing elements of self-deception, it will hopefully be a worthwhile contribution to the subject.

Overview of approaches to self-deception and wilful ignorance

Self-deception is a term which describes a familiar, every-day phenomenon, and which tends to attract negative normative evaluations. It is therefore unsurprising to discover references on self-deception from a variety of perspectives; theological, moral, psychological as well as philosophical. It is also standard procedure to use paradigm examples to ground discourse on self-deception.

In the 18th Century, Bishop Butler provided a concise depiction of self-deception as a way for people to excuse their bad behaviour and attitudes.[8] Butler takes issue with the self-deceivers: the near-bankrupt who deliberately does not inquire into their finances, in order to maintain their extravagant habits and avoid paying their debts, and more generally, the harm done by the amoral culture and 'rakes' in 18th century Europe. Butler portrays the self-deceived as intentionally closing their eyes to the truth, suspecting the truth but avoiding it, and curiously, lacking insight into those particular issues, even though they may in other ways be honest. This characterisation of the self-deceived remains relevant.

Existentialism

Two centuries later, we find that self-deception is an issue for Existentialists. Self-deception undermines the Existentialist's primary normative aim of living authentically. Morality has become focused on personal responsibility rather than obedience to a doctrine. Self-deception, and morality are personal issues for the individual.

For Friedrich Nietzsche,[9] self-deception is the outcome of *ressentiment*. This derives from the attitudes of patience, humility and acquiescence of those without power. These attitudes mask their anger, envy and desire for revenge towards the powerful. As there is no action the powerless can take to bring relief, the base emotions transform and are revalued as self-deceptive moral superiority.[10]

The issue of authenticity is prominent in the work of influential, mid-20th Century, Existential philosopher Jean-Paul Sartre. Sartre's position is that one is radically responsible for one's existence. To realise this, and to live authentically, produces unbearable anxiety. Sartre argues that individuals engage in self-deception or *mauvaise foi* (bad faith) inevitably,[11] as a way of escaping this Existential 'angst'. Sartre's illustrative, paradigm examples are now iconic; the over-conscientious waiter and the flirt on an early date with her boyfriend. Importantly, Sartre stated the paradoxical nature of self-deception; that the self-deceived must somehow know the truth they are hiding from themselves and that they must be both deceiver and deceived. Further, because for Sartre there are no 'opaque' areas of consciousness, the project of self-deception must occur in consciousness. This is an early, explicit statement of the paradoxes of self-deception, which form the fulcrum of much of the later debate in Analytic Philosophy.

Self-deception in Analytic Philosophy

Taking up the issues of the paradoxes of self-deception, Analytical Philosophers have transcribed them into doxastic terms. They become: 'how can I believe p and ~p at the same time?' and 'how can I deceive myself, when I must be aware of the attempt?' Two major

[8] Butler, Bishop Joseph. *Sermon X*. 1726

[9] I am aware that there is some debate about whether Nietzsche is properly identified as an Existentialist. Resolution of this issue is beyond the scope of this thesis.

[10] Nietzsche, Friedrich. 'First Essay' *On the Genealogy of Morals.* Penguin 2014 and 'First Part. Section 19. On the Adder's Bite,' *Thus Spake Zarathrustra* in *The Viking Portable Nietzsche* Penguin. 1977. See also: Elgat, Guy. 'Slave Revolt, Deflated Self-Deception' *British Journal for the History of Philosophy,* 2015 23:3 524-544.

[11] Sartre, Jean-Paul. *Being and Nothingness: An Essay in Phenomenological Ontology.* Translated by Sarah Richmond. Routledge. 2018. 87-113.

philosophical approaches have on the one hand, attempted to explain these paradoxes, as in Intentionalism, and on the other, to develop a theory which avoids them, as Motivationalists do. Despite each side achieving ascendancy for a period, no definitive resolution has been found to the differences between the major theorists in either camp, resulting in an impasse between these theories. Such an impasse creates an obstacle to a straightforward examination of a possible relationship between self-deception and wilful ignorance. I will consider some of these theorists later in this section.

Eric Funkhouser's 2019 book *Self-Deception* provides a thoroughgoing review of the philosophical literature on self-deception, a critique of key concepts, an examination of different theoretical positions, and an analysis of the contemporary debate.[12] However, it is not possible to rely solely on this *tour de force* of self-deception theory, despite the fact that it provides a definitive overview. Nor is it the case that Funkhouser's analyses are always uncontroversial.

Intentionalism

Intentionalists model self-deception on interpersonal deception. Self-deception involves the doxastic paradoxes, mentioned above, and the task is to explain them.

Donald Davidson

Donald Davidson argues self-deception is intentional and that the dual belief is necessary for self-deception. It is awareness of the possible truth of a disagreeable proposition which causes the self-deception. Davidson sees self-deception as the cognitive parallel of *akrasia*, because the self-deceived breach the requirements for inductive belief formation. Where there is evidence for *p* and for ~*p*, they believe the proposition with the least evidence.[13]

Davidson admits that he cannot see how self-deception can happen unless there is some sort of partitioning of mind, but such partitioning does not need to be as drastic as the concept of a Freudian unconscious. Davidson is one of the three major philosophers I consider in detail later.

Motivationalist Theories

Deflationist or Motivationalist accounts argue for desire motivated belief-formation in self-deception. They thus avoid the classic paradoxes of self-deception.

Alfred R Mele

Alfred R Mele claims his 'deflationist' explanation of self-deception avoids both paradoxes.[14] Self-deception is a type of belief motivated by desire. It should not be explained using the model of interpersonal deception and there is no need for the 'dual belief' requirement. The self-deceived need only wrongly believe a proposition. As self-deception is a species of motivated desire, there is no need for the strong agency argument which sometimes appears in Intentionalist accounts. I will consider Mele's views in more detail later.

Twisted Self-Deception

Mele also claims that his Motivationalist approach can explain 'twisted self-deception' in which people deceive themselves into believing something that they do not want to be true.[15]

---

[12] Funkhouser, Eric. *Self-Deception* Routledge. 2019

[13] Davidson, Donald. *Problems of Rationality. Philosophical Essays Vol. 4.* Oxford University Press. 2004.

[14] Mele, Alfred R. *Self-Deception Unmasked* Princeton University Press. 2001

[15] Mele, Alfred R. Twisted self-deception, *Philosophical Psychology*, 12:2, 1999:117-137, DOI: 10.1080/095150899105837

Funkhouser acknowledges the phenomenon of twisted self-deception.[16] I will argue later that there are serious problems with this concept.

Importantly, Mele's theory stands in opposition to Intentionalism. Mele's theory was seen as the accepted doctrine for a considerable time, although more recently it has again been questioned. If the Intentionalists' criticisms are valid, there may be consequences for the soundness of other theories developed with too great a reliance on this type of Motivationalism.

Other approaches to self-deception.

Before venturing further into the issues raised by the major theories, other theoretical approaches should be noted. These, while attracting support, will not be helpful in resolving the issues addressed here. Some theories have developed consequential to the major theories, in an attempt to support, supplement or amend them. I will note but not be relying on these theories.

Theories consequent on Intentionalism and Motivationalism

Theories proposing partitioned mind or divided self

Approaches which *prima facie* offer the possibility of explaining the paradoxes in self-deception come from divided mind or divided self theories. These propose a variety of autonomous or semi-autonomous divisions, sub-systems and sub-personal aspects of the individual.

Below, I will consider two quite different approaches as examples of divided mind theories: Amelie Oskenberg Rorty, who seeks to normalise self-deception and Sigmund Freud, whose theory of the unconscious, some propose, may offer an explanation of self-deception.

Amelie Oskenberg Rorty

Amelie Oskenberg Rorty contends that the concept of a whole self is over-stated, and that it derives from the norms of theoretical schemas in Western thought.[17] Further, 'strategies of self-manipulation', including rationalisation, compartmentalisation, selective attention and internalisation are of the same category as self-deception. We often disguise, submerge or refuse to acknowledge a negative or ambivalent fact or attitude in favour of a positive in our lives, and this is like self-deception. Without some form of self-deception, Rorty claims, our work, our relationships and our goals would collapse. As well, our families, work colleagues and society generally support these processes.

Rorty further argues that many theories of depth psychology, cognitive psychology or sociology posit divisions, homunculi or sub-systems of the self. These sub-systems are not entirely integrated, may have relatively independent motivations (where motivations may be goals, desires or 'reasons'), and may be the source of the apparent deception in self-deception. In some cases, the sub-systems are internalisations of socially-determined and supported roles.

Rorty reflects Daniel Dennett[18] and others, who posit 'levels' of explanation for psychological and other phenomena, sometimes distinguished as the 'personal' and 'sub-personal' levels. Although there is a possibility of 'vertical' explanation between the personal and sub-personal level, it is not guaranteed. Explanations may not be integrated and it may be as if various instincts, intentions or 'homunculi' have a will, intention or desire alien to consciousness. Intelligible or understandable explanation mostly occurs 'horizontally', within a level, rather

[16] Funkhouser. *Self-Deception.* 110-112.
[17] Rorty, Amelie Oskenberg. 'User-friendly Self-deception' *Philosophy* 69 1994: 211-28
[18] Dennett, Daniel 'Intentional Systems Theory' Tufts University https://ase.tufts.edu/cogstud/dennett/papers/intentionalsystems.pdf

than 'vertically', between levels. Explanation at the level of neuroscience, for example, does not always capture lived conscious experience.

It is not necessary here to take issue with the notion of sub-personal, intentional systems. It is possible to grant Rorty's point that the self is much less integrated than we believe. That still leaves unaddressed any explanation of self-deception, as such explanation must occur 'horizontally'. As soon as a split or division of the self is proposed, especially a 'vertical' split, whereby one part believes p and the other ~p, this refers to a different type of phenomenon.

Sigmund Freud and the Unconscious

Depth psychological theories, which propose an unconscious aspect of the mind, are also examples of divided mind theories. These theories are said to rely on the sub-personal for explanation,[19] and therefore may assist in explaining self-deception. In such theories, the extent of opacity between the unconscious and consciousness is rarely complete and differs between theories.[20,21] In particular, Funkhouser identifies the theories of Freud as offering a 'tri-partite division of the mind' which could assist with explaining the paradoxical nature of self-deception.[22]

Freud proposes an unconscious which can be described as being largely 'opaque' to consciousness.[23] Its means of expression are symbolic: parapraxis,[24] dreams,[25] psychopathology[26] and a variety of defensive measures.[27] The unconscious is dynamic, as the instinctual urges towards expression are countered, 'censored' and repressed by contrary urges. These repressed phenomena may be ego-dystonic or originate in trauma, often experienced at pre-linguistic stages of development. None of these phenomena comprise self-deception. Freudian theory explains psychological disturbance or psychopathology with the involvement of the sub-personal level, but does not describe or account for self-deception.

Explanations from the unconscious do not assist us to understand or explain self-deception, and are at best peripheral to the current project. It is not necessary either to embrace Freudian theory, nor to reject it absolutely,[28] in order to consider self-deception.

Diachronic Partitioning Theories

Another type of partitioned mind theory is proposed by some Intentionalists, who argue that self-deception occurs over time, suggesting that the self is divided temporally.

---

[19] Gardner, Sebastian. 'Psychoanalysis and the personal/sub-personal distinction', *Philosophical Explorations*, 3:1, 2000 96-119, https://doi.org/10.1080/13869790008520983

[20] José Luis Bermúdez 'Personal and sub-personal; A difference without a distinction', *Philosophical Explorations*, 3:1, 2000 63-82. https://doi.org/10.1080/13869790008520981

[21] Jackson, Frank. 'Psychological explanation and implicit theory', *Philosophical Explorations*, 2000 3:1, 83-95. https://doi.org/10.1080/13869790008520982

[22] Funkhouse.r *Self-Deception.* 42ff

[23] Freud, S. (1915). 'The Unconscious.' *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIV (1914-1916): On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works*, 159-215

[24] Freud, S. *The Psychopathology of Everyday Life.* Penguin. 2002. See especially chapters I, II, V and VI.

[25] Freud, S. *The Interpretation of Dreams.* Penguin.1976.

[26] Freud, S. (1926). 'Inhibitions, Symptoms and Anxiety.' *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XX (1925-1926): An Autobiographical Study, Inhibitions, Symptoms and Anxiety, The Question of Lay Analysis and Other Works*, 75-176. Freud, S. *On Psychopathology.* Penguin. Harmondsworth. 1979. Freud, S. *The Wolfman and Other Cases* Penguin London 2002

[27] Freud, S. (1915). 'Repression.' *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIV (1914-1916): On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works*, 141-158

[28] As, for example, Sartre does. Note that conversely, it is not necessary to reject Freudian theory in order to appreciate Sartre's contribution to understanding self-deception.

Jose Luis Bermudez describes how this might happen.[29] He might, for example, decide on an initial project in general terms to enhance his career. This decision might cause him to sign on to a university committee, although when he does so, his initial intention is not clearly in his mind. He may have even forgotten it. The intention still operates as the background for his coming to this decision.

Sorenson argues that the events extending over time mean that there may be some overlap in the contradictory beliefs, but there is no reason to argue that the self-deceiver believes the contradictory beliefs simultaneously throughout that time. [30]

In these views, the intention is retained but the requirement for the self-deceived to contemporaneously hold contradictory beliefs is relinquished. Further, while these and similar situations describe dual beliefs, it is questionable whether they comprise self-deception.

Self-Deception as Failure of Self-Knowledge

One way of refining our understanding of self-deception is to try to more closely identify what we might be self-deceptive about. Some philosophers who have proposed that self-deception is a failure of self-knowledge seek to supplement Mele's Motivationalism by defining the object of self-deception.

Richard Holton

Richard Holton adds to Mele's sufficient conditions for self-deception a proposed necessary condition. [31] Holton takes issue with Mele's motivated belief view, arguing that mere motivated desire may produce a mistaken belief but will not necessarily result in self-deception. In order for motivation or desire to result in self-deception, it must be focused on a belief about oneself. The self-deceived make a mistake about their own beliefs about themselves. The subject matter of self-deception is a mistaken *de se* belief.

Jordi Fernández

Jordi Fernández, arguing against Intentionalism, also sees self-deception as a failure of self-knowledge.[32] Fernández focuses on the challenge to first-person authority in self-deception. In forming beliefs we look at behaviour and evidence as grounds for that belief. We can bypass this process when we form second-order beliefs or beliefs about our own beliefs. In subjectively determining what beliefs they have, the self-deceiver misuses their position of first-person authority to bypass both reason and evidence and thus makes a mistake about their beliefs. However, the mistake is not lost on others, who detect the self-deception.

Fernández, Holton, and others who propose that self-deceivers make a mistake in their second-order beliefs about their own beliefs have to contend with Funkhouser's pithy and most telling criticism: are the self-deceived so reflective?[33]

Eric Funkhouser

Funkhouser also argues that self-deception is a failure of self-knowledge and that the self-deceived have a false second-order belief about what they believe.[34] He avoids introducing an

[29] Bermudez, J. 'Self-deception, intentions and contradictory beliefs' *Analysis* 60 no.4. (2000)309-19
[30] Sorenson, R. 'Self-deception and scattered' events ." *Mind.* 94, no. 373 (1985): 64-69
[31] Holton, Richard. 'What is the Role of the Self in Self-Deception?' *Proceedings of the Aristotelian Society,* 2001
[32] Fernández, Jordi. 'Self-deception and self-knowledge' *Philosophical Studies* (2013) 162:379-400 DOI 10.1007/s11098-011-9771-9
[33] Funkouser. *Self-Deception.* 211
[34] Funkhouser, Eric 'Do the self-deceived get what they want?' *Pacific Philosophical Quarterly*. 86, no.3 (2005): 295-312.

over-reflective aspect into the concept of self-deception by proposing that the first order beliefs are opaque to consciousness, acquired without conscious reflection. Consciousness is unaware of the true belief. Thus the paradoxes are avoided.

Dana Nelkin

Dana Nelkin[35] notes that Motivationalism has been criticised for being overinclusive. As well, Motivationalism does not identify what is specific to self-deception. Nor does it differentiate self-deception from like phenomena such as wishful thinking. Nelkin proposes to add a necessary condition which counters these criticisms: a 'desire to believe' condition. The self-deceived have a desire to believe the self-deceptive belief and it is this that causes the biasing belief-forming strategies.

Self-Deception and Tension:

One defining characteristic of self-deception is tension, which may derive from the cognitive load caused by the self-deceptive project, encompassing evidence manipulation and manoeuvres to avoid the truth.

Robert Trivers notes this effect.[36] He points out that whatever evolutionary purpose self-deception serves, there is also a cost in terms of cognitive load and possible follow-on effects such as impaired immunity. By contrast, Fernández moves that tension so that it is external. The tension is between the self-deceiver and those observing the effects of the self-deception.[37]

Normative Aspects of self-deception

Overwhelmingly, the normative attitude towards self-deception has been one of negative judgement. Yet, there are dissenters.

Neil Levy seeks to follow up the normative implications of accepting a Motivationalist account of self-deception.[38] He argues that if self-deception is not intentional, then it is not appropriate to hold the self-deceived responsible for it.

Fernández argues that others find the self-deceiver's position objectionable because they detect the lie in self-deception and they sense the 'epistemic dishonesty'.[39] I would argue that epistemic dishonesty is a concern limited to some philosophers. Rather, as indicated above in relation to our paradigm example, others care because self-deception and its foreseeable consequences are disturbing.

Evolutionary Biology

Robert Trivers approaches self-deception from the perspective of evolutionary biology. [40] He argues that deception produces evolutionary advantage in nature. In human beings self-deception is an adaptation which functions to shore up our ability to deceive others, and so to achieve dominance. This view has been challenged from within the philosophy of evolutionary biology by D. S. Neil Van Leeuwen,[41] who argues that self-deception is not itself adaptive, but

[35] Nelkin, Dana K. 'Self-deception, motivation and the desire to believe.' *Pacific Philosophical Quarterly.* 83 (2002) 384-406.
[36] Trivers, Robert. *Deceit and Self-Deception: Fooling yourself the better to fool others.* Penguin. 2011.
[37] Fernández, Jordi. 'SD and S-Knowledge' p399ff
[38] Levy, Neil. 'Self-Deception and Moral Responsibility.' *Ratio.*17, no.3(2004):294-311.
[39] Fernández, Jordi. 'SD and S-Knowledge' 399
[40] Trivers, Robert. *Deceit and SD.*
[41] Van Leeuwen, D. S. Neil. *The Spandrels of Self-Deception: Prospects for a Biological Theory of a Mental Phenomenon.* Philosophical Psychology. 2007, 20:3, 329-348, DOI:10.1080/09515080701197148

is a by-product or 'spandrel' of other cognitive abilities which are selected for because they are strongly adaptive.

Trivers assumes a Motivationalist account of self-deception, and like those accounts, his theory is overinclusive. His illustrative examples of deception cover a fascinating range, from the camouflage-enhancing single-organ adaptation of stick insects to instances of 'questionable arguments used to convince the UN to condone modern warfare'.[42] Trivers grounds his arguments in a variety of experiments including bias-formation, subliminal priming and cognitive dissonance. Ultimately, Trivers attempts too much. He broadens the notion of deception and attenuates the concept of self-deception so that it seems to be a variety of assertive, if sometimes toxic, self-promotion.

What do the self-deceived want?

We return to Funkhouser, not to his *tour de force* on self-deception, but to his article, mentioned above, 'Do the self-deceived get what they want?'[43] What is all the alleged false belief, information manipulation and avoidance behaviour about? *What is it that the self-deceived want?* Funkhouser's answer is: *the self-deceived want to believe*. Further, this is the source of the lack of coincidence between proposed first and second order beliefs in the self-deceived. Don wants to believe he isn't bald, Emily wants to believe her husband is not having an affair, and so on. This is one article which considers and offers an answer to this curious and important question.

Focus of this thesis in relation to self-deception

As a result of the above overview of theoretical positions on self-deception, it is possible to eliminate of some types of explanations from consideration. These include explanations from psychoanalysis, those proposing psychological or temporal divisions of the self, evolutionary biology and those positing second-order 'beliefs about beliefs'. The reasons vary. Some approaches prove not to be about self-deception at all, while others involve serious conceptual difficulties. The theories I will be considering here address the question of how self-deception can occur in the consciousness of a relatable, functioning human being. There remains much to be considered within these parameters. As well, while I acknowledge a normative feature in self-deception, I will not be focusing that aspect.

Wilful Ignorance

As is the case with the term 'self-deception', the term 'wilful ignorance' is familiar, and has a colloquial use. By contrast with self-deception, there is a formal application of the term 'wilful ignorance' in criminal law, and this can provide a definitive basis for understanding. Examples drawn from case law may illustrate the concept.

In cases of wilful ignorance, a person claims ignorance when it is difficult to see how they could be ignorant. In criminal cases, where a defendant claims ignorance, it is a defence to the 'knowledge' aspect of a crime, provided it is not wilful. Thus in cases where the ignorance is deliberately maintained in order to avoid conviction, this defence is void. Douglas Husak focuses on this interpretation of wilful ignorance. [44]

The problem of how a defendant can be said to 'know' when they have deliberately avoided knowing is what motivates Alexander Sarch's consideration of the problem.[45] Sarch is also

[42] Trivers, Robert. *Deceit and SD* 257
[43] Funkhouser, E. 'Do the self-deceived get what they want?'
[44] Husak, Douglas N. *The Philosophy of Criminal Law: Selected Essays.* Oxford University Press, 2010
[45] Sarch. A *Criminally Ignorant.*

concerned that repeated negligence will attract less culpability than one act of wilful ignorance. This is a problem for legal philosophy, and Sarch, Yaffe,[46] and others have engaged in an energetic debate about the issue. These theorists, along with Douglas Husak,[47] take the view that wilful ignorance involves an understanding short of actual knowledge that an incriminating situation exists. However, the further debate, regarding comparative culpability between wilful ignorance and negligence is beyond the parameters of what can be considered here.

Others have moved the consideration of wilful ignorance outside of the courtroom and focused on our responsibilities as social and global citizens. Michelle Moody-Adams discusses the issue of 'affected ignorance'.[48] She argues that this occurs when an individual refuses to look at a situation involving pain or disadvantage for another, even though it is difficult to believe they did not know what was happening.

Some parallels between wilful ignorance and self-deception are now apparent. Both phenomena involve first-person claims of belief or ignorance which are questioned by others, both involve manipulation of one's awareness of evidence and both occur in otherwise relatable, functioning people.

A direct consideration of the relationship between self-deception and wilful ignorance is not straightforward, even if there were only one acceptable explanation of self-deception. The relationship has already been considered by a few philosophers, with incompatible results. As mentioned earlier, Kevin Lynch, who is strongly influenced by a Motivationalist perspective on self-deception, has argued that they are 'different psychological types'.[49] Jan Willem Wieland argues that a more appropriate way of approaching the issue is to look at the attitude of the individual involved. [50] In his view the distinguishing attitude, characterised by lack of concern, is culpable. As a consequence, Wieland does not make a distinction between self-deception and wilful ignorance.

While I am not alone in seeing the possibility of a relationship between the concepts of wilful ignorance and self-deception, no definitive determination on this issue has as yet been made. The question remains open for exploration. Hopefully, the considerations here will clarify some issues and indicate what might be needed to answer this question.

[46] Yaffe, Gideon Yaffe, Gideon J 'The Point of Mens Rea: The Case of Willful Ignorance.' *Criminal Law, and Philosophy*. 12 (2018): 19-44. See also: Zimmerman, Michael J. 'Recklessness, Willful Ignorance and Exculpation.' *Criminal Law and Philosophy*.12 (2018):327-329

[47] Husak, Douglas N. *Phil Crim Law*.

[48] Moody-Adams, Michelle. 'Culture, responsibility and affected ignorance.' *Ethics* 104, (1994) 291-309

[49] Lynch, Kevin. 'Self-deception and Willful Ignorance.' *Philosophical Studies.* 173, no. 2 (2016): 505-23.

[50] In particular, see: Wieland, Jan Willem. 'Willful ignorance.' *Ethical Theory, and Moral Practice.* 20, no. 1 (2017): 105-19.

Part Two

Self-deception

Introduction

The concept of self-deception has a long history in Western thought and is still identified as being a 'puzzling' phenomenon. While many theorists have assessed self-deception to be negative, no doubt there are instances of self-deception that are harmless.

While there is no definition of self-deception, there are many theories of self-deception. These theories are largely incompatible and perhaps, incommensurable. Yet, there is no possibility of answering the main question in this thesis unless the concept of self-deception can be adequately identified or described. Therefore, the aim in this section is to identify the theory of self-deception which has the most explanatory power. In order to do that, it is necessary to identify the characteristics of self-deception that a theory should explain. If this is achieved, it will be possible to articulate some 'desiderata' for a theory of self-deception.

Paradigm Examples

A first approach to identifying these characteristics is to consider paradigm examples of self-deception.

I have briefly considered the paradigm example of Emily, who deceives herself about whether her husband Daniel is having an affair. Here is her story:

Emily

Three years into her marriage to Daniel, Emily finds herself more often alone at home. Daniel's work was always demanding, but since his recent promotion he works late at the office a few nights a week. Even when he is at home, Daniel seems to be distracted. His 'work phone' is often ringing, sending him into another room for twenty or thirty minutes at a time. Now, he is frequently away for the week-end on work-sponsored conferences. He is accompanied by is new personal assistant, Lily, the recent college graduate who looks stunning in her Vogue knock-off wardrobe, which only a singleton could afford! Emily knows Daniel had a reputation as something of a 'player', but the day he proposed he told her that was over and he was now devoted to her. Emily believes him. Emily refuses to be jealous. Yet, when she meets her friends for coffee, she jokes that she sometimes wishes Daniel had chosen a less attractive personal assistant. She sees the concern on their faces when she tells them about Daniel's working hours and knows they are thinking Daniel is having an affair. She believes Daniel is faithful to her and tells her friends so.

Ryan

Ryan met Amy at the University orientation for new students and was immediately smitten. They shared a few casual words about how overwhelming first-year university felt. She smiled at him. Ryan worked up the courage to ask her out for coffee. She was polite, but said 'no'. Disappointed and a little confused, Ryan cannot stop thinking about her. It must have been fate that they met, he muses, because they are both majoring in the same subject. When he sees her, he tries to start up conversations with her. She is polite but always makes an excuse to hurry away. He asks for her phone number but she refuses. He tries to contact her on social media, but she ignores him. He believes that she must feel for him as he does for her. Why smile at him like that if she did not? He has heard that women 'play hard to get' and after considering her actions, he concludes that is what she is doing. She really loves him but is testing him. When he tells his undergraduate friends about her, they warn him to 'back off' from Amy. He does not want to seem like a stalker.

Liam

At 30 years old, Liam believes he is going to be a great writer. He always did well in English literature at school. He has started several novels but each time is plagued by 'writer's block' a few chapters into the work, so abandons the project. He supports himself with a series of part-time jobs well below his potential, which allow him time to work on his writing. He is part of a local aspiring writers critique group in his small town. The other members seem to be in awe of his work. He knows he is certainly a better writer than anyone else in this group. The group members tell him his ideas are original, even inspired. He doesn't want to crush his creativity by joining writers' societies or enrolling in courses to learn writing techniques. He doesn't research what publishers are looking for. He won't 'sell out' his art for profit. Besides, when he finishes his novel, his talent will be recognised by any publishers. If only he could get over his writer's block!

Evelyn

Evelyn believes she is healthy for her age. While her friends are slowing down, some even developing chronic conditions, she keeps active. She minds her grandchildren twice a week, volunteers at a local charity thrift shop, never misses her weekly art classes and goes on long walks or hikes with a group. She occasionally has what she calls 'turns', which are short-lived episodes of chest pains accompanied by difficulty breathing. Recently, she had one of these 'turns' while she was at a family gathering. Her daughter urged her to go to the doctor. 'No need to worry,' Evelyn replied when she caught her breath. 'It's just part of getting older. Besides people rely on me. I'm too busy to go to the doctor.'

Each self-deceiver holds their self-deceptive belief against the weight of evidence. In each case, the evidence points to an undesirable proposition: Emily's husband Daniel is unfaithful, Amy really is not interested in Ryan, Liam is not likely to become a great writer and Evelyn is developing a chronic heart condition.

In three of the four cases, the first-person authority of the self-deceiver is disbelieved by those around them. Emily's friends, Ryan's undergraduate colleagues, and Evelyn's daughter challenge the self-deceptive belief. It is not clear in Liam's case who might question his belief that he will become a great writer, but we might speculate that anyone outside of limited circle of the local critique group would see little evidence to justify it.

Even though they believe against the evidence, in no case is evidence for the self-deceptive belief entirely lacking, nor without an alternative, plausible explanation. Emily, we saw, relied on Daniel's recent promotion. As well there is his previous promise and three years of marriage.

Ryan relies on his misinterpretation of Amy's politeness and the information that women 'play hard to get'. Liam points to his writer's block, while his self-deceptive belief is confirmed by the local critique group. Evelyn has the evidence of the lifestyle changes in her contemporaries, so tells herself that like them, she is 'just getting older'.

Our self-deceivers may also appear to take efforts to maintain their self-deceptive belief. This could be seeking out evidence for the belief or avoiding contrary evidence. Liam will not seek feedback from an editor or publisher. Evelyn avoids going to a doctor. Ryan puts a lot of weight on Amy's smile and the fact that they are in the same class. Emily has a promise, three years of marriage and her own value, a refusal to be jealous. These efforts to maintain their self-deception could contribute to the tension often observed in self-deceivers.

All of our self-deceivers are otherwise highly functioning people. They are self-supporting, have commitments to jobs, studies or voluntary activities, and have good relationships with family, colleagues, friends or people with similar interests. Their self-deceptive belief is one aberration in an otherwise understandable and relatable life.

In relation to the task of finding the theory of self-deception with the most explanatory power, it is now possible to identify some features of self-deception which the theory must address. That is, either the theory explains these features, or it shows why an explanation is not necessary. These features are:

- The self-deceptive belief is believed against the evidence, while an undesirable, conflicting proposition is more likely to be true.
- The self-deceiver's first-person authority is questioned by others.
- Self-deceivers take some effort to maintain the self-deceptive belief, usually involving manipulation or avoidance of evidence. This may contribute to tension.
- The selectivity question regarding the selection of the issue about which to deceive oneself requires serious consideration. This is especially so because self-deception is an aberration in an otherwise highly functioning and relatable person.

These issues identify some of the puzzling aspects of self-deception which have been considered in the literature. I have examined the above, fairly standard, paradigm examples in order to tease out the key features of self-deception. This has resulted in a list, which itself needs further explanation. In the next section I will consider what theorists have said about these issues, in order to understand these features in more depth and refine these concepts.

The key features of self-deception

I have identified some distinctive features of self-deception. In this chapter, I will consider these characteristics in more detail, in order to more clearly understand what is meant by them and to refine the manner in which they are used in this project.

The puzzle of self-deception

The puzzle of the concept of self-deception is inherent in the term, which is grammatically reflexive. It may derive from the practice of modelling self-deception on interpersonal deception.

Interpersonal deception has the basic structure of the lie. It occurs between people. It may occur within private relationships, in public relationships and in business relationships. When deception occurs in business relationships it is fraud. When it occurs in court testimony it is perjury. Laws protect us against these deceptions and their effects, identifying them as criminal, and punishing offenders with fines, and imprisonment. Thus, deception can be a serious matter.

In interpersonal deception the deceiver believes one thing but persuades the deceived to believe another. Yet it is difficult to see how this can happen in a single mind. Self-deception seems to suggest puzzles, even paradoxes.

The paradoxes of self-deception

Jean-Paul Sartre stated the problem of the paradoxes concisely:

> *I must know in my capacity as the deceiver the truth which is hidden from me as the one deceived. Better yet, I must know the truth very exactly in order to conceal it more carefully. This…in the unitary structure of a single project.*[51]

Analytical philosophers have approached the puzzle by focusing on the two paradoxes of belief which emerge from this statement. They are:

- *The static paradox*: how can I believe two contradictory beliefs at the same time?
- *The dynamic paradox*: how can I persuade myself to believe something I do not believe, when all that is needed for the self-deceptive project to fail is for me to be aware of it?

The Static Paradox

The 'static paradox' is central to the philosophical debate about self-deception. The 'static paradox' of self deception, is that the self-deceived must hold contradictory beliefs and it is difficult to see how this can occur.

In attempting an explanation, Donald Davidson and Eric Funkhouser have each insisted that the self-deceived do not entertain contradictory beliefs about an objective situation.[52] A logical contradiction in the form 'p and ~p' is *a priori* an impossibility. This does not describe the static paradox.

It is the beliefs held by the individual that are contradictory. We attribute to the self-deceived a contradiction in the content of their beliefs. The static paradox is represented as below:

- A believes that p; and
- A believes that ~ p

[51] Sartre, Jean-Paul. *Being and Nothingness*, 49
[52] Funkhouser. *Self-Deception* p31; Davidson. *'Problems'* 199ff

Donald Davidson offers this curious example which Erik Funkhouser echoes[53]:

- Don believes he is bald; and
- Don believes he is not bald.

While this explanation avoids the logical impossibility, it presents us with a psychological problem. It seems impossible that a person could hold two such contradictory beliefs, consciously, fully 'front of mind', with equal force and sustain that position for any length of time.

Other contradictory beliefs

In practice, we do hold contradictory, or at least conflicting beliefs, without too much difficulty. In these cases, while it may be a matter of inconsistency, it is not self-deception. I may hold conflicting, even contradictory beliefs when:

1. I can live with the conflict. I may believe both that wine is good for the health and that alcoholic drinks are not good for the health. This conflict will not stop me having wine with my dinner.
2. I am mistaken. I may believe that mindfulness techniques are calming but Buddhist meditation techniques are not.[54] As these are the same techniques, I am mistaken.
3. I don't care. I listen to the Hollywood gossip section on morning television but simply shrug when my work colleagues call me a 'Hollywood gossip tragic'. I don't really care about Hollywood gossip and I don't even care if people think I do.

If these examples do not shed light on the static paradox in self-deception then what is it that makes self-deception a puzzle? Perhaps there is something about the notion of 'belief' which causes the puzzle.

The nature of belief

Philosophers use the term 'belief' in diverse ways. Some point out that our beliefs derive from and depend upon a network of supporting beliefs.[55] Philosophers tend to associate conscious belief-formation with both rationality and empirical reality or evidence.[56] At the extreme, 'belief' that is sufficiently justified and true may be said to be 'knowledge'. The tradition of defining knowledge as 'justified true belief' is longstanding in philosophy. [57] This is the domain of epistemology. The statement 'I believe that…', short of the certainty of knowledge, implies introspection and some reflection. This is the domain of doxastic philosophy.

However, to have a belief does not mean that one can 'believe' what one chooses. To 'believe' without any relationship to evidence is to engage in wishful thinking or fantasy.[58] If one brings this into everyday experience in any sustained manner, it is likely to be a full-blown delusion.[59]

---

[53] Davidson. *'Problems'* 199. Funkhouser. *Self-Deception.* 32 and 'Do the self-deceived get what they want?'
[54] For an overview see:
https://en.wikipedia.org/wiki/Mindfulness#Popularization,_%22mindfulness_movement%22
[55] Harman, Gilbert. *Change in View. Principles of Reasoning* The MIT Press. 1986. Davidson, Donald. *Subjective, Intersubjective, Objective. Philosophical Essays. Vol 3*. Oxford University Press. 2001
[56] See also discussion in Funkhouser. *Self-Deception* 31ff; Davidson. *'Problems'* 189ff; Harman, Gilbert. 'Notes on Practical Reasoning' *Cogency* Vol 3 No 1 Winter (2011) 127-145
[57] More recently with the exceptions provided by so-called 'Gettier examples'.
[58] Funkhouser. *Self-Deception* 35ff Davidson *'Problems'* 178
[59] See for example Mele, Alfred R. 'Self-deception and delusions' *European Journal of Analytic Philosophy* 2 No. 1 (2006)

As well as being a reference to the product of reflection, the term 'belief' is used in relation to the effects of perception. I am said to 'believe' the coffee cup is on the table in front of me, the light will go on when I flip the switch, I will catch the 8am bus if I am at the bus stop by 7.55am.[60] These are beliefs which get me though my life and I do not have to reflect about them. Indeed, they seem only to gain the categorisation as 'belief' by an unnatural abstraction. In my daily life, I pour coffee into the cup, flip on the light and stand at the bus stop at 7.55am with neither reflection nor introspection. Sartre has called this consciousness 'pre-reflective' or 'non-positional' consciousness and has contrasted it to reflective consciousness.[61] More recently, Kahnemann,[62] from a different tradition, has distinguished two systems of awareness along broadly similar lines. The conclusions to be drawn from these theorists, despite the diverging background, is that much of our lives are lived, much of our engagement with the world happens, without taking the effort to reflect and without introspection.

Kahnemann takes things further. He shows that we bring a vast amount of information to any situation. Looking at a photograph of a woman's face, we immediately see and understand a variety of things about her. Significantly, we read her emotions and know immediately that she is angry. We anticipate she is about to say something in an angry or aggressive tone. That is, our understanding projects into the future. We do this without introspection or reflection. Kahnemann calls this 'System 1',[63] and this system is responsible for getting us through much of our lives.

When we do something requiring introspection or reflection, Kahnemann shows, this involves a different 'system', which he designates 'System 2'. When System 2 is engaged on a task, it takes a lot of energy and often involves stopping System 1 activities.

At the same time, it is not necessarily the case that System 2 is concerned with high-level logic or abstraction. System 2 requires attention and energy. It may concern rationality and it may concern reflection, a trawling through of memory or other complicated mental activity.

Further, the 'Systems' are not isolated from each other. The effort taken by System 2 to learn a new skill, for example, transfers to System 1 over time.[64]

These considerations do not necessarily provide answers to the static paradox nor to the phenomenon of self-deception. They do show that the puzzle of the beliefs in the static paradox may arise in part because of the disparate ways in which the term 'belief' may be used or the failure to explicitly identify the type of beliefs involved.

We cannot as yet dismiss the issue raised by the paradox. It may be that these considerations attenuate the puzzle of the static paradox. We know that the self-deceived appear to have contradictory beliefs. The self-deceived do not decide to live with these beliefs, they are not mistaken, they are not engaging in fantasy or wishful thinking and, importantly, these apparent beliefs are not likely to be about an issue they don't care about.

It is not clear how much reflection is involved, nor at what stage of the self-deceptive project reflection is required or diminished. There is evidence to be considered. If the self-deceived considered the evidence and opted for the proposition with the most evidence, there would be no problem. In self-deception however, the self-deceived apparently declare their self-deceptive belief against the evidence.

---

[60] Fernández, Jordi. 'SD and Self-Knowledge.' 389

[61] Sartre. *'B & N'* 8-16

[62] Kahneman, Daniel. *Thinking, Fast and Slow* Farrar, Straus and Giroux. New York.2011.

[63] This is a short-hand name awareness which does not require reflection

[64] Kahneman is one of a number of theorists who have made a distinction between these modes of consciousness. See also: Harman, Gilbert. *Change in View* Chapter 1.

The Dynamic Paradox

The dynamic paradox in self-deception refers to the psychological processes required to deceive oneself.[65] It is modeled on interpersonal deception, where the deceiver believes p while taking steps to bring it about that the deceived believes ~p and those steps succeed. All that is necessary for the attempt to fail is that the deceived becomes aware of the intention to deceive.

Therefore, the dynamic paradox can be stated as: how can I succeed in my intention to deceive myself when awareness of the deceptive intention is all that is necessary for the deception to fail? Further, how could I not be aware of my own intention? Self-deception is a specific type of attempt to have oneself believe what one does not believe. Therefore, it is useful to distinguish it from similar phenomena.

There are many psychological mind-manipulation processes which involve having the subject intentionally change a belief from p to ~p, and we would not portray these processes as being self-deceptive. Neil Levy,[66] amongst others, identifies these processes as 'clear-eyed' change processes and distinguishes them from self-deception. By contrast, for Amelie Oskenberg Rorty,[67] these processes are similar to or related to self-deception, involving as they do change from one belief to another. Pascal's famous advice to non-believers in Christianity appears to be more a clear-eyed change process than self-deception. [68]

Goal-setting can be reduced to p and ~p beliefs. In goal-setting, I intend to bring about a change. That means I am aware of where I start and where I hope to end. I do this in full consciousness but necessarily, the p and ~p beliefs are separated in time. Goal-setting is not self-deception.

There exists a cluster of psychological therapies which aim at replacing a negative, ~p belief about oneself or the world with a positive, often opposite p belief. Early therapies appeared in the late 19th century. Prominent 20th century schools include Albert Ellis' Rational Emotive Therapy,[69] which also inspired Aaron Beck's Cognitive Behaviour Therapy and their iterations. More recent examples include Happiness Therapy, therapies promoting gratitude and some aspects of Mindfulness Therapy.

Whether it is a matter of keeping a formal journal, introspection or regular daily thought-changing practices, all of these theories have common themes relevant to this consideration of changing our beliefs. Each of these practices requires that the subject identifies a negative belief and through intentional action brings it about that she acquires another, positive belief. These practices are aimed at changing attitudes and at transformation. The idea is that insight and self-knowledge, plus practice, will lead to change. While there may be a time during the process when the negative belief and the replacement belief are equally in the subject's mind, that is temporary. The negative belief at the beginning of the process is a single belief as is the positive belief at the end of the process and it is in precisely this that they differ from self-deception.

What is suggested by the above is that self-deception seems to involve some effort on the part of the self-deceived to arrive at the self-deceptive belief, against the belief more likely to be true. By definition, unlike the person engaging in clear-eyed change techniques, the self-deceived are seeking neither change nor insight.

---

[65] Funkhouser. *Self-Deception* 34ff
[66] Levy, Neil. 'Self-Deception and Moral Responsibility' *Ratio* 17 no 3 (2004) 298
[67] Oksenberg Rorty, Amélie. 'User-Friendly Self-Deception.' *Journal of Philosophy.* 69, no.268 (1994):211-28.
[68] Pascal *Pensees* Dutton and Co. New York. 1958 Kindle Edition from Project Gutenberg http://www.gutenberg.org/ebooks/18269
[69] Ellis, Albert and Robert Allan Harper. *A New Guide to Rational Living* Engelwood Cliffs, NJ:Prentice Hall 1975.

Perhaps the self-deceived *cannot* be seeking insight. Insight would destroy their self-deception. Without such insight they are able to retain their self-deception, possibly irrationally, or against the evidence. As well, it seems that the self-deceived are not seeking change. This raises the question whether in self-deception, the effort is expended in maintaining the self-deception, rather than at eliminating it. This view is broadly consistent with that of Eric Funkhouser, who concludes that the aim of self-deception, what the self-deceived want, is to believe the self-deceptive belief.[70]

First Person Authority

First person authority refers to the practice of attributing veracity and infallibility to a person's reports about their own sensations, intentions, desires and beliefs. It is a particular aspect of the more general philosophical debate about self-knowledge,[71] and shares with it some of the broader issues which have arisen in relation to the issue of self-knowledge.

When a person honestly says 'I have a headache' or 'I am sad' or 'I believe UFOs exist', we accept these statements as true statements about their feelings or beliefs. The speaker, whilever they are speaking in the first person singular, has a presumed epistemic privilege concerning these statements. Yet, when we detect self-deception in others, these first person statements are disbelieved.

There are reasons for questioning this default position of attributing epistemic authority to first person singular accounts of beliefs, intentions and feelings. Here it is worthwhile considering some of those reasons. As they are mostly to do with theories of knowledge, it is useful to be aware of them, so that these concerns can be distinguished from the immediate challenge to first-person authority which occurs in self-deception.

While self-knowledge is a perennial concern of philosophy, the focus of much of the debate can be said to have originated with Descartes' introspective method of inquiry.[72] His method implies that a person's mental states comprise an inner world, and that this world is substantively different to the external world. On this view, a person's 'inner world' is directly knowable to that person due to the immediate proximity of the subject to the object of introspection (their own mind). [73] This explains the claimed epistemic infallibility of 'I' statements about consciousness. Descartes' model of self-knowledge, that it is analogous to perception, is 'observational' and this in turn raises a number of concerns.

In the early 20th century, Behaviourism presented a challenge to the Cartesian view of mind and self-knowledge,[74] by apparently rendering redundant the possibility of autonomous agency and the mind-body dualism. On this view, our relationship to our own minds, beliefs, feelings and intentions originates externally. Currently, there is still a strong 'externalist' position on self-knowledge, informed by social sciences, which argues that many of our beliefs are derived from interaction with our social situation.[75] These views seemingly diminish the claims of first person authority.

Gilbert Ryle represents an attempt to address the conflict in theories regarding first-person authority. Ryle puts forward what has been called an 'expressive account' of first person

[70] Eric Funkhouser. 'Do the Self-Deceived Get What They Want?' 305
[71] This use of the term 'self-knowledge' is distinct from another use of the term which refers to knowledge of 'the self'.
[72] Descartes, Rene. *Discourse on Method* Dover. New York. 2003
[73] Wright, Crispin. 'Self-Knowledge: the Wittgenstein Legacy' in Wright, Crispin, Barry C Smith and Cynthia Macdonald (eds). *Knowing Our Own Minds* Clarendon Press, Oxford, 2001 13-47
[74] That is particularly the Behaviourism of B F Skinner and B Watson..
[75] Martin, M G F. 'An Eye Directed Outward' *Know Own Minds* 99-123

authority.[76] Ryle argues against Cartesianism, the 'dual' knowledge theory, claiming it creates other problems including questions about the nature of self-knowledge. However, Ryle is not a Behaviourist. He does not deny the privacy of mental processes, nor that we are in the best position to know our own minds. Ryle focuses on the communication of spoken 'avowals' of one's own beliefs, feelings and intentions. It is through the avowal that others understand the subject's mind. Depression is avowed with consistent body language, voice tone and in an appropriate setting. A query whether this is the way to London avows that the questioner wants to go to London, and so forth. These are assessed, not as being true or false but as being sincere or insincere. This is what makes them unassailable.

Ludwig Wittgenstein's account of first person authority has also been called an 'expressive account'.[77] Wittgenstein argues that the rules of grammar express rules for meaningful language. This precludes grammar applying to what only the subject can know. Focusing on the grammatical function of first person avowals, Wittgenstein states that when I claim to 'know' what I am thinking, this is not correct usage. First person avowals are verbal expressions which are akin to natural expressions of those same states such as grimaces or grins.[78] Thus the epistemic question does not arise, as their incorrigibility and infallibility is in their immediate expression.

First Person Authority and Rational Agency

Another line of discussion of this issue focuses on the relationship between self-knowledge and the subject's role as responsible agent.

Inspired by P E Strawson, Akeel Bilgrami argues that we consider that a subject's first person statements about beliefs and intentions are accurate just to the extent that we consider her to be a rational agent.[79] Thus, we take her actual mental states to be in accord with her first person statements about them. It is not a causal relationship, but relies on the ascription of rationality.

Tyler Burge argues that only a being with the possession of the concept of the first person pronoun is in a position to review and either accept or amend their own attitudes and rational beliefs. Thus, Burge acknowledges that first person authority, being subject to amendment, can change. [80]

Donald Davidson acknowledges the asymmetry between the acceptance of a first person propositional statement and the same statement in the third person.[81] Davidson elsewhere explains why it is necessary to assume that in intersubjective situations we are dealing with rational subjects.[82] It should also be noted that Davidson acknowledges limited fallibility in first person authority, as for example when unconscious phenomena contradict conscious

[76] Ryle, Gilbert. *The Concept of Mind.* Penguin. 2000. See also:Harre, Rom. 'Gilbert Ryle; The Concept of Mind' in Shand, John (ed) *Central Works of Philosophy Vol 4. The 20th Century: Moore to Popper* Acumen: Chesham U.K. 2006. Deutscher, Max. 'Some recollections of Ryle and remarks on his notion of negative action' *Australasian Journal of Philosophy.* 60 No 3 (1982) 254-264. DOI:10.1080/00048408212340671. Deutscher, Max. 'Infallibility, Hypocracy, Artless Talk' *work in progress* 2021.

[77] Child, William. 'Wittgenstein on the First Person' Kuusela, Oskari and McGinn, Marie (eds) *The Oxford Handbook of Wittgenstein* Oxford 2011 DOI: 10.1093/oxfordhb/9780199287505.003.0018

[78] Bar-On, Dorit and Long, C L 'Avowals and First Person Privilege' *Philosophy and Phenomenological Research* (2001) Vol 62 No 2. 311-335

[79] Bilgrami, Akeel. 'Self-knowledge and Resentment'. *Know Own Minds* 207-242

[80] Burge, Tyler. 'Reason and the First Person'. *Know Own Minds* 243-270

[81] Davidson, Donald. 'First Person Authority'. *Subjective, Intersubjective, Objective. Philosophical Essays Vol 3.* Oxford University Press. 2001. DOI: 10.1093/0198237537.001.0001. 3-14

[82] Davidson, Donald. See: *Problems*. 189ff.

propositional attitudes.[83] In relation to first person authority, Davidson states this is a necessary feature of language, which is essentially interpersonal. In attempting to communicate, as a speaker, one's aim is to make oneself interpretable to the other. The speaker may succeed more or less. However, in speaking, he must know what he means in order to try to be interpretable. Therefore there is a presumption that the speaker usually knows what he means and 'if he knows that he holds a sentence true he knows what he believes'.[84]

Second Order Beliefs

There is considerable debate which questions the claimed infallibility of first person authority by introducing the notion of second-order beliefs; that is, my beliefs about my beliefs. These considerations are particularly relevant to one type of approach to self-deception, which proposes that in self-deception the self-deceived make a mistake in their second-order beliefs about their beliefs.[85]

If self-knowledge involves knowledge of my beliefs, then the question arises concerning what the basis of such knowledge may be. Christopher Peacocke is one theorist who argues that a conscious belief is the reason for a second-order judgement that I have that belief. There is no need to resort to explanations from inference, nor from observation.[86]

It is possible to question the validity and authority of such self-ascriptions of belief and Michael Martin does so, referencing the inter-action between beliefs about the world and self-ascription.[87]

Cynthia Macdonald argues that unlike perception, which involves object, intermediary and subject, self-knowledge involves only two factors. [88] These are the content-based state, and the second-order belief state which validates it. Self-ascription of a particular state would be a criterion for having that state.

While these theories have merits, there is a concern that they may introduce a revised and possibly more complex version of the 'Cartesianism-Empiricism' duality. It had been argued that this duality had been disposed of in earlier explanations.

One philosopher who argues for a 'multiple criteria' account of self-knowledge is Elizabeth Fricker.[89] Special access accounts of self-knowledge need not be restricted to Cartesian accounts, but might include functionalist accounts. Further, both grammar and empirical factors 'collaborate' to hold our mental state concepts together, and also in forming a basis for our judgements about our own mental states.

For the current project, it is necessary only to note the various interpretations of, and objections to, the concept of first person authority. I take the debate to have been moved considerably by Ryle and Wittgenstein, in avoiding the problems of reductionist empiricism and Cartesian Idealism. As well, I note with reservations, Davidson's not-incompatible explanation. Those

---

[83] Davidson, Donald. 'First Person Authority.' 4. Note this particular exception to infallibility is controversial. Psychoanalysis, while confronting, does not involve the imposition of the analyst's interpretation on the analysand. Rather the analysand is supported in developing a new perspective or understanding A change in propositional belief consequent on psychoanalysis is similar to a change in a propositional belief following a re-examination of a situation involving newly discovered factors. It does not necessarily negate first person authority at any particular time along the process.

[84] Davidson, Donald. 'First Person Authority.' 10

[85] For example theories proposed by Holton and Fernandez, mentioned above.

[86] Peacocke, Christopher. 'Conscious Attitudes, Attention, Self-Knowledge'. *Know Own Minds.* 63-98

[87] Martin, M G F. 'An Eye Directed Outwards'. *Know Own Mind.s* 99-122

[88] Macdonald, Cynthia. 'Externalism and Authoritative Self-Knowledge' *Know Own Minds.*123-154

[89] Fricker, Elizabeth. 'Self-knowledge: Special Access versus Artefact of Grammar – A Dichotomy Rejected.' *Know Own Minds.*155-206

reservations focus on his comments about the unconscious, which may conflict with the requirement of rationality for the attribution of first person authority. Theories promoting second-order beliefs may, I would argue, be over-explained and misconceived. They seem to usher in new versions of problems already dispatched.

Although there are these differences, the concept of first person authority is clear enough so that a challenge to it is understandable, and able to provide one of the desiderata to be considered when evaluating theories of self-deception. That is, how the particular theory of self-deception either accounts for, or realistically dismisses, the challenge to first person authority which occurs in self-deception.

Selectivity issue

A key issue to be considered is the 'selectivity issue', which has been described above as answering the question: 'Why does the individual deceive themselves about this issue and not another?'

Bermúdez states that self-deception is 'paradigmatically selective'.[90] Bermúdez[91] sees the 'selectivity problem' as being an issue for Motivationalist positions, including that of Mele. In Bermúdez' view this problem comprises a 'fatal defect' for deflationary theories, unless supplemented by an Intentionalist explanation. Otherwise, there is nothing to explain the connection between the desire that p and self-deception.

Mele[92] counters these Intentionalist challenges with an argument that if they are valid, then the Intentionalist position also involves a selectivity issue.[93] In two similar or connected situations, where one succeeds in the self-deceptive project and the other does not, what accounts for the difference? Mele argues that the Intentionalist, who requires an intention to self-deceive, cannot answer this.

As it is Intentionalists who insist on an answer to the 'selectivity problem', it may seem to be unfairly biased to include this as a criterion in assessing the explanatory power of theories of self-deception. However, Mele's answer to the challenge suggests that this is an issue which may apply to Intentionalist and other theories. Further, as indicated earlier, in considering whether a theory of self-deception meets the desiderata, the theory under consideration may either explain the issue or render the need for explanation unnecessary.

Tension

I have mentioned that tension is an aspect of self-deception. I am using this term in a broad sense, to indicate not only possible overt signs of tension, but also specific causes of tension. These are the measures the self-deceived take in order to maintain their self-deception. They include misunderstanding or misconstrual of evidence, and avoidance. Writers on self-deception, from Bishop Butler to Sartre to Funkhouser have detailed these manoeuvres which the self-deceived take to maintain their position.

Further Question: What is the purpose of self-deception?

A further essential question, which may not be teased out of an analysis of the paradigm examples, but which might be asked is a teleological question. This question goes beyond the

---

[90] Bermúdez, José Luis. 'Self-deception, intentions and contradictory beliefs' *Analysis* Vol. 60, No 4 (2000) 309-319. https://www.jstor.org/stable/3329185

[91] Bermúdez, José Luis. 'SD, Intentions' 317

[92] Mele, Alfred R. 'Self-deception and Selectivity' *Philos Stud* (2020) 177: 2697-2711. https://doi.org/10.1007/s11098-019-01334-9

[93] Mele, Alfred R. 'SD and Selectivity' 2702-2703

'why' of self-deception or what motivates the self-deceptive project. It asks: what purpose, if any, does self-deception serve? What is all the effort to believe against the evidence, to cause oneself tension, to have one's first-person authority questioned, about? What does it achieve? What is self-deception *for?*

In the introduction to this section, I noted that if we are to consider what, if any, is the relationship between self-deception and wilful ignorance, it was necessary to have a clear idea of each concept. Although it seems to us that we intuitively know what self-deception is, there is no accepted definition and a number of incompatible theories. Therefore, to answer the original question it is necessary to choose a theory of self-deception, and it was decided that the theory with the most explanatory power would be the best theory of self-deception for this project.

The next problem was how to decide which theory had the most explanatory power? In order to decide this, four paradigm examples were analysed and this elicited a number of characteristics of self-deception. The theory with the most explanatory power will be the one that best explains those characteristics of self-deception. In this chapter I attempted to expand and refine the understanding of those characteristics as used in this thesis. As well, I have added two questions: 'why?' and 'what is its purpose?' It seems to me essential that these questions are answered by any theory of self-deception.

In the next chapter, I will provide an overview of the main approaches to the issue of self-deception in the philosophical literature, and then, a critical summary of the methodology of philosophical argument in relation to this subject.

Approaches to Self Deception

Approach to Theory

The structure of arguments explaining self-deception.

As is evident from the above review of the different approaches to the topic, arguments regarding self-deception tend to be inductive rather than deductive. As there is no agreed definition, explanations begin with what is said to be 'known' about self-deception. Without a definition, what is 'known' remains anecdotal, and is presented in the form of 'paradigm examples'.

Argument begins from paradigm examples and evolves towards a theory. There exists a plethora of paradigm examples, many of which are proposed on an apparent assumption that they will be recognised as being self-deception. A further problem for this project is that some paradigm examples are said to exemplify both self-deception and wilful ignorance. A non-exhaustive, but indicative, list of paradigm examples is attached at *Appendix A*.

In some cases, the claim to explanation is modified to a claim to explain only a limited set of cases of self-deception or a particular type of self-deception.[94] This may have the effect of forestalling a type of criticism based on the production of more examples alleged to counter the proposed paradigm examples. There may be a cost in terms of limiting explanatory power.

The next common step in the argument is to construct proposed 'sufficient conditions' for self-deception. Sometimes 'necessary and sufficient conditions' are constructed. Often these conditions are overly complex, while the reasoning supporting them is neither incisive nor decisive. There is seldom agreement regarding the 'conditions'. A table showing the lists of these conditions is attached at *Appendix B*.

Rarely does a theorist specifically relate the proposed sufficient and/or necessary conditions back to the paradigm examples. The reader is left to make these connections. However, such relating back would provide helpful explanation.

Each step along the way in this way of theorising is open to interpretation, objection, and counter-examples. Problems inherent in inductive arguments concern the type of reasoning involved and standard of proof. Alfred Mele has attempted to bring empiricism into his work, by relying on the results of his informal questionnaires to students and formal experiments conducted by others.[95] However, his work also utilises the approach indicated above.

This method of conducting philosophical explanation has resulted in an embarrassing deadlock between theories of Intentionalism and Motivationalism. There may be many differences, but the unresolvable one is that Motivationalists claim that self-deception can be explained without recourse to the paradoxes, while Intentionalists insist that self-deception involves avoiding a proposition which is the source of the paradox.

It is possible to make facts or a depiction of a situation, such as a paradigm example, fit more than one proposed explanation. That is, the same paradigm example may tend to confirm or corroborate two different theories. A further problem that derives from the abovementioned stalemate between theories is that once a particular theory is stated, it is very difficult to find something that might disconfirm or even falsify it. If the response to an objection to a theory is merely that the theory is correct, then we may be dealing with dogma, not philosophy.

---

[94] For example Mele's 'garden variety' self-deception and Fernández' claim to explain 'some types of self-deception' in 'SD and Self-Knowledge.' p379

[95] Mele, Alfred R. 'Recent work on self-deception' *American Philosophical Quarterly* 24 no. 1 (1987) 1-17

Additional Criteria for Assessing Theories of Self-Deception

In addition to the abovementioned criteria for assessing the explanatory power of theories of self-deception, we can add criteria for assessing the type of theory we are dealing with, aiming to avoid dogma. Thus we might add the following criteria:

As well the theory needs to be amenable to:

- Confirmation; and
- Disconfirmation.

There are many desirable norms of explanation, including but not limited to simplicity and clarity, which contribute to elegance in explanation. These qualities may enhance a theory, but they are not strictly necessary for understanding it.

The assessment of theories of self-deception is a different project from understanding self-deception *per se.* Clearly, the self-deceiver is not engaging in the abstraction and reduction involved in theorising but rather is engaged in living their life. Whatever type of rationality or irrationality may be involved in self-deception, it is not to be understood by the same principles of analysis that I will apply to theories of self-deception.

In the next section, I will consider in more depth representatives of each of the three major theories of self-deception, with a view to identifying which has the most explanatory power. This will provide a model of self-deception which can be compared to the concept of wilful ignorance. This will ultimately allow us to answer the question posed in this thesis.

Comparison of major theories of self-deception

The overview of theories of self-deception considered in the previous chapter identified three major approaches to self-deception, and a philosopher who is most representative of each position. I will now consider each of these in more depth, with a view to determining which theory has the most explanatory power, as assessed with reference to the previously identified desiderata for a theory of self-deception. The philosophers and theories under consideration here are:

- Existentialism represented by Jean-Paul Sartre
- Intentionalism represented by Donald Davidson
- Motivationalism represented by Alfred R Mele

Existentialism - Jean-Paul Sartre

A key concept in Jean-Paul Sartre's *Being and Nothingness* is his concept of 'bad faith' (*mauvaise foi*),[96] which is also translated as self-deception.[97] Sartre's focus is on the individual in their lived situation. He acknowledges both the static paradox and the dynamic paradox, although he does not reduce them to propositional statements. His concise depiction of the problem of self-deception, as quoted earlier, is:

I must know in my capacity as the deceiver the truth which is hidden from me as the one deceived. Better yet, I must know the truth very exactly in order to conceal it more carefully. This…in the unitary structure of a single project.[98]

Sartre also acknowledges the persistence of self-deception once established:

A person can live in bad faith, which does not mean that he does not have abrupt awakenings to cynicism or to good faith, but implies a constant and particular style of life.[99]

Sartre's concept of self-deception arises out of his ontology, in which a central question is what type of beings are human beings, if they are to be capable of self-deception? Sartre contends that we are conscious beings, and therefore free, in a contingent and therefore determined situation. We are also absolutely responsible for our lives, a responsibility which produces *angst*. As illustration, Sartre asks us to imagine walking a narrow path high on a mountain next to a sheer precipice.[100] I am afraid, Sartre explains, because at any moment I might slip and fall due to the treacherous terrain. I am also acutely aware that I keep myself on the path, but there is nothing in my immediate past to determine the future. I am totally free. As I am totally free, at any instant I might choose to throw myself off the path and down the precipice. I am absolutely responsible for each moment that I stay on the path. I am radically responsible for

[96] Sartre, Jean-Paul. *Being and Nothingness: An Essay in Phenomenological Ontology.* Translated by Sarah Richmond. Routledge. 2018.
[97] There has been some discussion concerning whether the two terms are co-extensive. In later works, Sartre may depict types of bad faith which are not self-deception, as argued by Joseph Catalano 'Lying to Oneself : A Sartrean Perspective' *Philosophy and Phenomenological Research,* Vol 50 No 4 (1990) 673 – 693. https://www.jstor.org/stable/2108229 However, here I am only concerned with the philosophical explanation in *Being and Nothingness.*
[98] Sartre. *'B & N',* 90
[99] Sartre. *'B & N'* 91
[100] Sartre. *'B & N'* 67

my life. This creates intolerable anxiety. Bad-faith, or self-deception, is a means of alleviating that anxiety.

Human beings are free consciousness in a contingent situation, and are always 'both' and 'neither'. In living in the world and trying to understand ourselves and our place in the world, we try to grasp this 'metastable'[101] duality in our situation. We inevitably fail when we attempt to focus on either contingency or consciousness/freedom, as the one refers to the other. Not only that, it dissolves and becomes the other.

Sartre's paradigm examples

Sartre provides examples illustrating the two basic manifestations of self-deception. The first, in the chapter in *Being and Nothingness* entitled *Bad-Faith*,[102] is the over-conscientious waiter.[103] The waiter is trying to pretend he is 'just' his social role, a waiter, and nothing else. More, he is trying to *realise* his status as a waiter, so he is playing at what it is to be a waiter, and in doing so he is trying to understand himself, his 'being'. Paradoxically, the more he tries to be consistent with this contingent status, the more he is aware that he is doing so freely. It is a choice, but he could choose otherwise. He could choose not to get up at five, not to sweep the floor or start up the espresso machine, even if it means he will be fired. We all do this, Sartre says. Not only waiters, but tradespeople, sales people, teachers, even philosophers writing in French cafés. This is not the privileged Sartre misunderstanding the situation of the oppressed waiter, as the example is often mis-read. Sartre provides this example to illustrate how we may fall into self-deception.

The same structure can apply if we deny all contingency and pretend we are total consciousness, total freedom. Sartre's example here is that of a young woman on an early date with a man.[104] She is not ready to 'make a decision' regarding the romantic aspects of the relationship. When the man finally takes her hand, *we know what happens next*, writes Sartre. (*T)he young woman leaves her hand there but* ***she does not notice****…because it happens by chance that she is at that moment all intellect…*[105] Like the waiter, the young woman is playing with her existence. In contrast to the waiter, she is pretending she is all freedom and consciousness with no 'contingency' involved. Sartre is describing flirtation, which is part of the game of romance. While the freedom of the intellect is what the young woman projects, there can be no doubt that the 'contingency' is very much in the picture. The young woman will 'have to make a decision'. Like the walker on the path overlooking the precipice, this involves serious responsibility as such a decision, particularly when made in the early 1940s in France, may drastically affect her life.[106] She cannot be unaware of her situation.

---

[101] Although the more familiar term 'unstable' might appropriately convey Sartre's meaning, Sartre means more than mere instability. 'Metastable' refers to the tendency to dissolve certainty in structure and refer to, or become, its opposite. Freedom refers to and ultimately becomes contingency and vice-versa.
[102] Sartre. *'B & N'* 87-113
[103] Sartre. *'B & N'* 103
[104] Sartre. *'B & N'* 97
[105] Sartre. *'B & N'* 98
[106] Note that Paris at the time was Nazi-occupied, the French in the Resistance lived in danger, the Catholic religion determined moral mores, contraception was unreliable, there were barriers to work and independence for women as well as discrimination against single mothers and their children. There are many consequent scenarios for the young woman other than a romantic one.

These examples illustrate the structure of self-deception and, Sartre argues, illustrate that self-deception is not only possible, but indeed inevitable for human beings.

The phenomenology of 'belief' and self-deception

However, Sartre still needs to account for self-deception as stated above, that the self-deceiver is both deceiver and deceived at the same time, and that they know this.

To explain this, Sartre examines the concept of 'belief'.[107] If it were a matter of facts, Sartre argues, self-deception could not arise. Our relationship to a 'fact' is based on evidence and a 'fact' is a certainty. We require a stringent standard of evidence in matters of knowledge and fact. We may be right, wrong or mistaken about a fact but the relationship is objective and clear.

In matters of belief, Sartre says, the object of belief is given indistinctly or in some cases, not at all.[108] In order to believe, we accept a lesser level of evidence, evidence which is not as fully persuasive as in matters of fact or good faith. The self-deceived manipulate, or play on this characteristic of belief. The self-deceived decide that 'non-persuasion' is the structure of all convictions. In accepting the non-persuasive evidence, or not-fully-persuasive evidence, the self-deceived also make a 'decision' to accept that evidence. This is not a considered, reflective decision, but a 'spontaneous determination of our being'.[109] Sartre says:

One puts oneself in bad faith as one goes to sleep and one is in bad faith as one dreams.[110]

A further aspect of self-deception follows from its status as mere belief, not knowledge. Sartre points out that all belief, just because it is only belief, entails the idea of its opposite. That is, if one knows that 'p', this is certain. If one merely 'believes' that 'p', this indicates that '~p' is also part of one's belief. *To believe is not-to-believe*.[111] In self-deception, which is a particular psychological structure utilising belief, in order for the self-deceived to adhere to the (mere) belief that 'p', the first act is to decide to believe. Then the self-deceived must also take the emotional effort to maintain that belief; selectively collecting evidence for 'p' and in relation to '~p' concentrating on the vagueness of the contrary evidence in order to dismiss, rationalise, or simply ignore it. Once established as part of one's life, self-deception is maintained as a matter of course. It is like a bad habit.

Translucent consciousness and self-deception

On Sartre's view, there is no need to resort to the concept of a Freudian unconscious to explain self-deception. Sartre rejects the notion of the unconscious as being 'explanation by magic'. In particular, the notion of the 'censor' and repression fails; it is difficult to explain without acknowledging awareness or consciousness of what is repressed. Sartre posits a 'translucency' rather than a 'transparency' to consciousness but rejects opacity.[112] 'Translucency' allows for the possibility that conscious awareness may be more or less distinct, or it may be mistaken.

[107] Sartre. *'B & N'* 114
[108] Sartre. *'B & N'* 115
[109] Sartre. *'B & N'* 114
[110] Sartre. *'B & N'* 115
[111] Sartre. *'B & N'* 116
[112] Sartre. *'B & N'* 91ff

Sartre's explanation of self-deception explains the static paradox and the dynamic paradox. He also explains the motivation for self-deception, although the explanation from Existential angst may lead us to stray from the boundaries of the current work.

What Sartre gives us, which is important for our consideration of self-deception, is an understanding that a self-deceived person is a person living a complex life in a situation. For Sartre, the self-deceived may be trying to understand, realise, escape or control a lived situation

Further, self-deception is the inevitable lot of human beings because without it they would be faced with the reality of their absolute responsibility for their life, and the absurdity of their existence. This realisation produces anxiety which cannot be tolerated. However, further analysis of this aspect of Sartre's Existentialist philosophy is beyond the scope of this project.

'The Other' and first-person authority

Sartre cannot explain the challenge to first-person authority, or can do so only with great difficulty. This indicator points to an interpersonal dimension to self-deception. Sartre develops a theory of relations with others ('the Other'),[113] which reflects the dichotomies he proposed in relation to consciousness, and is fraught with frustration and tension. There is little room for the interpersonal understanding of the situation which, for example, Emily's friends demonstrated when they questioned her first-person authority in the example introduced earlier, in Chapter One.

Sartre's view of human beings as consciousness in contingency might lead to solipsism, except for the idea of the Other. For Sartre, we are all conscious beings organising our world around us. This is a successful enterprise until I become aware of another's consciousness, which occurs when someone looks at me.[114] I become aware that there is another conscious being, organising the world in their way, and that I am an object in their world. I have no way of knowing what I am for the Other, but I also cannot escape the fact that I am that object. Sartre famously illustrates this by asking us to imagine looking through a keyhole at a scene in the room behind the door.[115] The keyhole-viewer is totally absorbed in the scene. Then she hears a creak of floorboards behind her. The Other has entered her consciousness, she is aware that as well as conscious subjectivity, she is also an object for that Other. She experiences that 'objectness' in an emotion; in this case shame, but in another situation, possibly pride. She realises she has no control over her being as it appears to the Other, but she cannot escape it, she must 'wear' it. This occurs before she reflects. This awareness of the possibility of the Other occurs whether there is in reality another person present or not. It occurs even if the creak of floorboards, while suggestive of a person, turns out to merely be the house settling. Thus, for Sartre, the Other robs one of one's subjective position as the centre of the world one is organising, and thus of an aspect of one's being.

This structure forms the basis for all relationships with Others, the aim of which is to possess the Other, not as an object, but as consciousness.[116] This brings about one of two projects which set up a metastable dichotomy and is doomed to fail. In the first type of project, one tries to get the Other to freely deny their subjectivity and acknowledge one's own subjectivity. This is

[113] Sartre. *'B & N' Concrete Relations with Others* 479ff
[114] Sartre. *'B & N' The Look* p 347ff
[115] Sartre. *'B & N' The Look* p 355ff
[116] Sartre.*'B & N' Concrete Relations with Others* 479ff

manifested in hatred, sadism and indifference. It is doomed to fail because if the Other gives up their subjectivity, they do so freely.[117] In the second type of project, one tries to turn oneself into the object and force the Other to assert their conscious freedom. Examples of this attitude are love and masochism. Of course, success means the Other has been forced to act against their freedom and has become object.[118] These projects are self-defeating and the instability of each means it falls into the opposite.

The defining characteristics of Sartre's depiction of our relationships with others are alienation and instability. He does not here give us an indication of how we might reach agreement, cooperation or relationship. Sartre does write about the plural dimension, the 'we', but in this he projects the unstable duality of relations with others onto the experience of *mitsein.*[119] Nor, significantly, does he show us how we come to understand what others believe, nor the truth of their claims about their beliefs. Self-deception is entirely a matter for the self-deceived individual. It is a barrier to the individual aim of authenticity. However, it is not a matter of concern for anyone else.

Remarks on Sartre's theory of self-deception

Sartre's Existentialist account of self-deception succeeds in explaining the paradoxes, because he avoids doxastic explanation in favour of explanation from lived experience. The selectivity issue is resolved by tying the instance of self-deception to avoidance of its opposite mode of being, that is, contingency or freedom. The choice of specific issue about which to self-deceive is entirely the responsibility of the self-deceiver. Thus Sartre obviates the necessity to explain the selectivity issue further.

The structures he proposes to explain our relationship with others affect his ability to explain one of the desiderata for a theory of self-deception, the challenge to first-person authority. It is not clear that Sartre can explain this aspect of self-deception and for the purposes of the question in this thesis, this represents a limitation on the explanatory power of his theory.

The weight to be given to this limitation cannot yet be determined. This is so, particularly given that Sartre's view provides answers to the more difficult, essential questions of 'why?' and the teleological question of the purpose of self-deception. Sartre equalises self-deception. It is inevitable for all of us. He presents self-deception as affecting everyone and subtly suggests living authentically as an alternative, however aspirational. Perhaps, too, his orientation is preferable: address self-deception in oneself first, not others.

Sartre has suggested an answer to some important aspects of self-deception, with some limitation. I will now consider another major approach to the question of self-deception, that of Intentionalism as represented by the theories of Donald Davidson.

[117] Sartre. *'B & N'* 482ff
[118] Sartre. *'B & N'* 501ff
[119] Sartre. *'B & N'* 543ff

Intentionalism - Donald Davidson

Intentionalists embrace the static and dynamic paradoxes. In the Intentionalist view, the task is to explain the holding of contradictory beliefs and how an agent can intentionally self-deceive. Intention, they argue, is necessary in order to account for the selectivity issue in self-deception. A prominent if in some ways uncharacteristic Intentionalist is Donald Davidson and I will now consider his views.

Donald Davidson

Donald Davidson offers the following concise set of conditions for self-deception:

> *'An agent A is self-deceived with respect to a proposition p under the following conditions. A has evidence on the basis of which he believes that p is more apt to be true than its negation; the thought that p, or the thought that he ought rationally to believe p, motivates A to act in such a way as to cause himself to believe the negation of p'.*[120]

Aspects of these conditions which are important to note include that they avoid the hard abstraction and reduction of the static paradox. The agent is in a situation and has evidence. The evidence indicates that p is 'more apt' to be true. There is no requirement, as there is in the strictly doxastic approach, that p be true. There is no requirement that A acquire the belief that p. The requirement is that A has the 'thought' that p or that he ought rationally to believe p. The 'thought' need not be a carefully considered, reflective decision. However, as discussed below, Davidson allows for mental causes. A 'thought' can be a cause.

The next word in the conditions, 'motivates' is crucial. The noun 'motivation' may be a desire, an emotion, a motivation or a combination of these. Each and all of these are irrational and as noted, can be a mental cause. Thus, for Davidson, in self-deception, irrationality intervenes in the rational process and causes the self-deceptive belief. Davidson is concerned with the mental structure of self-deception and to understand that, he needs to explain how a basically rational person can engage in the irrationality of self-deception. By acknowledging the irrational 'motivation', Davidson's account has the possibility of explaining the dynamic paradox.

Further, by tying self-deception to the self-deceived's thought that p or that he ought to believe p, Davidson explains the selectivity issue. The self-deceptive belief is tied to the belief that the self-deceiver is motivated to deny.

Davidson's conditions for self-deception emerge from his broader theories, and I will now attempt to sketch the most relevant of these.

Rationalism and Naturalism

Davidson is identified as a Rationalist,[121] but his rationalism encompasses neither Idealism nor Empiricism. Knox Peden argues that Davidson's position in contemporary philosophy is parallel to that of Spinoza in relation to Kant and Descartes.[122] Davidson's aim, Peden contends, is to 'fortify Rationalism' even while he disputes the dualism between a conceptual scheme (whether idealist or empiricist), and content. Davidson aims for an 'unmediated'

---

[120] Davidson. 'Deception and Division' *Problems of Rationality ('Problems')* 208

[121] See for example Manning, Richard R, 'Rationalism in the philosophy of Donald Davidson', in Nelson, Alan (ed). *A Companion to Rationalism* John Wiley and Sons. 2012.

[122] Peden, Knox 'Donald Davidson's Spinozistic extravagance' *Critical Horizons*, 18:4, (2017) 347-358, DOI: 10.1080/14409917.2017.1374916.

connection between reality and our understanding of the world. Peden suggests that this poses a challenge which could undermine the key conceptual schemes which comprise some of the central 'conceits' of contemporary philosophy.[123]

Davidson asserts that human beings are rational beings and this is what distinguishes them from other living beings and inanimate objects.[124] In the alternative to fitting his theory into a 'conceptual scheme', he formulates a three-way basis for communication. Thus he has been said to propose a synthetic *a priori* account of knowledge of the world, others and oneself.[125]

Davidson has been criticised from a naturalist position, particularly in relation to his explanation of interpretation.[126] This position has been further examined with a view to identifying underpinning meta-philosophical differences in the conception of naturalism between Davidson's view and other more conventional naturalist views.[127]

As may be anticipated from his attempts to avoid theory-specific distinctions, Davidson's work does not easily fit into the understanding of established schools of thought and has attracted criticism from these diverse approaches. This may be why some have suggested that the route to understanding Davidson is through examining his views from a meta-philosophical perspective.[128]

Davidson's focus on rationality should not necessarily incite a criticism that he is over-rationalistic. This is because, firstly, he incorporates an account of the irrational into his theories, including his explanation of self-deception. As well, motivations and desires feature as causes in his explanations. Finally, he points out that he is concerned with general processes while admitting some 'indeterminacy' in particular instances.

Holism and the web of belief

Davidson argues for a holistic theory of belief.[129] This, he admits, is influenced by his mentor, W V O Quine, although others argue that Davidson's differences with and critiques of Quine bring this affiliation into question.[130] The theory is that for a person to hold a particular belief, that belief depends on a 'web' of other beliefs, which relate to it and support it. Davidson goes further, aiming for a unified theory of experience which encompasses meaning, belief, and objectivity. His holism extends to mental content, including beliefs, values, judgements, and intentions. This forms a basis from which his theory of meaning and radical interpretation evolves.

---

[123] Peden, K. 'Spinozistic Extravagance' 354

[124] Davidson. 'Rational Animals' in *Subjective, Intersubjective, Objective.*95-106 Note there is an energetic debate regarding Davidson's refusal to attribute rationality to dogs and similarly relatable animals. While empirical developments in the field of evolutionary biology and evolutionary psychology may eventually challenge Davidson's view, for the present purpose this stipulation by Davidson is useful to identify the boundaries of his concept of rationality.

[125] Gluer, Kathrin. 'The Status of Charity I. Conceptual truth or *a posteriori* necessity?' in Bagrahmian, Maria (ed). *Donald Davidson Life and Words.* Routledge 2013 155

[126] Fodor, Jerry and Lepore, Ernest 1994. 'Is Radical Interpretation Possible?' in J. E. Tomberlin (ed.), *Philosophical Perspectives , Vol. 8, Logic and Language*. Ridgeview. 101–19.

[127] Sinclair, Robert. 'What is Radical Interpretation? Davidson, Fodor and the Naturalization of Philosophy'. *Inquiry* 45 pp 161-184 https://doi.org/10.1080/002017402760093252

[128] Such as Robert Sinclair and Knox Peden, referenced above.

[129] Davidson. 'On the very idea of a conceptual scheme' *Inquiries into Truth and Interpretation* pp183-198

[130] Bagrahmian, Maria (ed). *Donald Davidson Life and Words.*

Reasons as Causes and Rationalising

In one of his earlier articles,[131] Davidson considers the relationship between an action and an agent's reason for it. He argues that, insofar as the reason explains the action, that reason can be considered a cause. Davidson acknowledges his departure from those, including the later Wittgenstein, who argue that causes have to do with physical or external relations, and cannot apply to reasons and actions.

He acknowledges that the 'web of belief' means there can be more than one reason for an action. Also, motivation plays a role in decision and action. Here 'motivation' includes desire or some 'pro-attitude'. Davidson explains his position as follows:

Giving the reason why an agent did something is often a matter of naming the pro attitude (a) or the related belief (b) or both; let me call this pair the primary reason why the agent performed the action.[132]

A primary reason allows us to understand why the agent acted as they did. Reasons justify and/or explain actions. Thus reasons 'rationalise' actions and render them understandable to others. A reason can, therefore, be a cause of an action.

Radical Interpretation and the Principle of Charity

Davidson proposes that it is not possible to assign meaning to speech without knowing what the speaker believes, and vice-versa. The way we manage to make sense of another person, he suggests, is because we use the 'principle of charity' when we interpret others' utterances. That is, interpretation is aimed at optimising agreement. This requires certain assumptions which form the principles of communication. We assume that the other holds their beliefs to be true, that their set of beliefs is rational (principle of coherence) and that there is a causal relationship between their beliefs and the objects of belief (principle of correspondence).[133] This 'principle of charity', places constraints on radical interpretation and furthers communication.[134]

Some have criticised Davidson's position regarding interpretation on the grounds that this theory does not reflect how we actually learn to interpret a foreign language.[135] As well, Davidson's theory does not accurately reflect the contextual nature of the way infants learn language. Others argue that Davidson's concentration on 'utterances' conflicts with his theory of semantic meaning, and holism.[136] Another issue is that Davidson's concentration on speech or 'utterances' ignores the role that the non-verbal plays in communication. However, Davidson is interested in interpretation in the broadest sense, and not in the process of language translation, while admitting indeterminism in relation to instances.

Triangulation

We understand ourselves, the world, and others. Davidson argues that these understandings arise inseparably. It is not possible to reduce the three aspects to two as in idealism or empiricism.

---

[131] Davidson. 'Actions Reasons and Causes'.*Essays on Actions and Events* Oxford Uuniversity Press 2001. 4.
[132] Davidson. 'Actions Reasons and Causes'. 4 *'Actions and Events'* Oxford Uuniversity Press. 2001
[133] Davidson. *'Truth and Interpretation'*. Oxford University Press. 2001
[134] Davidson. 'Radical Interpretation' and 'On the very idea of a conceptual scheme' in *'Truth and Interpretation'*.
[135] Sinclair, Robert. 'What is Radical Interpretation? Davidson, Fodor, and the Naturalization of Philosophy', *Inquiry*, 45:2, (2002) 161-184, DOI: 10.1080/002017402760093252
[136] Camp, Elizabeth. 'Conventions' Revenge: Davidson, Derangement and Dormativity' *Inquiry*, Vol. 59, No. 1, (2016) 113–138. http://dx.doi.org/10.1080/0020174X.2015.1115277

There are three basic problems: how a mind can know the world of nature, how it is possible for one mind to know another, and how it is possible to know the contents of our own minds without resort to observation or evidence. It is a mistake, I shall urge, to suppose that these questions can be collapsed into two, or taken in isolation.[137]

Understanding is a matter of triangulation between the three; one's understanding of the world, of others and of oneself. [138] They arise together.

The Puzzle of Irrationality

Akrasia

If rationality is the way in which we derive understanding of ourselves, the world and others, then how do we deal with instances of irrationality? To explore this, Davidson first utilises the notion of *akrasia*, or weakness of the will.[139] Weakness of the will occurs when a person acts against their own best interests. A motivation intervenes to cause the action. It is the reason for the action, but it does not justify or explain it to others. Thus it is a 'reason' for the action, without 'rationalising' it.

Davidson distinguishes between the non-rational, which is outside of human effect, and irrationality, which is a 'rational process gone wrong'.[140] We may not have difficulty understanding irrationality when we encounter it. What we do have trouble with is when we encounter the failure, within a single person, of coherence or consistency in the pattern of beliefs attitudes, emotions, intentions, and actions.

Consider, Davidson says, partially borrowing from one of Freud's cases, a man who sees a large stick across the public path when he is walking in the park. Let's call him Henry. Henry moves the stick to the nearby hedge, out of the way of other pedestrians who might trip over it, and goes on his way. Later, he realises that it may be more of a danger where he has left it in the hedge, so he returns to move it and places it back on the path.

We understand what Henry is doing because we understand that the reason for his action is to minimise danger to other people in the park. The reason explains his action, it gives the intention with which he acted. We can say of Henry that all things considered, he acted according to his best judgement.[141]

Suppose, however, that Henry has an important meeting for which he should not be late. He is on the way to the meeting when he realises that the stick that he placed in the hedge could be a danger to pedestrians. He does not have time to remove the stick and be on time for the engagement. He goes to the park and removes the stick anyway.

---

[137] Davidson. 'Three Varieties of Knowledge' *Subjective.* 208

[138] Davidson. 'Three Varieties.' 213

[139] Davidson. 'Paradoxes of Irrationality'. *Problems.* 169ff

[140] Davidson. *Problems.* 169

[141] Note that some disagree with this proposition. For example, Marcia Cavell, relying on Freud's original case, argues that it is not possible to understand these obsessive actions without also understanding the unconscious urges and drives which prompted them. Certainly Freud's original case was troubled by more than *akrasia*, and exhibited other more inexplicable behaviours indicative of psychopathology. The entire complex of distressing behaviour included this more commonplace example. That does not mean that Henry, as described by Davidson, cannot be understood at the level of consciousness and rationality. As described, Henry does not seem to be much more irrational than those of us who occasionally feel they must go home to check whether they have really locked the door. See: Cavell, Marcia 'Dividing the Self'. Preyer, Gerhard, Siebelt, Frank and Ulfig, Alexander (eds.) *Language Mind and Epistemology:On Donald Davidson's Philosophy* Synthese Library. Springer Science and Business Media.1994.

Henry had two courses of action open to him; attend the important meeting on time or move the stick and be late. He had reasons for both, but better reasons for attending the meeting on time than for moving the stick. Yet, there is a rationality in his actions and no inconsistency in choosing one over the other.

The inconsistency is evident only if Henry also believes that he should act according to his own best judgement, everything considered. If he believes that, then he has contravened that second-order belief in this case. That is, his usual course of action is in line with the second order belief, but in this case, he contravened it. In doing this, he is not acting according to his own best judgement. This is an example of weakness of will, or *akrasia*.

What has also happened in this case is that there is a mental cause of action which is not a reason for the action. A motivation, desire or emotion has intervened to cause the action. These causes cannot 'rationalise' the action. For example, Henry might fear that someone may trip on the stick. It is a cause of the akratic action, it is not a reason for the action. Given that Henry has to be on time for his meeting, this fear does not justify Henry's action, even from his own point of view.

Wishful Thinking

Davidson next considers wishful thinking as a cognitive attitude with some parallels to *akrasia.* He distinguishes wishful thinking from self-deception. Wishful thinking does not need to have any basis in objective reality. Wishful thinking occurs just because one wants it to be true. The agent is not required to do anything to create or sustain wishful thinking.[142]

Self-Deception

When he turns his attention to self-deception, Davidson argues that it is a cognitive parallel to *akrasia.* Whereas *akrasia* is weakness of will in relation to action, self-deception is 'weakness of the warrant' in relation to beliefs. In *akrasia* the agent acts against their own best judgement. In self-deception, the agent acts against their own principles for valid or justified belief formation.

Those principles are shared principles for rational belief formation which require that belief should be determined by the weight of evidence for and against it. When we are deciding between mutually exclusive hypotheses, we should give credence to the hypothesis most highly supported by all the available evidence. In self-deception, these belief-formation principles are contravened.

As with *akrasia*, self-deception is a breach of a rule generally observed by the agent. In self-deception, the agent does not live up to their own epistemic standards. The difference is that in self-deception, the cause of the subject's action is a belief.

This is not all that is involved. This breach of epistemic standards is intentional. It is something the self-deceived person *does*. However, they do not do it rationally. The self-deceived does not have a reason for their breach of belief-forming standards, but there is an element which causally accounts for the breach: the 'weakness of the warrant'. At this point, Davidson proposes his definition of self-deception, quoted above.

Davidson's paradigm example of self-deception is Carlos who, on past experience of multiple failures of his driving test, believes he will fail his next driving test. However, taking into account the evidence of his charm and personal connections along with the fact that it would be too painful if he were to fail again, he induces himself to believe he will pass.[143]

---

[142] Davidson. *'Problems'*. 205-207
[143] Davidson. *'Problems'*. 209

It has to be said that as with others of Davidson's examples, this is an unfortunate one. It does not intuitively strike one as clearly being self-deception. There is slim and largely irrelevant evidence on which Carlos bases his belief that he will pass the test. Further, most people who face tests, whether for driving or something else know that charm alone is unlikely to achieve a pass. Therefore, it more closely resembles wishful thinking.

Weakness of the warrant in self-deception is self-induced. It is precisely because the agent is aware of the ~p that the agent believes p. It is this awareness of the negation that brings about the belief p, even though it is not rational.

Not only does the awareness of ~p bring about the self-deceptive belief that p, but it requires continued awareness of ~p to sustain the self-deception. Davidson claims:

When reality (or memory) continues to threaten the self-induced belief of the self-deceived, continuing motivation is necessary to hold the happy thought in place. If this is right, then the self-deceiver cannot afford to forget the factor that above all prompted his self-deceiving behaviour: the preponderance of evidence against the induced belief.[144]

Davidson has provided a comprehensive explanation of how self-deception can come about.

Important in Davidson's contribution to our understanding is that he explains why there must be contemporaneous contradictory beliefs in self-deception. It is important to anchor the self-deceptive belief to its opposite and to recognise that the 'unwelcome' belief, the ~p, is what triggers the self-deceptive belief. Otherwise self-deception would be indistinguishable from any number of wish-fulfilment or fantasy beliefs.

In doing so, Davidson answers the 'selectivity problem' which is an enduring issue in self-deception. Self-deception is not an all-pervasive phenomenon in any self-deceiver. It is characterised by a rational person who *in this instance* is irrational. It is because the phone calls and absent week-ends suggest that Daniel is having an affair that Emily believes on the basis of his earlier promise that he is faithful. It is the awareness of somatic evidence suggesting some medical condition that causes Evelyn to believe her symptoms are just part of getting older. It is because Amy ignores him that Ryan spins the fiction that she is playing hard to get. It is the growing stack of unfinished novels that are going nowhere that causes Liam to believe that it is his only writer's block that stands in the way of his being a great writer.

This restriction also brings the possible beliefs about which one can be self-deceived closer to home. We will be self-deceived about beliefs relating to or impacting on ourselves at a personal or close to personal level. It is these things which elicit a desire or emotion. It is these things we care about. We will not be self-deceived about events and things in the world.

Supposing at election time I review the parties' policies, vote for the party closest to my own values, and it never occurs to me to consider whether possible corruption in that party will affect their policy implementation. I am not self-deceived. I am deceived by the party in question. If I have considerable experience in politics and know that members of that party might be corrupt, but fail to check that before voting, I may be at fault for carelessness in failing to make inquiries prior to exercising my right to vote. I am not self-deceived. It is possible to be deceived, but not self-deceived about an objective situation in the world which has no personal relation to me, where the evidence is readily available. There is no 'cause' which is not rational in these cases, although there may be mistake or lack of care.

A further positive contribution to our understanding is Davidson's starting point that our understanding of others comes from shared rationality. Self-deception is an instance of

[144] Davidson. *'Problems'*. 210

irrationality in a predominately rational person. This is important as a means of setting the boundaries on our consideration of self-deception. The types of beliefs that can be considered self-deception must by this definition be possible for a rational person to believe (which is not the same as saying they are true).

This by implication means that beliefs which cross the boundary into dysfunction or pathology are not self-deception. It means that full-blown delusions are not self-deception.[145] This boundary on self-deception is important for another reason. While we hold the self-deceived responsible for their self-deception, we do not hold individuals responsible for the effects of psychopathology.

Davidson's Partitioned Mind requirement

Davidson suggests a theory proposing a partitioned mind might help in understanding how the self-deceived subject can deal with two conflicting beliefs. However, for this purpose, he explicitly rejects the empirical aspects of Freud's theories.[146] He also suggests a more subtle division of mind might answer the issue of how the contradictory beliefs can be held by the one person. Yet, his notion of partitioned mind remains undefined. Such a theory may rely on subliminal aspects of consciousness or theories which propose an accessible 'subconscious' rather than a fully unconscious area of mind.[147] Nevertheless, if we are to accept partitioned mind and want to discover how that might work, we need to look elsewhere.

I have considered Davidson as a key representative of the Intentionalists, who embrace the paradoxes. Opposed to Intentionalists are Motivationalists who argue that it is unnecessary to consider the paradoxes in self-deception. I now turn to this approach.

[145] Although, as noted above, it is possible to theorise that self-deception taken too far will lose its relationship to evidence and may become delusion.
[146] Davidson. *'Problems'*. 181
[147] For over a century there have been many experiments aimed at identifying unconscious influences and subliminal priming on behaviour and conceptual thought. An overview and critical analysis can be found in: van den Bussche, Eva. 'The Mechanisms of Subliminal Stimuli:A meta-analysis and new experiments.' Doctor of Psychology thesis, Catholic University of Leuven, 2009. Chs 1 and 7.
https://www.researchgate.net/publication/28360860

Motivationalism

Alfred R Mele

Alfred R Mele's Motivationalist theory proposes a 'deflationist' explanation of self-deception, which he developed with the aim of avoiding the paradoxes of Intentionalism.[148] Mele's theory dispenses with the practice of modeling self-deception on interpersonal deception, and thus requires neither the static paradox nor the dynamic paradox to explain self-deception. Self-deception, Mele argues, is a species of motivated false belief which 'happens to' the self-deceived.

Mele's *Self-Deception Unmasked*,[149] in which he sets out his theory of self-deception as motivated biased belief formation, is a consolidation of much of his earlier work. His account explains how both self-deception and so-called 'twisted self-deception' are motivated. Either desire or emotion or a combination of both comprise the motivation for the formation of the false belief. Emotion is particularly relevant in 'twisted self-deception', which is considered separately below.

Paradigm and other examples of self-deception

Mele's approach to theory is informed by empiricism.[150] Initially, he tested his proposed paradigm examples on students to identify intuitions about self-deception.

His paradigm examples of what he calls 'garden variety' self-deception are:

1. Sid who believes that Roz really is interested in him but playing 'hard to get'.
2. Sam who believes his wife Sally is faithful to him despite contrary evidence.
3. Gordon's parents who believe their son, a US Marine, is innocent of the treason with which he has been charged.
4. The father who believes against the evidence that his son is not abusing drugs.
5. Don, whose academic article has been rejected, believes the academic journal editors cannot have understood the article and wrongly rejected it.

Mele makes mention of more generalised examples derived from surveys conducted by Gilovich.[151] These show that 94% of university professors believe they are better than their colleagues and all of one million high school seniors thought they were above average in social skills, with 25% believing they were in the top 1%. This, Mele concludes, suggests there is a strong motivation to deceive oneself.

The processes of biasing belief in self-deception

Mele relies on explanations from empirical psychology regarding the way evidence can be biased in the processes which we normally use to test hypotheses. These, he argues, are relevant to self-deception.

---

[148] Mele, Alfred R. *Self-Deception Unmasked* Princeton University Press. Princeton. 2001
[149] Mele. *SD Unmasked*
[150] Mele, Alfred R 'Recent Work on Self-deception' *American Philosophical Quarterly* Jan 24:1. (1987) 1-17 Stable URL: https://www.jstor.org/stable/20014170
[151] Gilovich 1991 cited by Mele in *SD Unmasked* 1

'Cold' biases in evidence processing include the comparative vividness of information, the 'availability heuristic', and the 'confirmation bias'. These biases are not necessarily motivated, but apply broadly to information processing.

'Hot' biases occur when desire or strong emotion colours the way evidence is processed or interpreted. These processes are 'triggered' and operate to establish and sustain a motivationally biased belief.

As well, Mele considers 'lay hypothesis testing', particularly as proposed by James Friedrich's PEDMIN model.[152] Importantly, this model asserts that 'pragmatic reasoners' are primarily concerned with avoiding crucial, costly, or harmful errors and less so with testing for truth.

To complete his theory, Mele supplements the PEDMIN model with the research undertaken by Yaacov Trope and Akiva Lieberman.[153] They found that once the major threat, in the form of the crucial error, was identified, the remainder of the evidence processing was 'often semi-automatic'.

Trope and Lieberman proposed variable 'confidence thresholds' for lay reasoners to cease to test for a proposition, as follows:

- An Acceptance Threshold, which is the minimum required to accept the truth of a proposition and cease testing for it, and
- A Rejection Threshold, which is the minimum required to reject the proposition and cease testing for it.

There is no objective criterion for these thresholds. For any particular question they are not 'equally demanding'. They move, depending on the cost of false acceptance or false rejection of the 'hypothesis', compared to the cost of obtaining more information.

Mele calls this entire theory, incorporating both Friedrich's PEDMIN model and Trope and Lieberman's work, the 'FTL model'. This evidence-biasing behaviour is strategic, but need not be fully conscious, nor agential. Importantly, there is a causal connection between the desire and the acquired false belief.

Mele's sufficient conditions for self-deception

Mele is now in a position to propose the following jointly sufficient conditions for 'S's' self-deception:

1. *The belief that p which S acquires is false. (This is a 'lexical' point)*
2. *S treats data relevant, or at least seemingly relevant, to the truth value of p in a motivationally biased way.*
3. *This biased treatment is a nondeviant cause of S's acquiring the belief that p.*
4. *The body of data possessed by S at the time provides greater warrant for ~p than p.*[154]

Mele further explains his jointly sufficient conditions. Point 1, he notes, is a 'lexical point' because a person is deceived or self-deceived only if p is false. However, it does not rule out the possibility that someone may deceive themselves into believing an improbable truth. Mele also states that point 4 is not a necessary condition. It may be that the evidence biasing process

[152] Primary Error Detection and Minimisation (PEDMIN) developed by James Friedrich.
[153] Trope, Y., & Liberman, A. *Social hypothesis testing: Cognitive and motivational mechanisms.* In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* 239–270. The Guilford Press. 1996.
[154] Mele. *SD Unmasked.* Chapter 3. 1

has ensured that the self-deceived does not possess evidence for the opposing proposition.[155] This opening up of the 'sufficient conditions' is one factor leading to a consistent criticism of Mele (considered in more detail below), that his theory is overinclusive.

Mele proposed one additional test for bias in evidence gathering in self-deception. That is the Impartial Observer Test. If the same evidence was presented to peers of the allegedly self-deceived person, would they reach the same conclusion? If they would not, that is an indication of self-deception.

Mele's anticipation of his critics

Mele attempts to anticipate and answer the types of criticism that will be directed at his work, particularly from Intentionalists. As well, he sends out challenges to potential critics, and succeeds in deflecting their responses.

Firstly, Mele notes that of his critics, none who adhere to the 'dual-belief' theory of self-deception have actually given examples of dual beliefs. They have instead resorted to 'mental exotica', including higher-order beliefs, galvanic skin responses (lie detectors) and various partitions of the mind.[156] Mele's contention that none of this explains self-deception is valid.

Mele reviews the George Quattrone and Amos Tversky experiment, [157] in which students were asked to keep an arm in cold water, as a test of dual belief and agency in self-deception. Students changed their assessment of whether the cold was bearable largely in line with information they had been given about the relationship between cold water tolerance and heart health. Mele concludes that the experiment demonstrates neither dual-belief nor agency in arriving at this possibly self-deceptive belief. However, the experimenters adopted a theory of self-deception which allowed for 'unconscious' factors in belief-formation. This experiment has also been interpreted as a test of unconscious or subliminal priming. It is therefore not clear that the results provide support for any theory of self-deception.

Overinclusiveness

One issue noted by critics, which Mele did not formally anticipate, is that his theory of self-deception is too broad. It accounts for phenomena which we might not want to accept as being self-deception, such as, the above examples of the academics who over-estimate their achievement or the students who over-estimate their social skills. For these and similar examples, identifying them as self-deceptive rather than simply mistaken or outright wrong is an unnecessary complication on explanation. Mere ego-inflation is not self-deception. Nor might we want to accept as self-deception Mele's paradigm example of Don, who believes his article was wrongly rejected. Given the effort and the academic stakes involved, Don's reaction may well, understandably, show disappointment, disbelief and emotion. Don may be mistaken, or wrong. It is unclear that he is self-deceptive.

The claim that Mele's theory is overinclusive is reinforced by other Motivationalists, who are his supporters. Earlier, I considered Richard Holton, who attempts to modify Mele's sufficient conditions by adding a necessary condition that self-deception must involve a *de se* belief.[158]

[155] Mele. *SD Unmasked.* Chapter 3 2
[156] Mele. *SD Unmasked.* Chapter 4.
[157] Quattrone, George A and Tversky, Amos 'Causal versus Diagnostic Contingencies: on self-deception and on the voter's illusion.' *Journal of Personality and Social Psychology* 1984 Vol 46: 2 237-248.
[158] Holton, Richard. 'What Is the Role of the Self in Self-Deception?' *Proceedings of the Aristotelian Society,* 2001.

Kevin Lynch,[159] claims that Mele frames self-deception in terms of 'unwarranted degrees of conviction'. This does not help to focus Mele's theory. Lynch has also attempted to differentiate 'escapism' from Mele's theory of self-deception,[160] arguing that Intentionalists are really describing escapism. Lynch later seeks to distinguish self-deception from 'stubborn belief'.[161] It is not necessary to critique the merits of these articles. For present purposes, it is clear that they are attempts to establish more definite boundaries around Mele's position.

Selectivity Problem

Bermudez raised the issue of the selectivity problem in an early response to Mele's work.[162] The problem is not unrelated to the issue of overinclusiveness. Bermudez claims that Mele cannot account for why an issue qualifies for self-deception while other issues do not. Bermudez further claims that Intentionalism has an answer to this. Although the strong 'dynamic paradox' may have its own problems, the intention of the self-deceived is linked to the self-deceptive belief and thus Intentionalism has a solution to the selectivity problem.

Recently, Mele answered this long-standing criticism by pointing out that it also applies to certain types of Intentionalism.[163] Citing Bermudez as its main proponent, Mele argues, correctly, that if self-deception requires an intention to self-deceive, the selectivity problem re-appears. These Intentionalist theorists have to answer the question 'why intend to self-deceive on this issue and not another?'.

There is, as noted in the previous section on Davidson, another response to the 'selectivity problem'. Davidsonian Intentionalism answers this by reference to awareness of the ~p situation or proposition and the so-called 'static paradox'. The self-deceptive belief is triggered by and tied to awareness of the unwelcome proposition.

The effect of overinclusiveness and the selectivity problem

Mele's theory held the prominent position as an explanation of self-deception for decades, and still does for Motivationalists. There is no doubt that Mele explains biased belief formation and thus contributes to our understanding of self-deception.

The issues considered here concern the explanatory power of a theory, including whether the theory obviates the need for one of the desiderata. In relation to the selectivity issue, Mele's theory has neither accounted for the selectivity issue, nor shown why it is unnecessary. This affects its explanatory power.

The selectivity indicator for the explanatory power of a theory of self-deception requires that we can explain, for example, why Sam self-deceives about Sally's faithfulness. Sam may know many things about Sally. Sally is pretty, she may inherit a fortune from her aging aunt, she is popular at work, she has probably maxed out her personal credit card, she gets a bit silly when she has too many tequila shots, she hates cooking, her hair colour is not natural and she may be unfaithful to Sam. Sam may feel strongly about any or all of these situations. Mele gives us no way of explaining why, of all the things he knows about Sally, Sam deceives himself about

---

[159] Lynch, Kevin 'On the tension inherent in self-deception' *Philosophical Psychology* vol 25: No 3 June 2012 433-450. DOI: 10.1080/09515089.2011.622364

[160] Lynch, Kevin. 'Tension in SD'

[161] Lynch, Kevin. 'Self-deception and stubborn belief' *Erkenntnis* 2013 78: 1337-1345 DOI 10.1007/s 10670-012-9425-0 https://www.jstor.org/stable/24010975

[162] Bermúdez, José Luis. 'Defending Intentionalist Accounts of Self-Deception.' *Behavioral and Brain Sciences.* 20, no. 1 1997: 107-08

[163] Mele, Alfred R. 'Self-deception and selectivity' *Philosophical Studies* 177 (2020) 2697-2711 https://doi.org/10.1007/s11098-019-01334-9

Sally's faithfulness. Although we may intuit an answer, it is not available through a consideration of Mele's theories.

The ~p, the contradictory or conflicting proposition, has an unacknowledged effect in each of Mele's paradigm examples. Most obviously, this is the case regarding US Marine Gordon's parents. One might expect nothing less than that they have faith in their son and believe that he is innocent of treason. If someone, perhaps out of spite or jealousy, merely made that allegation, Gordon's parents would refute it firmly and immediately and call it out for what it is, without self-deception coming into the question. Yet, in this example as described by Mele, Gordon has already been charged. There must be at least enough evidence against him to make a case to be tried by the Military Courts Martial. The evidence of his guilt is the basis for his parents' disbelief, and their self-deception. Mele does not acknowledge the role of this evidence in self-deception.

The challenge of twisted self-deception

An early challenge to Mele's Motivationalist account of self-deception came from David Pears' example of 'twisted' self-deception.[164] In twisted self-deception, it is claimed, the self-deceiver deceives herself into believing something that she does not want to be true.

David Pears' paradigm example is the unjustifiably jealous husband,[165] who Mele later names 'Jeff'. Jeff deceives himself into believing his wife is having an affair, even though he desires that she is faithful to him.

Mele admits that his theory that self-deception is a false belief motivated by desire is threatened by the possibility of twisted self-deception. He considers whether emotions can play a part in motivating the false belief. He concludes that neither desire nor emotions alone can account for twisted self-deception. However, if the emotional state primes the desire, then the FTE theory can account for twisted self-deception as motivated false belief. [166]

In this example, the motivation for the jealousy is not straightforward. The aim of Jeff's unwarranted jealousy, Mele and others claim, is to establish a closer relationship with his wife, by increasing vigilance in his relationship and eliminating rivals.

As noted above, in the monograph consolidating his theory,[167] Mele includes both desire and emotion as elements required to explain self-deception, including twisted self-deception.

Response to Mele's theory of twisted self-deception

In response to Mele, others have proposed instances of people who seem to desire something that it is difficult to believe they desire. Many of these are more likely examples of psychopathology than twisted self-deception.

These include the case of Dolores, a mother who convinced herself that she was responsible for her daughter's death from leukaemia, because she let her daughter play with an ill cat.[168] It has been variously proposed that in this case there was a desire to self-punish, or a desire to wrest some control from the uncontrollable workings of fate. This tragic case cannot be distanced from the psychological effects of grief. Dolores' beliefs are pathological, and Dolores is not alone in developing psychopathology, however temporary, following severe loss. Admittedly, with a stretch, the beliefs may be fitted into Mele's conditions for self-deception,

---

[164] Mele, Alfred R. 'Twisted Self-Deception'. *Philosophical Psychology,* 12:2 (1999) 117-137, DOI:10.1080/095150899105837

[165] Pears, David (1984) *Motivated Irrationality* OUP 42-44 as cited by Mele in *Twisted SD*

[166] Mele. *Twisted SD* 133

[167] Mele. *SD Unmasked*

[168] Mele. *Twisted SD* 129

mentioned above. However, there does not seem to be anything that distinguishes these beliefs as self-deception, and much that suggests Dolores' beliefs are so irrational, so slimly tenured to any evidence, that they approach delusion. That is beyond the scope of correction without the healing effects of time, or professional intervention.

In another example, it is proposed that a person with a 'negative self-schema' will desire to fail, and enter into a belief that they will fail, in order to avoid disappointment.[169] Such desires, it is argued, arise out of various unconscious motivations or perhaps 'complexes'. Again, it is difficult to see that this comprises any type of self-deception, twisted or straight. Rather, irrespective of whether the belief arises from unconscious motivations or not, it appears to be 'clear-eyed' thought manipulation to avoid disappointment.

The example proposed by Mele and Pears, of the unjustifiably jealous Jeff has additional problems.

Jealousy, anger and fear are emotions experienced by all human beings and demonstrably, some animals.[170] Neuroscientists tell us that fear, anger and jealousy are emotions triggered in the part of the human brain which responds in a primitive way to threats, the amygdala. These emotions have primitive survival value, stimulating the 'fight or flight' response as well as the release of cortisol. Jealousy is triggered when the threat is to something which is 'ours', so there is an intrinsic possessive aspect to jealousy. Jealousy also stimulates the part of the brain which is aroused by social slight.[171]

In order to live our complex lives, we have ways of dealing with these emotions when we experience them. Reason, personal and social control modulate our reactions. Our lives would be impossible if this were not the case.

Just as the extreme of anger is uncontrollable rage, so jealousy can result in extremes which are pathological. This suggests that the 'increased vigilance' which Pears and Mele mention in the example of Jeff may be euphemistic. 'Vigilance' in relationships translates to suspicious behaviour and control. We know what it means for the wrongly accused partner: a life of hurtful accusations, controlling behaviour and in the extreme, violence.

Further, excessive jealousy is a feature of a number of disorders identified in the DSM,[172] as well as being a feature of neurological disorders. It is better explained medically.[173] Certain types of brain damage to the frontal lobe can interfere with the ability to inhibit or control emotions, especially those basic vestigial survival emotions which include anger and jealousy.

---

[169] Mele. *Twisted SD* 128-129

[170] See: Bucher, Benoit & Arahori, Minori & Chijiiwa, Hitomi & Takagi, Saho & Fujita, Kazuo. (2020). Domestic cats' reactions to their owner and an unknown individual petting a potential rival. 10.21071/pbs.vi9.12176. Christine R. Harris, Caroline Prouvost. Jealousy in Dogs. PlosOne July 23, 2014 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094597

[171] Envy, on the other hand, occurs when we compare ourselves to others. It is different from jealousy, but the terms are often used interchangeably. Envy is a socially-triggered emotion and is partially stimulated in the frontal cortex of the brain. When we react to one of our colleagues gaining an undeserved promotion, or a school friend building a better house than ours, that is envy. It can persist over time and of course can make us miserable. It does not have the instinctual survival value that fear, anger and jealousy have.

[172] American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5.*

[173] For example: *delusional disorder, jealous type; Othello syndrome.* See also by way of summary: 'Apart from patients with schizophrenia and alcohol abuse, delusional jealousy can be frequently found in patients with organic psychosis, as for example in Alzheimer's disease. ' (Letter to Editor by Michael Soyka, MD Psychiatric Hospital, Munich, 1998. https://neuro.psychiatryonline.org/doi/pdf/10.1176/jnp.10.4.472 There is also evidence that pathological jealousy (Othello syndrome) is one symptom of various neurological disorders (Graff-Radford et al, 2012).

People suffering from these disorders cannot control their emotions, including jealousy, and we do not hold them responsible for their condition.

Along the spectrum between jealousy which is easily dealt with and that which is pathological there can be little theoretical space for unjustified jealousy which is intense enough to comprise twisted self-deception. This must be so, when we remember that by definition, the jealousy is both unjustified and something that Jeff cannot want to be true.

The lack of an example of twisted self-deception that is not better explained through pathology reasonably leads one to question whether twisted self-deception is a real phenomenon.

Desiring a negative and self-deceiving about it

This does not mean one cannot desire a negative which is also self-deceptive. It merely means that while such a desire may be difficult for others to understand, it is not twisted. I argue that it is entirely possible to self-deceive regarding what others see as a negative.

If this is possible, then perhaps the 'twisted' aspect of self-deception derives from an unacknowledged normative element projected onto the motivating desire in self-deception, rather than the structure of self-deception. That is, the projection of positivistic values into the notion of desire, which perhaps underlie social mores. Desire, accordingly, should be aspirational, directed towards growth, improvement, reaching one's highest goals. These aspirations are not equally available to all sections of society.

It is, therefore, possible to desire something which others find difficult to understand, without the desire being 'twisted'. Here is an everyday example:

Sarah

Sarah values her family and her circle of friends. Her family holds regular family dinners and week-end 'get-togethers'. She often meets her friends for lunch or a coffee. Make that a coffee and cake.

Sarah lives with being overweight. She knows that this could lead to obesity and more severe health issues. Increasingly, she cannot wear the clothes she would like to because of her size. She believes she would look more attractive without the excess weight. So it seems she could not possibly want it to be true that she cannot lose weight.

Sarah is fully aware that dieting might delay or prevent these consequences. She has been 'yo-yo dieting' for years. She experiences the deprivation, the tests of willpower and the impact on her family and social life. With each diet, despite early weight-loss results, eventually, she experiences the rebound effects comprising even more weight gain.

Recently, Sarah has done internet searches concerning the impact of metabolism and other issues on weight. She now believes that because of her particular metabolism type, she cannot lose weight. She does not seek out clinical confirmation of this belief, 'Dr Google' is sufficient.

Sarah deceives herself into the belief that she cannot lose weight. That means Sarah is absolved of responsibility for her condition. It is not that she does not want to lose weight, she claims. It is not that she has not tried to lose weight. It is simply that she cannot lose weight.

Sarah has deceived herself into a negative belief about herself that is likely to have further negative consequences. Yet there is no psychopathology in her situation. She is an intelligent woman who is leading a life socially engaged with family and a circle of friends. She has considered evidence and come to this self-deceptive belief about herself. Just because we cannot see why she would want it to be true that she cannot lose weight doesn't mean it is a matter of twisted self-deception. Apart from the apparently negative value we attach to her

desire, this example has the same structure as the four paradigm examples identified at the beginning of Part One, two of which echo Mele's paradigm examples.

The Possibility of Group Self-Deception

If it is possible for an individual to desire the negative proposition, is it possible for a group or social cohort to each desire the same negative proposition in relation to themselves? Rorty has noted that self-deception happens 'with a little help from our friends'.[174] Wieland has noted the societal rationales behind the excuses we concoct for self-deception or wilful ignorance.[175]

Such a social phenomenon, affecting a cohort of individuals, may have been documented. It resonates with the concept of so-called twisted self-deception, but as argued above, it is not really a matter of being 'twisted'. Betty Friedan identified and explored this phenomenon in her pioneering and world-view-shattering critique of mid-20th Century societal roles in the USA, *The Feminine Mystique*.[176]

Friedan explored the multiple societal pressures on women during nearly two post-World War II decades. These were primarily pressures to be content with fulfilling solely a domestic role in the family, not only for child-bearing years but for all of her adult life, whether she wanted to or not. There were significant barriers to, and negative sanctions against, women who might try to explore their potential outside of that restrictive role. In the early 1950s, an average of 60% of female college and university students in the USA dropped out of their degrees to get married before they were 19 years old.[177] This suggests that for many of them, tertiary education was to be taken less seriously than taking on that socially acceptable domestic role.

For many women, the pressure was *not to pursue* the positive ideals of demonstrating her intelligence, reaching her full potential and trying to be as successful as she could be. These were men's goals. Women were encouraged to limit themselves. There were multiple mass-media outlets to provide role models for happy domesticity and stereotypes of pathetic and bitter women who tried something else.[178] The wide-spread dissatisfaction amongst women, which resulted in hidden addiction and other problems, has since been acknowledged.[179] Friedan and her contemporaries first publicly expressed it. Let us see how it might play out.

It is 1954. Brenda has just quit college to marry Eddie, who is about to graduate. The plan is for Brenda to have a temporary job until the children come along. It has to be temporary because women are required to leave work if they become pregnant. By then, they will have a house in the suburbs with a mortgage, Eddie's career will be taking off, and they won't need Brenda's wages. She can be a full-time mother and housewife. Brenda decides to embrace this role and only considers this role for herself now and into the future.

Brenda and a significant number of other Brendas, Peggy-Sues, Shirleys and their contemporaries individually engaged in the same set of self-limiting beliefs. Clearly, many women freely and without conflict chose the home-maker role, and still do. For others, who

---

[174] Rorty, Amelie Oskenberg. 'User-friendly SD' 14

[175] Wieland, Jan Willem. 'Willful Ignorance.' *Ethical Theory, and Moral Practice.* 20, no. 1 (2017): 105-19.

[176] Friedan, Betty. *The Feminine Mystique* Penguin 1963

[177] Reese, Hope. 'Anger boiled up and Betty Friedan was there. The Feminine Mystique at 50' *The Atlantic* January 28, 2013. https://www.theatlantic.com/sexes/archive/2013/01/anger-boiled-up-and-betty-friedan-was-there-feminine-mystique-at-50/272575/

[178] See: *Is this TV's most toxic stereotype?* Caryn James. BBC Culture. 16 June, 2021. https://www.bbc.com/culture/article/20210615-why-tvs-most-toxic-stereotype-endures

[179] Casey, Scott. 'A cup of tea, a Bex and a good lie down' *Australian Pharmacist* June 12, 2018 https://www.australianpharmacist.com.au/cup-of-tea-bex-good-lie-down/
Dowsett Johnson, Ann. 'Alcohol as an escape from perfectionism' *The Atlantic* October 15, 2013 https://www.theatlantic.com/health/archive/2013/10/alcohol-as-escape-from-perfectionism/280482/

had perhaps different aims, there was at this time little other choice than to relinquish their personal ambitions. That this may have been self-deception, even if supported by strong social mores, was suggested ten years later.

Betty Friedan was organising a class reunion and contacted members of her university cohort who had dropped out to marry. From their comfortable suburban homes, which had everything they were supposed to desire came, one by one, the acknowledgement that they had believed there should be more, they could do more, and they could contribute outside of their roles as mothers and housewives. Each of these women contented themselves with limiting their expectations and denying their ambition to do something else.

The situation was reflected across English-speaking Western countries, including Australia, and persisted until the 1970s, when social changes brought in changes of attitudes.[180]

This example raises the question of collective or group self-deception. This question has attracted some attention from philosophers, but there is more to be said in this area. In addition, it suggests an answer to the question whether so-called 'twisted self-deception' is more likely to be 'desiring a negative' when an individual or group experiences strong barriers against achieving their aims in a particular social situation.

Mele has explained much about biased belief-formation, and has proposed a theory which attempts, from an empirical basis, to explain self-deception without resorting to either paradoxical doxastic explanations or, as he puts it, 'mental exotica'. His theory has with-stood certain challenges, including some from Intentionalism. A major concern is that while Mele explains much, his theory is overinclusive, not allowing us to differentiate self-deception from like phenomena. A further issue is the selection issue, which Mele's theory cannot satisfactorily explain.

It is unfortunate that Mele has accepted the challenge from so-called 'twisted self-deception'. I have taken the position that there is no such phenomenon as 'twisted self-deception', as any proposed examples seem to be psychopathology or ordinary self-deception. Mele's attempts to explain the twisted variety detract from his main theory and give the impression of being ad-hoc explanations. The examples offered of this concept introduce new confusions between self-deception and various pathologies, which exacerbates the overinclusiveness problem. So-called 'twisted self-deception' I have argued, is more a result of a misapplied value-system, one which does not recognise that it is possible for people to desire a negative, nor that indeed, some individuals are pressured into this due to society's current mores. It would have been better to re-assert the structure of self-deception, whether the desire is positive or negative.

The result of this analysis is that Mele's theory greatly contributes to our understanding of self-deception. It avoids certain issues arising from a purely doxastic approach to self-deception. However, due to the overinclusiveness issue, and in relation to the selectivity issue, it has less explanatory power than one particular type of Intentionalism; that described by Davidson.

[180] Rowland, Robyn. *Australian attitudes to the traditional cultural sex role stereotype*. Doctor of Philosophy thesis, Department of Psychology, University of Wollongong, 1979. https://ro.uow.edu.au/theses/1649

The explanatory power of the three major theories of self-deception

I have considered these prominent representatives of three major theories regarding self-deception, presenting their main arguments and identifying some concerns with each of them. Earlier, I extracted a number of 'desiderata' for a theory of self-deception from four standard paradigm examples of self-deception.

The next task is to assess whether, and to what extent, the major theories meet the identified desiderata for a theory of self-deception, with a view to identifying the theory with the most explanatory power.

The table below indicates how each theory meets each of the desired requirements for a theory of self-deception, and also shows whether and how each theory explains the paradigm examples.

Table Showing Theories of Self-Deception, Explanatory Desiderata and application to Paradigm Examples.

Theory	***Existentialism: Jean-Paul Sartre***	***Intentionalism: Donald Davidson***	***Motivationalism: Alfred R Mele***
Desiderata			
Static Paradox	Acknowledged. Sartre argues that the self-deceived rely on the implicit negation internal to the notion of 'belief' as opposed to 'knowledge'. If I say 'I believe' I am also acknowledging that 'I do not believe' in so far as my belief is not certain. This allows the self-deceived to maintain their belief.	Explanation. Self-deception is an irrational attitude in a rational agent. That is, between 'p' and '~p' the self-deceiver breaks the rules of inductive inference and believes the proposition with the least evidence.	Denied. It is not necessary to explain. Self-deceivers engage in motivated biased decision-making. There is no reason to ascribe the '~p' belief to them.
Dynamic Paradox	Acknowledged. The above also explains the dynamic paradox for Sartre's theory.	Explanation. The irrationality occurs when an 'intention' (for example a desire, motivation etc.) causes the self-deceptive belief. Davidson sees this as being parallel to *akrasia* or weakness of the will and calls it 'weakness of the warrant' for the self-deceptive belief.	Denied. It is not necessary to explain. Self-deceivers engage in motivated biased decision-making. This can sometimes take a 'twisted' form in which the self-deceived convince themselves to believe a proposition about something they do not desire, but in this case they desire something associated with that desire.
Challenge to First Person Authority	Not relevant. For Sartre, while I am always responsible for myself, relationships with other people, 'the Other', are always difficult. The 'Other's' view of oneself constantly challenges and 'transcends' one's ability to grasp it.	Explanation. Others attribute rationality to the self-deceiver and see the irrationality in the self-deceiver's position.	Does not explain. The implication is that others who do not have the motivation can see the biases.

Selection Issue	Not relevant. Everyone is capable of and indulges in self-deception or bad faith. Bad faith can become a way of living. The challenge is to avoid it and live authentically. However, that requires courage.	Explained. The self-deceiver is aware of the proposition they are denying (the ~p). It is this awareness which stimulates and maintains the self-deception.	Cannot explain. Motivationalism cannot explain why a person deceives themselves about 'p' and not about other issues. Without being able to explain the 'selection' issue, Motivationalism remains overinclusive.
Tension	Explanation. The self-deceiver must take measures to maintain their self-deception. These may become habitual. Even then, there may be short-lived abrupt awakenings to the self-deception.	Implied.	Implied.
Reason or Purpose	Explanation. The over-riding purpose is flight from anxiety, freedom and one's radical responsibility for one's life.	Not stated. Davidson is interested in the structure of self-deception and not the purpose.	Explanation. Relevant lay decision-making processes operate to avoid 'costly' mistakes.
Confirmation or Disconfirmation	There are many challenges to Sartre's theory, particularly from psychoanalysis and materialist philosophies.	There are challenges to Davidson and Intentionalism from Motivationalism and theories proposing unconscious influences. These raise their own explanatory difficulties, so are not conclusive.	Does not seem possible due to inductive arguments. Mele has at least once amended his proposed sufficient conditions on an *ad hoc* basis to include 'twisted self-deception'. Is there a limit to these amendments? The theory may expand to absorb challenges.

Relevant Paradigm Examples from this thesis.	**Sartre**	**Davidson**	**Mele**
		Emily believes Daniel is faithful	
	Sartre's issues with the 'Other' may mean that without significant additional explanation, it is difficult to construe this as self-deception	Davidson accounts for this example by acknowledging that Emily must be aware of evidence he is unfaithful and this stimulates the self-deceptive belief.	This paradigm example does not seem to be fully explained without the selection criterion.
		Ryan believes Amy is playing hard to get	
	As for Emily and Daniel, this example may require additional explanation for it to be seen as self-deception.	Davidson identifies similar examples as being paradigmatic. Ryan must be aware of Amy's apparent avoidance of him (the ~p proposition) which stimulates this belief.	Mele acknowledges similar examples as being paradigmatic. His motivated desire explanation accounts for this example.
		Liam believes he will be a great writer	
	Liam is focusing on the 'freedom-consciousness' aspect of being and denying 'contingency'	Admittedly this paradigm example is close to wishful thinking or delusion and the '~p' belief might not be a mere contradiction. However, Davidson's theory can explain this example with some reservations.	Mele's motivated desire explanation accounts for this example.

Evelyn believes her chest pain symptoms are 'just part of getting older'		
	According to Davidson, Evelyn must be aware of the '~p' proposition, as she has the sensations of chest pains	Mele's motivated desire explanation accounts for this example. However it is difficult not to see that Evelyn must be aware of the ~p (that the pains indicate heart problems). Otherwise she is just mistaken.
Sarah believes she cannot lose weight		
Sarah is focusing on 'contingency' and denying her 'freedom'.	Davidson explains this example without recourse to a 'twisted' explanation. Sarah is aware of the 'p' but believes '~p'.	This example would be construed as twisted self-deception but I argue it is not twisted self-deception but is self-deception.
Peggy-Sue believes she is content with giving up her studies for a domestic role for life.		
Peggy-Sue is focusing on 'contingency' and denying her 'freedom'. Sartre's approach allows for the same self-deceptive belief to apply socially or collectively.	Davidson explains this example as being aware of the '~p' but believing the 'p' proposition.	This example would be construed as twisted self-deception but I argue it is not twisted self-deception but is self-deception.

Conclusion to Part Two – Self-Deception

The question to be answered concerns the relationship between the concepts of wilful ignorance and self-deception. I am also considering whether these phenomena are examples of self-induced, purported ignorance.

I noted that this task was not assisted by the fact that, while much has been written on the topic, there is no agreed definition of self-deception. It was further noted that rival theories of self-deception, specifically Intentionalist and Motivationalist theories, have reached an impasse in argument – a concerning state for philosophical argument to be in. This situation means that the present task becomes one of assessing whether, and if so, how a relatively closely defined topic (wilful ignorance) is related to an undefined, controversial, topic (self-deception). It seemed that it was necessary to decide on a preferred theory of self-deception if the task was to have any chance of success. In this case the choice would be for the theory that had the most explanatory power.

I then proposed four paradigm examples, which are standard in the literature. These served as a basis for extracting 'desiderata' or indicators by which to assess the explanatory power of various theories of self-deception.

Following an overview of the variety of theories of self-deception, it was determined that the major theories are Existentialism, Intentionalism and Motivationalism. I selected a prominent representative of each theory for further analysis.

Above I present a table with the results of that analysis. Ultimately, the results indicate that there are good reasons to prefer the Intentionalist position, as represented by Donald Davidson, as having the most explanatory power in relation to self-deception. This is not, however, a statement about validity. All these theorists deepen our understanding of self-deception.

Existentialism, as represented by Sartre, cannot be discounted. Sartre provides us with classic, memorable scenarios of self-deception which epitomise the human situation. His explanation of belief and the mind manipulations of the self-deceived are telling. His theory of bad-faith, or *mauvaise foi*, arises from his ontological theory, including his theory of relationships with others. This takes us beyond the boundaries of this thesis, while affecting his concept of bad-faith.

Motivationalism, as represented by Mele, emphasises the fact that self-deception is no intellectual, rational exercise. The self-deceived are motivated by strong desire and emotions, which influence their beliefs. Yet so are those engaging in wishful or magical thinking and in some cases, even those who are merely mistaken. Mele's focus is too wide. Despite his major contribution in refusing to focus on the doxastic issues of the so-called paradoxes of self-deception, Mele does not allow us to distinguish self-deception from other like phenomena.

The conclusion to this investigation of theories of self-deception is that Intentionalism, as articulated in the theory of Donald Davidson, provides the theory with the most explanatory power. Davidson's explanation is succinct, is not focused on doxastic puzzles, and is the one that most approaches a definition in its clarity. His theory accounts for the static and dynamic paradoxes, and the challenge to first-person authority. Importantly, his theory explains the selectivity issue. He acknowledges the role of desire or emotions in self-deception, so is not over-rational in his explanation. Finally, despite some reservations identified with his own examples, his theories provide a high level of explanation for the paradigm examples identified earlier.

Conclusion Additional Issues

Self-deception is one area of irrationality in an otherwise rational person.

For a person to engage in self-deception, the issue it has to be about is something that is:

- Not clearly resolvable. That is, there is evidence for p and ~p but in neither case is there enough evidence available to the self-deceived to count as knowledge.
- Something important enough for them to care about it.
- Something that negatively affects their psyche, or related areas such as self-concept or way of living.

It is not about avoiding an 'unwelcome truth'. The propositions in self-deception do not have the status of full truth. The propositions which are avoided are not merely unwelcome. They stand to change or put at risk something which is closely associated with the self-deceiver's psyche, or their way of living. Further, the issues involved are unlikely to be something over which the self-deceiver has control. Finally, the self-deceiver must in some sense have some awareness or understanding of the avoided or disavowed proposition and its potential effect. We have seen Davidson's admission that some sort of divided mind is necessary for self-deception. It is also argued that the type of awareness, or Kahnemann's 'System 1' processes allow us to foresee what is likely to happen. This may be the source of the '~p proposition', and its avoidance in self-deception. As well, we noted that while 'clear eyed' change strategies involving conflicting beliefs seek change, it is unlikely that the self-deceived seek either change or insight.

I am proposing, based on the above analysis, that in many instances, self-deception is an attitude aimed at preservation and conservation taken towards something that puts at risk the self-deceiver's psyche, self-concept, or something closely related to them. The risk may be a threat, or a shock, which would mean that in some cases, one aspect of self-deception is that the self-deceiver might be 'buying time' to accustom themselves to it.

This explanation is relevant to the paradigm examples.

If Daniel is being unfaithful to Emily, her life will change dramatically. This is a life that they have built together for more than three years. Emily's status will change from 'wife' to divorced woman. The lifestyle and assets she now enjoys will be halved. Her future will be different from what she now plans. The children she sees in her future will not look at all like Daniel. That is, assuming she re-establishes the possibility of relationship and family. The emotional upheaval caused by Daniel's betrayal may be far-reaching: hurt, grief, anger and unwillingness to trust again. Breaking ties and starting again with another partner will require huge resilience.

If Evelyn's chest pain is confirmed to be symptoms of a heart condition, her life will change dramatically. She will no longer be the active senior and loved grandmother she sees herself to be. There will be physical limitations and she risks becoming an invalid. She may feel she is not a help to her family, but a burden on them.

Liam relies on and needs the adulation of his local writing critique group to shore up his self-concept as a great writer, albeit one plagued by writers' block. A rejection or serious criticism from an editor or publisher would threaten that self-image. He would be a 30-something with a casual job, rather than a creative person with a promising future. In social situations, his response to the question, 'What do you do, Liam?' would be very different.

Ryan will cease to be a man in love and just be another, rather overwhelmed, undergraduate.

I now turn to Mele's examples. Gordon's parents will suddenly become the parents of a traitor rather than a son in an honourable profession dedicated to protecting the country. Apart from their fear for their son's fate, that will reflect on them, on their parenting, but also will

drastically change any hopes they had for their son's future. Rather than having an easy parental pride, their focus will be on clearing his name. Instead of hosting family dinners for their son, his wife and their grandchildren they will be visiting him in jail. Instead of being proud of his achievements, they will avoid mentioning him. His shame taints them personally.

Sarah will have to face criticism from herself and others for not watching her weight. She will be back on the treadmill, avoiding social occasions, watching what she eats and having to account for everything she does.

The women in the 1950s cohort who graduated knew, in choosing to leave college and forego the career their studies might lead to, they were choosing limitation, for years. Under pressure of social mores, they were only able to admit this years later. The realisation may bring deep emotions of resentment and questions of 'what if?'

Bishop Butler's imminent bankrupt who refuses to look at their finances is holding onto a lifestyle which they cannot afford. The alternative, poverty, or even being reduced to the middle class, is unthinkable for them.

Although Sartre has characterised bad faith as a 'flight' from the freedom and responsibility for our lives, his examples have some of the same flavour. The young woman on a date is delaying making a decision she knows she must make. That decision stands to change her life completely. She is right to delay until she knows this is the right decision. When Sartre was writing, in wartime France, the wrong choice could be devastating.

Each one of these self-deceivers understands the implications of the '~p', the 'proposition' they are avoiding. In some sense of the term, they 'know' what they are avoiding and they ascertain the risk of the negative proposition. It would bring about a change in their psyche, their self-concept or the life they have established for themselves.

Having identified the theory with the best explanatory power and differentiated the key aspects of that theory, it is possible to focus more precisely on the question of this thesis. The question concerns the relationship between wilful ignorance and self-deception, where 'self-deception' is understood to be a phenomenon characterised in Davidsonian-style Intentionalism.

I will now move on to Part Three of this thesis, in which I will consider in detail the phenomenon of wilful ignorance.

Following that, I will be in a position to propose an answer, as far as the available theories allow, concerning the relationship between wilful ignorance and self-deception. I will also consider what implications the answer might have.

Part Three

Wilful Ignorance

On the understanding that the structure and phenomenology of self-deception is most accurately depicted by Donald Davidson's account, I will now consider wilful ignorance with a view to answering the question of its relationship to self-deception.

As is the case with self-deception, I will begin by considering paradigm examples. Perhaps one of the most cited examples of wilful ignorance is that of Albert Speer.

Paradigm Examples

Albert Speer and wilful ignorance

In Hitler's Germany, Albert Speer was one of the highest ranking Nazis and Hitler's favoured architect. As such, Speer designed the monumental buildings he intended both to be a tribute to Hitler and even in ruins to remain powerful reminders of the 'Third Reich'.[181] He was also responsible for designing at least some of the concentration camps. Later, in his capacity as Minister of Armaments and War Production, Speer used slave labour from the concentration camps to meet the demand for arms. At his trial in Nuremberg, he succeeded in distancing himself from actual knowledge of the atrocities that happened in the concentration camps, claiming that he made requests for labourers, but actual procurement was the responsibility of others.[182]

On 5 June, 1977, Speer swore an Affidavit in Munich, concerning his awareness of concentration camps in Nazi Germany. The English translation of the Affidavit was sent to Speer for his approval. Speer took issue with the translation of the German word *Billigung* as 'tacit consent', advising that the closer meaning was 'looking away'.[183] As discussed elsewhere, this is an aspect of wilful ignorance.

In that Affidavit, Speer related what his friend Karl Hanke told him about concentration camps, as follows:

---

[181] See information on the value of ruins for Speer and others at https://en.wikipedia.org/wiki/Ruin_value

[182] Nuremberg Trials transcripts, Speer's testimony. The Avalon Project. Yale Law School Library. https://avalon.law.yale.edu/imt/06-20-46.asp as one source of the Trial's transcripts.

[183] The Affidavit was requested by D Diamond, Director of the South African (Jewish) Board of Deputies, asking Speer to assist in their law case against the publishers of a holocaust-denying pamphlet *Did Six Million Die? The Hoax of the Twentieth Century.* Sereny, Gitta. *Speer's Battle with Truth* Independent (online) 9 September 1995. https://www.independent.co.uk/arts-entertainment/speer-s-battle-with-truth-1600342.html

In Summer 1944, the District Leader of Lower Silesia, Karl Hanke, paid me a visit. Hanke had distinguished himself by bravery in the Polish and French campaigns. He was certainly not an easily frightened person. Therefore it was of particular moment, when, at that time, he told me in a shocked manner, that monstrous things were happening in the concentration camps of his neighbouring district, Upper Silesia. He said he was there and would never be able to forget what atrocities he had witnessed there. Admittedly, he did not mention any names, but he must have meant Auschwitz in Upper Silesia. From the agitation of this battle-hardened soldier, I could derive that something unheard of was happening, if it could cause this old party leader of Hitler's to lose his composure.[184]

Affidavits are required to be limited to fact, not reflection and interpretation. This recounting of the meeting with Hanke is more detailed and specific than what appears in Speer's memoirs (discussed below). As stated here, Hanke specifically told Speer that 'monstrous things' were happening in the concentration camp and that he had witnessed 'atrocities'. Speer, who had been friends with Hanke since the early 1930s, relied on his knowledge of Hanke, including his military service and ruthlessness,[185] to assess the severity of what Hanke had witnessed. Describing Hanke's demeanour with such terms as 'shocked', 'agitation' 'never be able to forget' and 'lose his composure' highlights that the non-verbal, non-reflective aspects of the meeting communicated as much as the words. It is clear that Speer believed Hanke.

The information in this Affidavit only partially confirms what Speer had written regarding this incident some years earlier, in his memoirs. There, Speer recalls that after visiting a concentration camp, Hanke advised him never to accept an invitation to inspect one under any circumstances.

This often-quoted passage is frequently interpreted as an example of wilful ignorance:

I did not query him, I did not query Himmler, I did not query Hitler, I did not speak with personal friends. I did not investigate – for I did not want to know what was happening there … During those few seconds, while Hanke was warning me, the whole responsibility had become a reality again … For from that moment on, I was inescapably contaminated morally; from fear of discovering something which might have made me turn from my course, I had closed my eyes. This deliberate blindness outweighs whatever good I may have done or tried to do in the last period of the war … Because I failed at that time, I still feel, to this day, responsible for Auschwitz in a wholly personal sense.[186]

It is clear that Speer correctly anticipated what he was likely to be told if he questioned Hanke further, or if he questioned Himmler or Hitler. He correctly anticipated he was likely to find atrocity if he visited a concentration camp. Speer had good reason to trust Hanke. Yet, it seems that he believed he could distance himself and claim ignorance, if he did not investigate further. He admits the reason he did not investigate was that he did not want to know what was happening there.[187] As well, he did not want to 'veer from (his) course'.

---

[184] The Nuremberg Trials: Testimony of Albert Speer. Jewish Virtual Library. American-Israeli Cooperative Enterprise (AICE) website. 1998-2021.
https://www.jewishvirtuallibrary.org/nuremberg-trial-testimony-of-albert-speer#N_2_

[185] Hanke's career included acts of brutality which earned him the moniker 'The Hangman of Breslau'. This emphasises how difficult it must have been to 'shock' him as described by Speer. See https://en.wikipedia.org/wiki/Karl_Hanke

[186] Speer, Allbert. *Inside the Third Reich*. New York: Macmillan 1970 p356

[187] At the Nuremberg Trials Speer successfully claimed ignorance, distancing himself from the atrocities. This was a consideration in his avoiding the death penalty. Yet, it is clear that his ignorance was carefully maintained. Arguably, the purpose was to enable him to avoid the ultimate penalty.

This passage also shows that Speer had considerable self-knowledge. He knew that he could not personally witness atrocities and keep a clear conscience. He knew that without a clear conscience he might 'veer from his course'. He liked his life with the power his favoured position under the Nazis conferred, along with the opportunity to pursue his architectural ambitions. Speer projected his understanding into the future and saw that what he had worked for – his status, position, and name, now and in perpetuity – would be threatened and morally tainted, unless he could maintain ignorance of the atrocities in the concentration camps.

Speer is unlikely at this point to have foreseen the Nuremberg trials, where his claimed ignorance provided a defence. More immediately, ignorance would excuse his participation in the Nazi project at high levels. He did not further investigate because his intentional ignorance allowed him to live the life he was leading but personally confronting the atrocities would mean he could no longer in conscience do so.[188]

Hannah

We met Hannah briefly at the beginning of this thesis. Here is her story in detail:

> Hannah is a woman with a good job in a large modern city. She needs to buy new clothes for work every season. Unlike her mum, she never has to iron because her clothes are synthetic, and synthetics are so much better than they used to be! Hannah has an optimistic outlook and genuinely believes that things are improving for everyone. She does not feel guilty for buying her seasonal wardrobe for work, because synthetics and 'fast fashion' mean that nearly everyone can afford clothes and offshore manufacturing means work for people in socially disadvantaged nations.
>
> When the government regulates against single-use plastic, Hannah does some internet research on the effects of plastic manufacture and use on the environment. She can't be part of that, so she decides to eliminate single-use plastic from her life. However, Hannah had not realised that plastic was everywhere and she had to become quite inventive to reduce her use of it, aiming at eventual elimination. After many months, Hannah succeeds.
>
> Chatting in the tea-room at work, a colleague mentions the social and environmental impact of 'fast fashion'. Hannah quickly leaves the room. She recalls her recent experience in eliminating single-use plastic from her life. Hannah knows that if she were to change her wardrobe so that she only bought natural fibres and clothes designed and made locally, where workers rights are protected by strong labour laws, her budget in time and money would be blown. She would even have to buy an iron and spend hours making her clothes presentable for work. Hannah sees this in an instant. She does not want to feel guilty about the garments she wears, but she can't feel guilty about something if she does not know about it.

In leaving the room and remaining ignorant, Hannah indicates that she has a considerable amount of worldly knowledge and self-knowledge.

She knows she wears 'fast fashion' and that she has made certain assumptions about it. The fact that it provides work for workers in lower socio-economic nations is, to her, a positive. Not that she believes the working conditions are optimal or comparable to those in Western

[188] Husak argues that it is for this reason that Speer is different from the other Nazi leaders. If he had known, he would not have acted as he did.

nations, but a job provides some measure of self-sustaining and independence, and so is better than having no income.

Hannah knows from experience that if her investigations revealed that fast fashion causes damage to the environment or sustains unconscionable working conditions it would play on her conscience. She knows that if she were to discover those things about 'fast fashion', then she would feel guilty about buying her clothes. She would likely put herself through a difficult process of 'fast-fashion elimination', along the lines of the process she has already undertaken in relation to single-use plastic. It would impact on her life. She is not wealthy, and because of her job she is already time-poor. Buying natural fibres made locally would impact her finances and time spent preparing for work. She does not want to do that.

She anticipates from the start of the conversation, before she leaves, what she is likely to hear and it is this which triggers wilful ignorance.

Hannah is intelligent, aware and engaged, both in her personal life and in relation to social or political issues. Her 'ignorance' is not complete 'not-knowing'. Nor is it a matter of recklessly or negligently failing to inquire, nor a matter of not caring. It is understandable that people may themselves be experiencing pressures which render them unable or unwilling to make inquiries about aspects of their lives or consumption. It is regrettable that some people simply do not care. Hannah is not uncaring.

While Hannah would not be prosecuted for wearing fast fashion, it might play on her conscience. If she doesn't know, then she is absolved of responsibility and her conscience is clear. Wilful ignorance serves to excuse behaviour that she would find morally unacceptable and allows her to live her life with a clear conscience. Like Speer, if she were confronted with the social and environmental issues related to fast fashion, she would act differently.

Drug transporting cases

United States v Jewell

In cases involving drug mules or transporters, defences of ignorance are sometimes discounted in favour of a finding of wilful ignorance. Wilful ignorance does not offer a defence and this finding leads to conviction.

One paradigmatic case is *United States v Jewell*.[189] Charles Demore Jewell was convicted of knowingly transporting marihuana into the United States from Mexico. Jewell was paid $100 to drive a car across the border. When intercepted at the border, the illegal drugs were found in a secret compartment within the car.

In his defence, Jewell claimed that he did not know that illegal drugs were in the car. He claimed that while travelling with a friend in Mexico, they visited a bar. There, a man they had never met before offered them the $100 to drive a car across the border into the USA. Jewell's friend declined. Jewell accepted.

Jewell's defence was not accepted. It was determined that his ignorance was wilful and he was convicted.

[189] *United States v. Jewell*, 532 F.2d 697 (9th Cir. 1976). Text of judgment found at Justia: https://law.justia.com/cases/federal/appellate-courts/F2/532/697/99156/

United States v Heredia

A case which is sometimes contrasted with *Jewell* is *United States v Heredia*.[190] The facts of the case, as described in the Harvard Law Review,[191] are reproduced below:

> *Carmen Heredia's troubles began when she agreed to drive her mother and aunt, accompanied by two of her children, from Nogales, Arizona, to her mother's home in Tucson. On entering the car, which was owned by her aunt, Heredia noticed a strong smell of detergent. Her aunt told her that she had spilled fabric softener in the car a few days earlier, an explanation that seemed unlikely to Heredia. As she was driving, Heredia noticed that her aunt and mother "both appeared nervous" and were "acting strangely." Knowing that her mother's boyfriend used narcotics and that her mother and aunt possessed unusually large amounts of cash, Heredia began to suspect that there were illegal drugs in the car. By this point, however, she was on an interstate highway and could not investigate further without stopping the car on the shoulder of the road. At a border patrol checkpoint, agents alerted by the strong odor of detergent emanating from the car found approximately 350 pounds of marijuana in the trunk. The marijuana was covered with dryer sheets — a method used by drug traffickers to mask the pungent smell of the drug. Heredia was arrested and charged under 21 U.S. C. § 841(a) with knowingly possessing contraband with the intent to distribute.*

The exculpatory factors here include Heredia's family relationships. Trust, or a desire not to injure relationships are strong incentives to believe the best interpretation and not to ask questions. A further issue is safety, if we believe Heredia's claim that she did not suspect anything until she was on a dangerous part of the highway. Given these aspects of the case, we cannot help feeling that the judgement against her in Carmen Heredia's case may have been harsh, or indeed, unjust. That is, perhaps being overly-trusting of her family, she actually was ignorant of the drugs until it was dangerous to raise the issue.

Definition

In the Introduction it was suggested that the derivation of the concept of wilful ignorance from the law may assist in providing a definition of wilful ignorance (which is sometimes also called wilful blindness.)[192] In particular, because there are serious consequences for a defendant in a finding of wilful ignorance, hopefully the concept can be clearly defined, in the interests of justice.

Criminal culpability in law depends on proving both the criminal act or *actus reus* and criminal intent or *mens rea*. Actual ignorance is a defence in relation to *mens rea*.[193] Deliberate preservation of ignorance, in order to be able to claim ignorance as a defence in a criminal case is wilful ignorance. In order to avoid this device, courts have held that *mens rea* can be found when the defendant either had actual knowledge, or engaged in wilful ignorance in relation to the relevant situation. The following quote by Glanville Williams is frequently cited:

---

[190] *United States v Heredia* 483 F.3d 913 (9th Cir. 2007)

[191] Harvard Law Review Editors. *United States v Heredia.* Feb 1 2008 121 Harv L Rev p1245 https://harvardlawreview.org/wp-content/uploads/pdfs/us_v_heredia.pdf p1246

[192] As noted earlier, there appears to be some connotative difference, but these terms are in practice used inter-changeably.

[193] Note that ignorance of the law itself is not a defence. If a defendant does not know that in this jurisdiction, robbery or murder is a crime, they will be convicted anyway. It is actual ignorance that the act in question occurred that provides the defence. If a defendant did not know her companions were committing robbery while she waited in the car, that there were people in the building she demolished or that the mushrooms she put in the dinner were poisonous, she can plead ignorance in her defence.

A court can properly find wilful blindness only where it can almost be said that the defendant actually knew. He suspected the fact; he realized its probability; but he refrained from obtaining the final confirmation because he wanted in the event to be able to deny knowledge. This, and this alone, is wilful blindness. It requires in effect a finding that the defendant intended to cheat the administration of justice.[194]

This definition imposes a high level of 'belief'. Importantly, it includes a statement of the motivation of the defendant. The defendant intended to cheat the administration of justice. As suggested earlier in the discussion of Speer and Hannah, a motivation in wilful ignorance is to provide a defence in law, or an 'excuse' which will allow one to avoid moral disapproval from others and one's own conscience.

This paragraph was quoted in relation to the rationale for culpability in findings of wilful ignorance by the High Court of Australia (HCA), in *The Queen v John Edwin Crabbe (1985)*.[195]

The HCA stated at paragraph 12:

When a person deliberately refrains from making inquiries because he prefers not to have the result, when he wilfully shuts his eyes for fear that he may learn the truth, he may for some purposes be treated as having the knowledge which he deliberately abstained from acquiring.

A satisfactory definition of wilful ignorance will address three issues. The first is the knowledge question. Related to that are issues concerning access to knowledge. Finally, there is the motivational issue.

The 'knowledge' issue in wilful ignorance

There are two formulations of the knowledge question. The more general, by reference to the examples above, is concerned with the level of belief a subject must have in order to be engaging in wilful ignorance. This is a 'doxastic' question. The question more specific to legal philosophy concerns whether the law is equating wilful ignorance with knowledge, or substituting wilful ignorance for knowledge. The further question is whether that serves the interests of justice, but this is outside of the parameters of this thesis.

There have been a number of attempts to answer the doxastic question. In *R v Crabbe* the HCA was asked to rule on whether the lower court judge was in error when he reportedly advised the jury that wilful blindness could be found if the defendant was aware of a 'possibility' that there were people in the bar. The HCA stated that a 'possibility' was not strong enough to indicate wilful blindness.

Lynch,[196] referring to the Speer example above, argues that Speer had an 'inkling' of the truth about concentration camps, rising to a 'suspicion'. The Williams quote above adds to the notion of 'suspicion' a realisation of 'probability'. Douglas Husak[197] requires the 'suspicion' to be warranted, which excludes suspicions brought about by mistake or delusion.

---

[194] Williams, Glanville. *Criminal Law: The General Part*. 2nd ed. 1961. 159.

[195] *The Queen v John Edwin Crabbe (1985) 156 Commonwealth Law Reports (CLR) 464* 26 March 1985. (*R v Crabbe*)
In this case, the defendant, a road train driver on an overnight stop, had been evicted for nuisance from a motel bar near Yalara in Central Australia. Under the influence of alcohol, he drove the cabin and one trailer of his road train into the bar, demolishing a wall, killing five people and injuring others. There had been appeals and two retrials before the matter was brought before the HCA.

[196] Lynch, Kevin. Self-deception and Wilful Ignorance. 505.

[197] Husak, Douglas. *Phil Crim Law* 209.

In *Jewell* the Appellant Court, relying on the US Model Penal Code, confirmed the level of belief should be 'awareness of a high probability'.[198] The judgement stated:

…(D)eliberate ignorance and positive knowledge are equally culpable. The textual justification is that in common understanding one "knows" facts of which he is less than absolutely certain. To act "knowingly", therefore, is not necessarily to act only with positive knowledge, but also to act with an awareness of the high probability of the existence of the fact in question …

…This is the analysis adopted in the Model Penal Code. (In that code) Section 2.02(7) states: "When knowledge of the existence of a particular fact is an element of an offense, such knowledge is established if a person is aware of a high probability of its existence, unless he actually believes that it does not exist." (Model Penal Code 27 (Prop. Official Draft 1962)).

Alexander Sarch proposes the following level of knowledge for wilful ignorance:

Belie(f) that there is a sufficiently high likelihood that p is true, short of practical certainty.[199]

Access to Knowledge

There are circumstances in which, even if a person had a belief amounting to a high probability that p is true, it would be unfair to expect them to investigate. This applies when the information is highly specialised, when it is onerous to investigate or when investigation involves risk to health, or danger to oneself or family.[200]

Motivation

In relation to the criminal law, the motivation for wilful ignorance must be that the individual wanted to be able to claim ignorance in order to avoid responsibility, blame or a conviction at court. A key aspect of the motivational factor in relation to the law is that it is conscious. It is also a cynical claim of ignorance.

The cynicism is not necessarily a strong feature of the motivation for wilful ignorance outside of the criminal law. Hannah, for example, may be more interested in preserving her lifestyle with a clear conscience than in persuading anyone else about her belief.

It is the issue of motivation that appears to distinguish Heredia's case from that of Jewell. Jewell's failure to inquire is cynical, while Heredia's suspicion grew slowly on seeing signs that something was not right. According to her account, she did not initially suspect anything. If this is the case then it is questionable whether Heredia was engaging in wilful ignorance.

Proposed definitions

Husak, relying on the three aspects of wilful ignorance mentioned above, proposes the following definition of wilful ignorance:

(A) defendant is willfully ignorant of an incriminating proposition p when he is suspicious that p is true, has good reason to think p true, fails to pursue reliable, quick, and ordinary measures that would enable him to learn the truth of p, and, finally, has a conscious desire to remain ignorant of p in order to avoid blame or liability in the event that he is detected.[201]

Allowing for these arguments, Sarch amends his earlier definition and now proposes:

---

[198] *United States v. Jewell*, 532 F.2d 697 (9th Cir. 1976). Text of judgment found at Justia: https://law.justia.com/cases/federal/appellate-courts/F2/532/697/99156/

[199] Sarch. *Criminally Ignorant* 18.

[200] See, for example: Sarch. *Criminally Ignorant*. 112ff , Husak. *Phil Crim Law.* 209.

[201] Husak D. *Phil Crim Law.* 209

(Wilful) ignorance may substitute for knowledge only when the defendant:

1) *had serious suspicions that the inculpatory proposition was true,*
2) *reasonably could have learned the truth but deliberately didn't, and*
3) *was not aware of anything that helped justify the decision not to learn more.*[202]

This new definition is expanding towards a statement of conditions, but nevertheless ties these conditions to criminal court proceedings by the use of the word 'defendant'.

The definition gives us this: the 'suspicion' must be 'serious', so it avoids the doxastic difficulty, concerning whether a 'vague suspicion' or even an 'inkling' would comprise wilful ignorance. It also avoids the issue that Sarch has identified, that the wilful ignorance doctrine attributes knowledge to claimed states of ignorance. It covers those conditions which have caused concern regarding the justice of cases featuring wilful ignorance, as for example, in cases such as *Heredia.*

An attempt to apply this definition outside of the courtroom to a more general situation, by substituting the term 'person' for 'defendant', may result in the definition being overinclusive. The courtroom location establishes the proximity of the wilfully ignorant to the matter at hand, which is identified as the 'inculpatory proposition'. Remove that location and the definition requires some alternative wording concerning proximity. There is also a problem concerning the determination of which matters the wilfully ignorant should have investigated.

While the definitions derived from court decisions and reasons clarify the concept of wilful ignorance with some authority, they are also necessarily made in hindsight in an attempt to reconstruct the defendant's state of mind. Further, they concentrate on doxastic issues, as the *actus reus* is admitted and therefore a given.

However, wilful ignorance takes place in the present. Nor is it the case that every instance of wilful ignorance involves an attempt to evade criminal justice. It is plausible that the majority of instances of wilful ignorance concern one's own moral standards or shared social values, as in Hannah's case. Finally, outside of the courtroom, any definition will have to connect the 'act' and the 'belief' aspects of wilful ignorance. If this connection is not made, clearly any definition will be overinclusive, as there are all types of issues about which we may, trivially or otherwise, choose to remain ignorant.

Although the consideration of wilful ignorance has now moved outside of the courtroom, it is useful to distinguish wilful ignorance from 'culpable ignorance', also called 'negligence'. In criminal matters a plea of ignorance provides a defence to *mens rea.* In wilful ignorance, that defence does not stand because wilful ignorance is contrived ignorance. By contrast, in negligence, whether considered as a tort or whether it rises to criminal liability, ignorance is not a defence.

Negligence: when ignorance is not a defence

The distinction between negligence and wilful ignorance is sometimes raised in the literature on wilful ignorance.[203, 204] This can be especially confusing if wilful ignorance and negligence are joined under the classification of 'culpable ignorance'.

In wilful ignorance one claims ignorance and, if this claim is believed, it is a defence to culpability regarding the 'knowledge' factor, which is necessary to establish *mens rea* in law.

[202] Sarch. *Criminally Ignorant.* 264.
[203] See: Sarch. *Criminally Ignorant*. Ch 6 and Lynch, Kevin 'Willful Ignorance and Self-Deception.' 508ff.
[204] In some texts, 'negligence' is called 'culpable ignorance'.

In negligence, ignorance provides no defence. It is irrelevant whether the act was done in ignorance, recklessly or with disregard for the duty of care one has for others. This is because negligence is determined in relation to a standard of knowledge which the defendant is presumed to have, which is defined as knowledge that one 'knew or ought to have known'. The criterion used to determine the appropriate knowledge is that of the 'reasonable person'. In relation to specialised knowledge, such as is required of the professions, standards are determined by professional bodies. In the absence of formalised standards, they are decided with reference to the accepted level of knowledge and practice in the profession at the time.

Doctors are required to know about current treatments and procedures, lawyers to be able to ascertain the likelihood a case will succeed and financial advisers to know the difference between 'blue chip' investments and speculative ones. In these cases, it is a matter of knowledge that the professional actually has or is deemed or expected to have. Failure to reach these professional standards leaves one liable to be sued.

Negligence, like other torts, has historically been a matter determined by precedent law. More recently, jurisdictions have sought to secure the understanding of negligence though enacting legislation. The Australian state of New South Wales is one such jurisdiction and that state's Civil Liability Act[205] offers a definition of negligence which applies generally and to specialised knowledge. It states:

(1) A person is not negligent in failing to take precautions against a risk of harm unless—

(a) the risk was foreseeable (that is, it is a risk of which the person knew or ought to have known), and

(b) the risk was not insignificant, and

(c) in the circumstances, a reasonable person in the person's position would have taken those precautions.

We see here an accepted definition of negligence as occurring in a case of risk of harm which was foreseeable, defined as a risk which the person 'knew or ought to have known'. This concept is applicable in civil cases where damages are sought.[206] It is important to note that the definition of negligence involves, as well as a standard of knowledge, the concepts of both 'risk' and 'harm', which limits its scope. Insignificant risks are ruled out.

There are any number of things that you and I might negligently do in which there is a foreseeable risk of causing someone harm: light barbeques in bushfire season, throw stones from the overpass onto traffic below, spike someone's drink as a joke, leave the swimming-pool gate open, take the Doberman to the park without a muzzle or a leash, to name a few. In these examples there is a risk of harm that is clear, foreseeable and easily avoided. If you or I did not know that, then we ought to have known it because the applicable standard is that any reasonable person in the situation would surely have known it.

There are of course limits on the concept of negligence. While it might have been foreseeable that a cricket ball would be hit for a six over the fence at Cheetham Cricket Club ground, it was not foreseeable that this would happen just when Ms. Bessie Stone was on the road outside and that the cricket ball would be hit across the road and strike her on the head with enough force to cause injury.[207] In the case for negligence brought by Ms. Stone, the Cricket Club was not

---

[205] *Civil Liability Act (NSW)* 2002 No 22 Part 1A Division 2 5B.

[206] See also: Blay, Samuel K N. *Torts Nutshell Books Series.* 6th Edition. Law Book Company Thompsons Reuters (Professional) Australia Ltd. NSW. 2010

[207] Corkery, Jim. 'Dangerous Sports and Obvious Risks - Anyone For Cricket?' (2011) *Sports Law eJournal.* Bond University. http://epublications.bond.edu.au/slej/12

found to be negligent.[208] An ordinarily careful person is not required to take all precautions against every possible risk.

Following this hopefully clarifying detour, I will return to a consideration of wilful ignorance and its relationship to self-deception. Following that, I will consider three views of wilful ignorance as it applies outside of the courtroom. Those include Kevin Lynch, who argues that wilful ignorance and self-deception are different psychological types, Michelle Moody-Adams, who considers wilful ignorance in relation to wrong-doing which may be socially sanctioned and Jan Willem Wieland who follows Moody-Adams' reasoning and further argues that, alternatively to Lynch's position, motive should be the determining factor in wilful ignorance, which is not distinguished from self-deception.

Wilful Ignorance and the duty to inform oneself

In *United States v Heredia*[209] the US 9th Circuit Court was asked to consider the issue of whether intention to avoid conviction is necessarily part of wilful ignorance, when they were asked to affirm the lower court's jury instruction in *Jewell*. In *Jewell*, the jury instructions included that the jury should consider this aspect of wilful ignorance in their deliberations. The jury instructions when Heredia's case was heard in the lower court did not. The appeal was denied. Thus, so-called 'judge-made law', which proceeds by precedent, can change, expand or limit the application of a concept.

The *Heredia* case raises the issue of the duty to reasonably inform oneself.[210] This is a moral duty rather than an epistemic one.[211] Failure in the reasonable exercise of this duty adds to culpability in this view.

However, there are limits on this duty. Such limits, proposed by Sarch and others,[212] include that it must not be onerous, knowledge must not be too specialised, it must not endanger the subject or their family's health, nor put them in any other danger. The courts have recognised this in relation to wilful ignorance.[213]

Wilfully Ignorant for Good Reasons?

Contrary to the view that there is a duty to reasonably inform oneself, Deborah Hellman has argued that one may be wilfully ignorant for good reasons.[214]

Hellman's examples focus on professionals. They include the lawyer who wants to vigorously defend her client, and so may not inquire as to their guilt. Similarly, the doctor who suspects their patient is on-selling prescription medicine may not make inquiries, out of concern for their patient. In these and similar instances, Hellman argues, there are good reasons for deliberately not inquiring further.

It is not the case that every deliberate refusal to inquire is wilful ignorance. We may refrain from inquiry out of concern for another's privacy, embarrassment, feelings or because it is none

---

[208] Although this case seems to speak to good sense for possible respondents in negligence actions, there is also a sense of sympathy for Ms Stone. Out-of-court, the cricket club agreed to pay Ms Stone's damages and waived recovery of their legal costs. See https://www.cricketcountry.com/articles/bolton-vs-stone-a-four-year-legal-tussle-kick-started-by-a-beautiful-straight-six-448083

[209] *United States v Heredia* 483 F.3d 913 (9th Cir. 2007)

[210] Sarch. *Criminally Ignoran.t* 110ff

[211] Sarch. *Criminally Ignorant.* 112

[212] Sarch. *Criminally Ignorant.* 119ff

[213] In *Heredia* the appeal court noted that defences of duress are available in such cases, but Heredia's legal team did not offer such a defence in the lower court.

[214] Deborah Hellman. 'Willfully Blind for Good Reason.' *Criminal Law, and Philosophy*. No. 3 (2009): 301-16.

of our business. Deliberate failure to inquire may also be a matter of negligence, to avoid danger or because obtaining information is onerous.

The examples offered by Hellman involve a confluence of legal and professional ethics requirements, which conflict with the professionals' reasons for failing to inquire. Legal practitioners have a primary duty to the Court as Officers of the Court, to their profession and arguably to the public. Even if it is a common tactic not to inquire directly of a client as to their guilt, a tactic is not necessarily wilful ignorance. Similarly, doctors are subject to a variety of legal, competence and ethical requirements. It is difficult to see how over-prescribing medication, which will not actually be used by the patient, is consistent with concern for the patient. In many professions, including increasingly in psychology and related professions, the duty is to inquire.[215]

Clearly many professions involve serious responsibilities. In individual cases, pride, ambition or misplaced compassion may be at play in decision-making. Legislative provisions and professional standards are intended to ensure that professionals adhere to an appropriate level of behaviour and ethics, even when understandable human concern prompts them to do otherwise.

It is a subtle but relevant distinction, that the professionals in Hellman's examples have made a decision to breach legal and ethical mandatory inquiry provisions. If the practitioner is disciplined, it will be for the failure to inquire, not for their knowledge or ignorance of their clients' activities. Inquiries properly and diligently made in good faith will absolve the practitioner of wrong-doing, while the practitioner's genuine ignorance of the client's wrong-doing will not be relevant. It is neither defence nor excuse. Unlike the examples mentioned earlier of Speer and Hannah, there is no personal detriment or advantage from their ignorance. The practitioners will be paid and continue to practice irrespective of what their clients are doing.

These are examples in which deliberate failure to inquire is a decision to breach law, ethics, or professional responsibilities. Actual ignorance is merely the effect of the deliberate breach.

The above discussion has considered wilful ignorance in relation to the current legal system in common law countries and the USA. I now turn to three philosophers who consider examples of wilful ignorance outside of the court system, in the light of the issues mentioned above.

Kevin Lynch

Kevin Lynch is one philosopher who has specifically addressed the question of the relationship between self-deception and wilful ignorance. He argues for them being 'different psychological types'.[216]

As an example of wilful ignorance, Lynch considers the paragraph from Speer's memoirs which was quoted above. Lynch's analysis is doxastic, focusing on the belief state required for wilful ignorance. Wilful ignorance, he argues, must not be knowledge and also cannot be complete ignorance or 'not-knowing'. He rejects legal definitions, such as those proposed by Husak, which require a high level of belief. Speer, Lynch argues, must have had a suspicion of

[215] I am aware of such a case in relation to a trainee psychologist and their supervisor. They jointly reached a decision not to inquire further into indicators of child abuse. Their stated concerns were that the client continue to receive treatment. This was not accepted. The practitioner was struck off the register of psychologists and the student not accepted for registration.

[216] Lynch, Kevin. 'Willful Ignorance and Self-Deception'. *Philosophical Studies* (2016) 173. 505–523 DOI 10.1007/s11098-015-0504-3

what was occurring in the concentration camps, but could not have known or believed the truth. Lynch determines that the wilfully ignorant must have at least an inkling of the truth, cannot have anything stronger than a 'suspicion that p', and that suspicion should be warranted.[217]

Lynch distinguishes wilful ignorance from disbelief, fantasy, delusion, psychopathy, laziness, absence of curiosity and obliviousness. He also distinguishes wilful ignorance from another form of culpable ignorance, negligence.[218] Negligent ignorance, he argues, is not willed, whereas by contrast and by definition, wilful ignorance is intentional.

Consistently with others considered here, Lynch argues that if the truth is obscured, or particularly specialised, or onerous to discover, then it is not wilful ignorance to fail to take measures to discover it.

Turning to the normative aspect of wilful ignorance, Lynch observes that the term 'wilful' has pejorative overtones of misconduct. While the terms 'intentional' or 'deliberate' may be used interchangeably with 'wilful' in some contexts, Lynch points out that they are not synonyms. While these terms have more positive applications, 'wilful', he argues, implies breaching acceptable norms. This accounts for the inextricably negative normativity of wilful ignorance.

Bringing these aspects of his analysis together, Lynch proposes sufficient conditions for wilful ignorance.[219] Although these conditions are quite detailed, and assume shared, but arguable, norms, they are convincing. However, it might be questioned whether they are more a description of the process of wilful ignorance and also, whether they improve on the more concise definitions as derived from the legal context, which Lynch rejects.

Lynch then turns to a consideration of self-deception. He identifies the common theme in all theories of self-deception as:

1. *In all theories of self-deception, S encounters evidence indicating that some true proposition, p, is true.*
2. *S strongly desires that p is not true.*

He concludes: *That is, S encounters unwelcome evidence.*[220]

Lynch considers different types of theories, first identifying 'unwarranted belief' theories. From his articulation of their position, and his identification of Mele as one such theorist, these are elsewhere called deflationary or Motivationalist accounts.[221] He characterises these theories as identifying a discrepancy between what the self-deceived believe and what they should believe, given the evidence.[222] The self-deceived do not intentionally deceive themselves, and if they were to become aware of an intention to deceive themselves, the self-deception would fail.

Next, Lynch identifies 'implicit knowledge accounts', in which the subject recognises the truth p but the knowledge is kept out of mind, shunned or ignored.[223] These accounts, as described by Lynch, seem to focus on the causes of the tension in self-deception, including evidence manipulation. As such, the individual philosophers are perhaps too diverse to be considered a group but rather, to each be describing a characteristic view of self-deception.

---

[217] Lynch. 'WI & SD.' 509
[218] Lynch. 'WI & SD.' 508
[219] Lynch. 'WI & SD'. 509-510 These conditions are reproduced in full at Appendix B.
[220] Lynch. 'WI & SD'. 513
[221] See Funkhouser. *Self-Deception.* 86ff for comparison.
[222] Lynch. 'WI & SD'. 514
[223] Lynch. 'WI & SD'. 514

Accounts which Lynch calls 'intermediate accounts' are characterised by a belief that p and ~p, and so are what others have called Intentionalist accounts. Lynch specifically mentions Davidson and Funkhouser as representatives of this view.

Lynch concludes that no account of self-deception supports the proposition that wilful ignorance and self-deception are the same phenomenon. His argument, however, is doxastic and reveals some of the difficulties involved in such an approach.

In previous articles, Lynch has expressed his preference for the Motivationalist approach to self-deception, as developed by Mele.[224] The key characteristic of Mele's theory of self-deception, as discussed in Part 2, is that self-deception is a matter of motivationally biased belief formation. Mele specifically argues against dual beliefs and intention on the part of the self-deceived. I have argued, with others, that this theory describes an aspect of self-deception but is overinclusive while not distinguishing self-deception. Lynch's finding that wilful ignorance is a different psychological kind from the phenomena described by a Motivationalist account of self-deception is not surprising. It is almost inevitable, given the differing attitudes to intention and to acknowledgement of the denied proposition.

Had Lynch used the more common terminology for what he calls 'intermediate accounts', identifying them as Intentionalist accounts, his dismissal of these theories might not have been so straightforward. These accounts, too, state that there is a discrepancy between what the self-deceived believe and what they should believe, given the evidence. Davidson, as shown in Part 2, specifically identifies a 'weakness of the warrant' in the beliefs of the self-deceived and shows how this can come about. Lynch further argues that 'there is no reason at all to attribute the beliefs that p and that ~p to the wilfully ignorant'.[225] This is a basic statement of the static paradox using an undifferentiated concept of the term 'belief'. The discussion in Part 2 of the ways the term 'belief' is used in theories of self-deception, suggests that there may be nuances to this term not evident in this statement regarding beliefs.

The doxastic approach to theory entails a consideration of levels of evidence. For a belief to be warranted, there must be some evidence. Different ranges of doxastic states and stronger or weaker levels of evidence feature in Lynch's conclusion.[226] His argument that wilful ignorance is intentional while self-deception is not, he concludes, accounts for the greater negative normative attitudes towards wilful ignorance.[227] Yet, a full consideration of Intentionalist theories of self-deception might have provided an alternative outcome. Arguments regarding behaviours of avoidance and denial in each phenomenon concern effects of the phenomena, not characteristics.

It is clear that Lynch's conclusion that wilful ignorance is essentially different from self-deception is valid, provided Motivationalist theories of self-deception are accepted. It is not so obvious that Intentionalist theories, which Lynch identifies as 'intermediate theories', are different from wilful ignorance. Indeed, the earlier discussion of Davidson's theory here in Part 2 suggests that this question remains open.

---

[224] See Lynch, Kevin. 'An agentive non-Intentionalist theory of self-deception' *Canadian Journal of Philosophy* 47 no. 6 (2017) 779-798 https://doi.org/10.1080/00455091.2017.1321909 and 'On the "tension" inherent in self-deception' *Philosophical Psychology*. 25 no. 3,. (2012) 433-450, DOI: 10.1080/09515089.2011.622364

[225] Lynch. 'WI & SD'. 516

[226] Lynch. 'WI & SD'. 521

[227] Lynch. 'WI & SD'. 521

As the question remains open, it is possible now to turn to two theorists who consider wilful ignorance outside of the legal context, and who examine how wilful ignorance operates in a societal or global context.

Michelle Moody-Adams

Michelle Moody-Adams considers the issue of 'affected ignorance'. 'Affected ignorance' is a term which she references to Aquinas and is described as: 'choosing not to know what one can and should know'.[228] This term is used inter-changeably with 'wilful ignorance'. Moody-Adams writes in this instance in the field of Moral Philosophy. She argues against the concept that cultural influences impact on moral agency and responsibility. She equally disputes that an individual's cultural background results in 'social incapacity', and that therefore they are absolved of moral responsibility. Moody-Adams rejects these approaches. Cultural limitations cannot absolve one of moral responsibility in her view.

Moody-Adams' argument turns to the issue of slavery in ancient Greece.[229] If, she argues, an individual can conceive of a negation, a 'not' or a 'no', then they can conceive that life, their social world, and society, could be different. In ancient Greece, slavery was not necessarily seen as being justified,[230] but it was perpetuated because the institution was unexamined by those who benefitted from slavery.

She then turns to consideration of the oppressive Nazi regime in pre-war Germany and during World War II. Citing Hannah Arendt's comments on 'the banality of evil', Moody-Adams reminds us that ordinary people can do evil, and can come to view their participation in it as routine. This, she claims, (amending Arendt's famous phrase) is 'the banality of wrong-doing'. *Evil can become routine; people who kill during the day may go home to protect their families at night.*[231] Moody-Adams reminds us that this is not an issue of the past, as the same can be said of current regimes which engage in torture. Yet, there is strong resistance to acknowledging these facts and this is a matter of wilful ignorance.

Such resistance was shown by social psychologist Stanley Milgram in his famous 1960s experiment,[232] Moody-Adams claims. In that experiment, ordinary people were ordered to administer electric shocks to others under direction from an authority figure.[233] Milgram found that, under pressure, ordinary people would administer the most extreme shocks to others, even when it appeared the other was in pain or could die. Provided the authority figure took responsibility, they continued. Moody-Adams questions whether the formal conclusion as explained by Milgram, tells the complete story. A reading of the experimenter's notes shows that people averted their eyes from those they were harming, and admitted that *I did not want to see the consequences of what I had done.* She sees it as being wilful ignorance, evidencing *a denial of the connection between one's actions and harm to the "victims".*[234]

Moody-Adams then differentiates four types of wilful ignorance.[235]

---

[228] Moody-Adams, Michelle. 'Culture, responsibility and affected ignorance.' *Ethics* 104, (1994) 291-309 at 296
[229] This issue is revisited, and amended, by Wieland, as discussed below.
[230] Except, Moody-Adams acknowledges, for Aristotle.
[231] Moody-Adams. 'Affected Ignorance'. 298
[232] Milgram, Stanley. 'Behavioural Study of Obedience.' *Journal of Abnormal and Social Psychology* Vol 67 No 4 (1963) 371-378
[233] Moody-Adams. 'Affected Ignorance'. 300ff
[234] Moody-Adams. 'Affected Ignorance'. 301
[235] Moody-Adams. 'Affected Ignorance'. 303

- The elaborate linguistic euphemisms which torturers use to mask the reality of the violence they do to others. These allow the torturer to deny the connection between their wrongdoing and the suffering or death of others.[236]
- The CEO of a corporation who demands that employees increase profits but does not want to know how it is done. This is wilful ignorance, an attempt to shield the CEO in ignorance of what she must anticipate is wrongdoing.
- A decision to 'ask no questions'. This example concerns a mother who accepts expensive gifts from her son, who has a modest income, while refusing to enquire where he got the money or whether he was involved in illicit drugs.[237]
- The university administrator who does not investigate sexual harassment claims because she does not believe her staff would do that. This type of wilful ignorance originates in our resistance to challenging ourselves, our beliefs and our situations.[238]

Responses to these situations are inadequate, Moody-Adams argues, if they on the one hand emphasise blame or on the other focus on socio-cultural explanations. For Moody-Adams, it is necessary to acknowledge 'the banality of wrong-doing' and the wilful ignorance that accompanies it, while acknowledging the difficulty of critically considering problematic socio-cultural assumptions.

Jan Willem Wieland

In Jan Willem Wieland's view, the essential characteristic of wilful ignorance is the avoidance of 'inconvenient' information. [239] Wieland argues against a doxastic approach in favour of a focus on motives.

Wieland acknowledges a doxastic 'puzzle' in wilful ignorance.[240] A subject S can be neither fully aware nor unaware that she is avoiding information. Yet any awareness means that S is not ignorant. Wieland acknowledges that Lynch attempted to deal with this puzzle by proposing that S has a 'suspicion'. Suspicions, Wieland points out, may be weaker or stronger and so cover a varied 'attitudinal space' between ignorance and full belief. Wieland then rejects this approach to wilful ignorance. Doxastic explanation at best describes, but does not explain, wilful ignorance.

Wieland proposes that what matters is that S is avoiding inconvenient information, but not due to external barriers. In rejecting Lynch's doxastic approach, Wieland also rejects a distinction between wilful ignorance and self-deception. Opting for the single term 'wilful ignorance', Wieland proposes that wilful ignorance is defined by S's attitudes.

Also rejected, in his view, is the notion that there is a causal link between the inconvenient information and the avoidance behaviour. One can be wilfully ignorant even without being aware of that information. The analogy Wieland offers is not answering your phone. Whether you don't answer because you don't want to talk to anyone at all, or because you don't want to talk to the person calling, the behaviour and result are the same. This may be true, but what is missing from this analogy is a consideration of motive. Motive, as was shown especially in legal cases, is essential to a consideration of wilful ignorance.

---

[236] Moody-Adams refers to reports on these practices by Amnesty International

[237] Note this example is different to the paradigm example of self-deception offered by Mele, in which a father merely refuses to believe his son is taking drugs. This example has a societal aspect while Mele's is personal.

[238] Moody-Adams refers to Mill.

[239] Wieland, Jan Willem. 'Willful Ignorance'. *Ethical Theory and Moral Practice* 20 (2017)105-119

[240] Wieland, Jan Willem. 'Willful Ignorance'. 114

Wieland acknowledges that his definition of wilful ignorance is very broad, and he is not troubled by this broadness. Otherwise, he points out, this would mean that only 'concerned consumers' can be wilfully ignorant, and this, he claims, is 'counter-intuitive'.

Here, Wieland has identified an unsatisfactory consequence of both wilful ignorance and self-deception. It is a worry for others, especially Sarch, as outlined above. It is part of the puzzle of both wilful ignorance and self-deception that neither involve complete unawareness. Wieland's response, to widen the definition to accommodate his intuitions, will not answer the problem.

It is only the 'concerned consumers' who have the possibility of being wilfully ignorant. Hannah leaves the tea-room *because* she has some awareness of what she is likely to discover regarding fast fashion and what that may entail for her. By contrast to Hannah, consider Hermione, who never asks questions and goes through her life using plastic, wearing fast fashion and destroying the environment. Whatever blameworthiness applies to Hermione, it is distinct from the blameworthiness of wilful ignorance.

In removing the causal connection between the inconvenient information and wilful ignorance, Wieland cannot address the issue of 'selectivity' in relation to wilful ignorance. His problem is parallel to the selectivity issue faced by Mele in relation to self-deception. There is no way of distinguishing blameworthy wilful ignorance from amongst all the issues, including the trivial, of which one might be deliberately ignorant.

Wieland has identified an essential attitude in wilful ignorance: avoidance of inconvenient information. Importantly, his overriding concern is normative, as is more evident in some of his other work.[241] This leads him to overinclusiveness.

On Wieland's view, wilful ignorance is ignorance due to one's own will rather than external barriers.[242] It does not occur because it is difficult to know better, but because it is 'inconvenient' to discover the truth.

Wieland then adopts for consideration modified versions of the four paradigmatic examples of wilful ignorance proposed by Michelle Moody-Adams, which we noted above. These are:

- The torturer, who masks her behaviour by using euphemisms for her actions and wants to torture to gain information;
- The banker, who refuses to hear about questionable practices in her firm because she wants to make a profit;
- The mother who does not inquire about whether her son's income is from dealing drugs because she loves her son and can't allow this to damage their relationship;
- The school administrator who refuses to consider cases of harassment at her school because it would reflect badly on the image and past performance of herself and the school.

Wieland is of the view that 'what all these cases seem to have in common is that the agent wants to remain ignorant because it is *convenient* to do so, while knowledge of p is inconvenient'.[243]

In this view, self-interest is a motive and an element of wilful ignorance. It is linked to one's self-concept and to a desire to maintain one's lifestyle or goals. It applies currently, or may be forward-looking, or backward-looking. That is, the wilfully ignorant can foresee issues in

[241] Wieland, Jan Willem. 'Willful Ignorance and Bad Motives.' *Erkenntnis.* 84, no. 6 (2019): 1409-28.
[242] Wieland, Jan Willem. 'Willful Ignorance.' 106
[243] Wieland. 'Willful Ignorance'. 110

future, such as a change to their practices or circumstances. They are also reluctant to re-visit the past because they may have to negatively re-evaluate their view of their past actions.

Wieland then proposes an account of wilful ignorance,[244] which is reproduced in *Appendix B*.

Wieland next considers the illustrative example of the 'ancient slaveholder', proposed by Moody-Adams. He notes the position for which Moody-Adams argues, which attributes radical responsibility for wilful ignorance to the slaveholder regarding the moral wrongness of owning slaves, and for turning a blind eye to suffering. If the slaveholder could imagine a better way of life, then she has the possibility of acting differently, even if slavery is accepted in her society.

He then draws parallels to the modern consumer. The parallel is deliberate and intended to underscore the point. If the ancient slaveholder is radically responsible for the moral wrong of owning slaves, then the modern consumer is, by analogy, radically responsible in relation to their consumption of goods which may be made under modern slavery or slavery-like conditions.[245] The modern consumer may be ignorant of the conditions under which their products (watches, smartphones, clothing) are made. They should have considered it, they could have checked labels and done basic research before buying, but not knowing is more convenient. They do not want to have to feel guilty about past purchases, they want to keep buying cheap products in future, and they want to maintain their lifestyle and self-image. Wilful ignorance is convenient.

There are, Wieland claims, certain attitudes which are symptomatic of wilful ignorance. These include indifference, arrogance, denial, trivialisation, and rationalisation. It should also be noted that wilful ignorance may be culturally embedded or supported. For example, the related activity (buying goods manufactured under the said conditions) is not illegal, and other people are doing it. In some cases, there may be a special status, such as a designer label, attached to the goods, which makes them socially desirable.

In taking this position, Wieland seems to imply conclusions we might not be willing to accept. We would like to distinguish the criminally wilfully ignorant from the wilful ignorance that we all might fall into. Is Hannah's failure to take note of the issues concerning polyester production and 'fast fashion' really comparable to Speer's wilful ignorance regarding the concentration camps? Hannah can only effectively change her own behaviour. Speer, by contrast, was complicit in the Holocaust and only took action against it once Hitler was facing certain defeat. Is Hannah's wilful ignorance comparable to the wilful ignorance of Heredia, who despite the cues, refused to consider that her family had placed illegal drugs into the car she was driving until it was too late to safely stop the car? Perhaps these are differences in degree only, but it is difficult to justify attributing equal blame to each of these.

Wieland's explanation of wilful ignorance is that it is 'convenient' for the wilfully ignorant. This explanation seems to be understated. It does not seem to be commensurate with the strength of the moral blameworthiness which Wieland levels at wilfully ignorant consumer behaviour. No doubt, that is the point. Our convenience is indeed morally blameworthy, on this view.

---

[244] Wieland. 'Willful Ignorance.' 111

[245] Modern slavery describes situations where coercion, threats or deception are used to exploit victims. Practices include human trafficking, slavery, forced labour, debt bondage and child labour. It does not essentially include substandard working conditions or underpayment of workers alone. See: https://www.homeaffairs.gov.au/criminal-justice/Pages/modern-slavery.aspx

Wieland has assisted our understanding in bringing the concept of wilful ignorance out of the courtroom and into our everyday lives. He encourages us to examine our conduct as responsible global citizens.

His concept of wilful ignorance and the attribution of blame for it is overinclusive. He does not explain how it is that we know what *not* to investigate. However, this, the 'selectivity' issue, is important, if we want to distinguish between genuine ignorance, negligent ignorance and wilful ignorance. In response, though, it may be argued that information on the manufacture of the goods we buy is available on the internet, for easy access to everyone likely to buy the goods in question. Companies involved in importing goods are often global. Researching the origins of their goods is not particularly onerous. This tends to render claimed 'genuine' or 'negligent' ignorance less likely. On examination, it is possible that in relation to current consumer behaviour, they meld into wilful ignorance.

Normativity and the Counter-factual

The counterfactual has provided a vehicle for analysing criminal culpability and blameworthiness in wilful ignorance, and both Sarch and Wieland make use of this device. David Luban[246] goes further, arguing that the concept of wilful ignorance itself involves a counterfactual as a sub-text, especially in the criminal law.

Luban reproduces the extract from Speer's memoirs, cited above. He identifies that Speer admits that he could have known about the Nazi atrocities but chose not to know in order to keep his conscience clear.[247] Speer also admitted that wilful ignorance was as bad as knowledge and accepted responsibility. Luban argues that Speer's admission implies that if he had not engaged in wilful ignorance, if he had known the truth, he would have 'veered from his course' and acted differently.

This, Luban argues, is the counterfactual sub-text of claims of wilful ignorance. We hold the wilfully ignorant less responsible if we believe knowledge would have caused them to act differently. By contrast, if the wilfully ignorant would have acted the same, even if they had known, we hold their guilt to be compounded. Not only did they commit the criminal act, they cynically prepared a cover-up or defence by engaging in wilful ignorance.

Luban illustrates this distinction by depicting three types of wilfully ignorant actors; the fox, the unrighteous ostrich, and the half-righteous ostrich. The fox does wrong and structures his knowledge so that he can claim he does not know. The unrighteous ostrich does not want to know she is doing wrong but would do it even if she did know. The half-righteous ostrich shields herself from guilty knowledge but would not do wrong if the shield were to fail.[248]

Sarch,[249] considers this illustration. He argues for culpable *mens rea* in the first two, but for recklessness in the half-righteous ostrich case. Legally, he clarifies, we hold people responsible for what they do and what they know. We do not hold them responsible for what they might have done, nor for state of mind alone. Otherwise, we are punishing for character, which is inappropriate.

Sarch takes issue with the distinction between negligence, characterised by recklessness, and wilful ignorance. He argues that repeated recklessness, rather than being considered negligence, should attract a *mens rea* finding similar to a finding of wilful ignorance. He finds it unacceptable that a person who continues to act recklessly is less culpable than the one-time

[246] Luban, David. 'Contrived Ignorance'. *Georgetown Law Journal* 87 4 (1999) 957 ff
[247] Luban. 'Contrived Ignorance'. 965
[248] Luban. 'Contrived Ignorance'. 967 ff
[249] Sarch. *Criminally Ignorant.* 96-97

wilfully ignorant offender. However, these considerations take us deeper into legal theory and outside of the parameters of the issues considered here.

Sarch proposes that a significant part of the blameworthiness in wilful ignorance derives from the wilfully ignorant breaching a duty of investigation. This, in addition to the culpable act itself, may be almost indistinguishable from the total blameworthiness of a knowing wrongdoer. This in turn is justification for the equal culpability thesis.

Sarch points out that there may be more or less 'bad motives' for maintaining ignorance which may be to:

- Provide a defence of ignorance if caught;
- Continue to receive the benefits from peripheral involvement in a criminal enterprise; or
- Personal considerations such as maintaining family relationships.

In proposing a criterion which would take all these issues into account, Sarch proposes that the overarching criterion for blameworthiness is manifestation of lack of concern for the rights of others and for the law.[250]

Wieland[251] sees this type of 'compositional' approach to blameworthiness in wilful ignorance as being unsatisfactory. Wieland interprets this account of wilful ignorance as attributing a lowered standard of culpability for the act, and adding to that for wilful ignorance.

Wieland offers a more direct, unified approach. He proposes that the equal culpability thesis applies because the motives of the wilfully ignorant are as bad as those who act knowingly, and not because they attract additional blame for their wilful ignorance alongside lowered blame for the act. The key element for Wieland is 'greater disregard or disrespect for others'. For Wieland, we can skip the 'compositional' aspect of blameworthiness and go straight to the 'disregard' aspect.

By way of introducing counter-factuals, Wieland presents us with 'Noah' who knows what he is doing is wrong and 'Will' who is wilfully ignorant. He places them in different scenarios:

- each is offered money to transport a suitcase on a plane to another country;
- each drives a train of 'passengers' in Nazi Germany;
- each travels frequently by plane for week-end breaks regardless of the effect on the environment.

Wieland also removes any differential in access to information and stipulates that it is easier to do the right thing.

In wilful ignorance cases such as these, blameworthiness is most directly determined by motives. If in transporting the suitcase both Noah and Will have the same motive, to make fast money, this renders them equally blameworthy for their action. The punishment is for the action of transporting the suitcase. If Will would still have done the wrong thing, even if doing the right thing were easier, then this points to a bad motive. Will is as blameworthy as Noah, who acts in full knowledge.

This theory, Wieland claims, offers a more satisfactory explanation of why some Noahs and Wills are more blameworthy than others. It also accounts for degrees of blameworthiness. The equal culpability thesis holds whenever the motives of wilful ignorance and knowing wrongdoers are equally bad.

---

[250] Sarch. *Criminally Ignorant.* 119

[251] Wieland, Jan Willem. 'Willful Ignorance and Bad Motives.' *Erkenntnis.* 84, no. 6 (2019): 1409-28.

Wieland's theory only indirectly can take account of the different culpability we tend to attribute to the 'righteous ostrich'. Both Speer and Hannah are examples of 'righteous ostriches'. Speer claims he might have 'veered from his course' and this fear is what activated the wilful ignorance. Hannah, we have seen, knows she will feel guilty and may engage on an onerous change in her clothes buying practices, which would impact her life. She knows this because of her experience removing single use plastic from her life. Wieland's account might want to hold Hannah responsible for her ignorance, and morally to blame for its effects.

Hannah is in a situation which we may consider to be harshly judged by Wieland's broad explanation of morally blameworthy wilful ignorance from convenience. Hannah may have a 'good job', but her budget is limited as is her time. She is under numerous pressures, including those from her own conscience. She has experience of the difficulty in making similar changes as occurred when she eliminated single-use plastic from her life. If her wilful ignorance is due to 'convenience', it is a 'convenience' not entirely of her choosing.

Let's follow Hannah's career a bit longer. She achieves a promotion and increase in pay. She can now afford to spend more on her wardrobe. Her clothes for work are expensive, but now she always buys good quality, designer clothes. She believes she is paying for the brand name, natural fibres, and the quality that comes from workers who experience good working conditions in countries where rights are protected.

To be certain, Hannah now begins to research her designer brands. She soon discovers that the situation with high-end designer fashion is in some instances worse than for fast fashion. Some designer brands are made in Northern China, and are in turn outsourced to prisons for political prisoners in North Korea.[252]

This is a difficulty with Wieland's approach, although it does not invalidate it. Moral blame in some socio-economic situations is not straightforward. Despite our greater access to information, the global nature of manufacturing also obscures information about the conditions and effects of production. The entities involved are nations and global companies. The issues are complex and relationships intertwined. Moral blame directed at individuals misses its mark.

Hannah, in her belief system, believes that fast fashion brings employment to people in low-income countries. She is not mistaken in that. Countries such as Bangladesh derive a high percentage of their national income from fast fashion production. There is no doubt that working conditions are not optimal, compared to countries with strong industrial relations systems, but in this case, neither are they 'modern slavery'. The work provides the workers, mostly women, with an income and independence for themselves and often their families. In the 'fast fashion' process, from cloth and clothing production, through fashion stores to charity shops, and shipment off-shore to be almost dumped on low-income nations, the issues are complex and of major proportions. One person cannot make an impact. As well, if consumers suddenly simply stopped buying goods, the impact in low-income countries could be devastating. An indication of the consequences can be discerned from one consequence of the effects of the coronavirus epidemic.

The recent COVID-19 epidemic has impacted Bangladesh in particular. Cancelled orders and delayed payments from global fast-fashion companies created serious economic hardship for

[252] See Phelan, Jessica. 'Made in North Korea; $300 ski jackets and a whole lot more' *The World* February 26, 2016. Also Wong, Sue-Lin and Wen, Philip. 'North Korea factories humming with 'Made in China' clothes, traders say.' Reuters, August 13 2017.

workers and industry in some countries, including Bangladesh.[253] Factories closed down and jobs have disappeared. There are reports of women whose jobs have disappeared surviving by becoming sex workers, which contravened their personal and religious values.[254] The fashion companies involved are of a global scale and are a huge target for an individual to take on.

That is not to excuse Hannah, nor anyone else engaging in wilful ignorance. It is rather to point out that in relation to many issues, individual action, or even large scale action at one point of the process is not going to be a solution.

In fact, there is some indication that 'fast fashion' consumers recently have been acting differently in an attempt to bring about change. They are requiring the companies they deal with to verify that their sources are not involved in modern slavery or environmental destruction, and they are demanding different materials and asking for different packaging options.[255]

The point of the above is that the normative view of blame for individual wilful ignorance will not always be appropriate or effective. Issues are too diverse, the players are global and unassailable, a fix in one area will create a problem in another. That does not excuse wilful ignorance in the individual. However, it does suggest that action on a larger, possibly international scale, is going to be required if it is to be effective.

[253] SBS. Dateline. 'Covid-19. Fashion's Great Unravelling.' 2021. https://www.sbs.com.au/ondemand/video/1876811843844/dateline-s2021-ep5-covid-19-fashions-great-unravelling

[254] SBS Dateline. 'Covid-10. Fashion's Great Unravelling' 2021

[255] There have recently been advances in the fast fashion industry, in relation to transparency of supply chain, introduction of natural fibres even in fast fashion, commitments to ensure appropriate working conditions and most recently, to pay for stalled orders from supply countries affected by the pandemic. See for example: Fashion Transparency Index https://www.fashionrevolution.org/about/transparency/ Saulene Rauturier. 'Good on You' website, 22.04.2021 https://goodonyou.eco/transparency-fashion-industry/ Gazziola,Patrizia, Enrica Pavione, Roberta Pezzetti and Daniele Grechi. 'Trends in the fashion industry. The perception of sustainability and circular economy. A gender/generation quantitative approach'. *Sustainability*. 2 April, 2020. MDPI. https://www.mdpi.com/journal/sustainability

Conclusion for Wilful Ignorance

Wilful ignorance is a concept derived from law, and it was anticipated this would result in a precisely depicted concept, or hopefully, a definition. Certainly, explanations and conditions such as have been considered here, particularly those of Williams and Sarch, have focused the concept and ruled out some areas of speculation. These included doxastic issues concerning the level of belief required for wilful ignorance. For legal purposes, the level of belief was settled at a very high probability, short of actual knowledge.

This relative clarity was soon overtaken by concerns from other quarters, partly inspired by the serious consequences for wilful ignorance in the legal context. These centred around concerns at the justice of attributing knowledge in *mens rea* crimes through the 'equal culpability thesis', examined by Husak and in further depth by Sarch. In more general contexts, moral blameworthiness was an issue. The focus shifted to the attitude of the wilfully ignorant, which was equally blameworthy as knowledge in certain circumstances. It seems that normative questions are inextricably linked to wilful ignorance. However, these questions, whether in law or morality, are not central to the questions considered here.

That is not the case for the issue of motivation. The motivation for wilful ignorance is essential to understanding this phenomenon. Wieland is a key theorist in this area, noted for distilling this complex concept to its essentials. Yet he specifically did not include motivation in his characterisation of wilful ignorance. In removing motivation from consideration, Wieland produces an explanation that is too broad. Importantly, Wieland insists on holding individuals accountable for their failures of knowledge and their attitudes.

Sifting through these points of view, in particular Wieland's contribution, what emerges consistently is that the motives for wilful ignorance tend to be self-focused, involving the preservation of lifestyle, self-image or self-concept. This is true almost by definition where wilful ignorance is engaged to avoid legal penalty in breach of criminal law. Further, these concerns may extend into the past and be projected into the future.

These motives come along with considerable knowledge about the issue concerning which ignorance is claimed. This is not reflective knowledge, as certain knowledge has been avoided. Yet is apprehension of the issue which triggers the wilful ignorance, the decision not to inquire further. The wilfully ignorant have a two-way understanding. They know a good deal about what they are avoiding. This is intuitive or lower-level awareness rather than reflective knowledge. As well, they know the implications of that knowledge. Cynically, some know they will be found guilty of criminal activity. More generally, though, the wilfully ignorant know what that knowledge will mean for them, their social standing or self-image and in many cases, their conscience. In this sense, wilful ignorance involves self-knowledge and is concerned with the wilfully ignorant individual, in their relation to social issues. The irony is that it is only a person who has some awareness, some level of concern, who can engage in wilful ignorance.

In this sense, wilful ignorance offends against a personal value, although most often this value has shared social or legal origins. This may be one reason that examples of wilful ignorance are overwhelmingly negative, and attract negative normative evaluation. The wilfully ignorant are averting their eyes because a norm, often societally based, but one which they share, is being breached. It is highly unlikely that wilful ignorance can occur for 'good reasons', because if the issues about which ignorance was claimed were good, then why avoid them?

It is now possible to summarise the conclusions regarding self-deception and wilful ignorance with a view to answering the two questions posed in this thesis. These concern the relationship between wilful ignorance and self-deception, and whether they each comprise 'self-induced, purported ignorance'.

Conclusion

Wilful ignorance and self-deception share many essential characteristics and this is sufficient to offer an answer to the questions asked in this thesis concerning the relationship between them, and the proposition that both involve self-induced, purported ignorance.

Provided an Intentionalist account of self-deception as described by Donald Davidson is the account with the most explanatory power, the considerations here indicate that self-deception and wilful ignorance are related phenomena and both comprise self-induced, purported ignorance.

There are some further parallels. Self-deception is personal, involving the self-deceiver's self-concept or close personal relationships. While the focus of self-deception may be narrowed to a single issue, the implications for the self-deceiver may be wide-reaching and can be devastating. By contrast, while wilful ignorance is an attitude concerning the wilfully ignorant's self-concept or lifestyle, it also involves their conscience. Wilful ignorance concerns one's role as a law-abiding or social or global citizen. Attitudes and motives are important in wilful ignorance and normative considerations are almost inextricable from the concept.

At the outset of this project, the question seemed to be relatively straightforward. It seemed that it might be answered by a comparison of theories and paradigm examples of each concept.

On initial consideration, a further problem was apparent. Despite the familiarity of the concept of self-deception, and the long history of commentary, resulting in a vast amount of writing on the subject, there is no concise definition of self-deception. Theory construction often comprises derivation of sufficient conditions from paradigm examples. Philosophical consideration of self-deception in recent decades has flourished, resulting in lively debate, and an impasse between two major schools of thought.

This, however, provided a basis for my analysis. I, too, proposed four standard paradigm examples. Rather than enter into an argument proposing conditions for self-deception from these examples, which had already been done thoroughly by a number of philosophers, I used the paradigm examples to identify desiderata or indicators which a theory of self-deception must address. In doing so, I emphasised that 'address' means to explain or alternatively, to demonstrate that there is no need to explain because the particular indicator is unnecessary. Given the further issues in the theories of self-deception, I proposed additional desiderata.

The phenomenon of self-deception is puzzling. The self-deceiver experiences tension and almost unique challenges to their first-person authority. This prompts the question of why one would go to all that trouble. A further, related question is teleological. What is the purpose of self-deception? What does it achieve?

These desiderata or indicators were used to identify the theory of self-deception with the most explanatory power. That is, the theory that best explains the distinctive characteristics of self-deception, and only self-deception.

A consideration of the overabundance of theories on self-deception was greatly assisted by Eric Funkhouser's recent book, *Self-Deception*, but the information in this book is not uncontroversial. A literature review identified several schools of argument regarding self-deception. Consideration of all of them would be onerous. It was possible to remove from further consideration, some identified as having serious problems, others as building on more comprehensively articulated positions with little value to add, and others as being too speculative. These are theories proposing that self-deception involved mistaken second-order beliefs, divided mind theories, those which offered diachronic theories building on Intentionalism, and evolutionary biology.

I identified three major schools of thought and a representative of each for more detailed analysis: Jean-Paul Sartre for Existentialism, Donald Davidson for Intentionalism and Alfred R Mele for Motivationalism. Following a review of each of their theories, it was possible to compare them against the desiderata identified earlier and the paradigm examples, to ascertain how well each theory explained self-deception. This narrowed the field of comparison. Results of the analysis of the three theories against the indicators are presented in Table 1.

This analysis indicated that there were good reasons to prefer Intentionalism, as represented by Donald Davidson, as the theory of self-deception with the most explanatory power.

Jean-Paul Sartre's theory identified the puzzles of self-deception, and offered an explanation which depicted self-deception as lived experience, through now iconic examples. Sartre's theory arises out of and is inextricably linked to his ontological theory. Importantly, Sartre's theories concerning relations between the individual subject and 'the Other' limit some aspects of his explanation of self-deception in relation to the identified indicators, especially the challenge to first-person authority.

Alfred R Mele contributes to our understanding by moving the focus away from consideration of doxastic attitudes and towards an understanding of the processes of self-deception. However, Mele's theory is overinclusive, explaining a variety of phenomena involving motivated belief formation, while somehow not providing any essential characteristic which would distinguish self-deception. Further, it was argued that the concept of 'twisted self-deception' is flawed. It would be better to simply acknowledge that the structure of self-deception can apply to apparently negative desires.

It emerged that Donald Davidson's theory of self-deception has the most explanatory power when considered against the desiderata. Davidson's statement explaining self-deception is simple and does not rely on extensive 'sufficient conditions'. He explains the apparent paradoxes of self-deception, tying them to the selectivity issue. It is, he contends, awareness that one ought to believe a particular proposition that activates the self-deceptive belief. This is because a desire or motivation intervenes to cause the self-deception. His theory places parameters around self-deception, stating that it is an irrational attitude in an otherwise rational person. Thus psychopathology is excluded from self-deception. He also differentiates other, similar phenomena such as wishful thinking. It is by reference to Davidson's wider theories that the essential rationality of the self-deceived is explained, as well as the challenge to first-person authority.

Not all of the identified desiderata were explained adequately. Only Existentialism offers an explanation of 'why', or the teleological question 'what is the purpose of self-deception?' However, that explanation is embedded in Existentialism's wider theory of the human

condition and we may not be willing to accept that in its entirety. Explanations which propose some undefined desire or motivation do not really assist. Even Funkhouser's conclusion, that 'the self-deceived want to believe', still leaves these questions unanswered. A review of the paradigm examples suggests that self-deception concerns an issue related to the self-deceiver's self-concept, self-image or close relationships. In self-deception, one or more of these are challenged. As well, the challenge concerns something the self-deceiver cares about.

Self-deception, I then argued, is aimed at preservation and conservation. Self-deceivers are seeking neither insight nor change. The self-deceived want to maintain the status-quo. Here, this is understood to include a desired future. It was shown that, while the self-deceptive belief focuses narrowly, the impact of that issue on the life of the self-deceiver can be wide-reaching, in some cases devastating.

The self-deceived are unlikely to reflect on the relevant issue. There is no need. It was shown that our knowledge and awareness extends beyond reflective consideration. One knows, because one is living one's life, what a change, or an apprehended insight, might bring. In self-deception, the assertion of the self-deceptive belief indicates, explicitly or implicitly, awareness of the greater evidence supporting the conflicting belief. If the self-deceived can be said to be ignorant, then this ignorance is purported. Further, as self-deception is intentional, it can be confirmed that self-deception involves self-induced, purported ignorance.

The consideration of wilful ignorance began with a focus on its application to the *mens rea* aspect of a crime. In these cases, ignorance is claimed and necessarily, it is self-induced. Thus, the inescapable conclusion is that wilful ignorance, too, is self-induced, purported ignorance. A review of theories of wilful ignorance which focus on global or social issues, whether or not they may be a breach of criminal law, confirms this. In particular, Wieland's analysis insists that such ignorance is intentional and blameworthy. However, the similarities between the two concepts do not end here.

Self-deception and wilful ignorance each involve an intentional attempt to control belief and deny awareness. In each case it is difficult for others to see how they can do so. The 'selectivity issue' is central to both. It is the awareness of the denied situation or proposition that anchors the asserted belief. Further, the effort taken by the self-deceived and the wilfully ignorant to manipulate evidence involves the same tactics and results in tension.

At this point it is possible to conclude that self-deception and wilful ignorance are at least related intentional phenomena which have parallel structures.

One apparent difference is that while self-deception involves issues which tend to be more personal, wilful ignorance involves a law, or a social issue, whether global or local.

For this reason, wilful ignorance has a more intense normative aspect to it, across a range including moral blame and criminality. In this project I have mentioned the need to distinguish the normative aspects from the structural, and also noted that normative issues seem to be inextricable from wilful ignorance.

In setting the question for this thesis, the aim was to provide an answer to the question of the relationship between self-deception and wilful ignorance which was grounded in sound reasoning. Similar questions regarding the relationship of the two concepts had been asked, but the answers were conflicting. Some, such as Lynch, argued that they were different types. By

contrast, Wieland avoided distinguishing the two concepts by focusing on attitude rather than doxastic states.

Admittedly, at times in this research on self-deception, I found the variety of approaches, many of them incompatible with each other and our naïve understanding of self-deception, to be perplexing. The arguments seemed to be unwieldly. Nothing had been decided. Finally, major theories had been unable to move beyond an impasse.

The arguments here reflect my attempts to bring some order to this subject area, in order to assist and clarify my own understanding. Philosophers describe this topic as paradoxical and puzzling, so there should be no surprise that it seemed at times a considerable challenge.

Yet, out of that challenge I am satisfied that my secondary conclusion, that there are good reasons to prefer a type of Intentionalism as a theory of self-deception with the most explanatory power, is sound. I do not anticipate that this argument will convince anyone who favours another theory, but I do say that I have reached that conclusion by critical analysis, and by considering the three major theories with objectivity.

At the least, I have identified a few issues with the philosophical approach to self-deception which contributed to the theoretical maze. Argument from paradigm examples to sufficient conditions has not been unequivocally successful. Perhaps, identifying characteristics of self-deception which ought to be explained by a valid theory allows for a different approach.

Hopefully, I have clarified some aspects of the theoretical consideration of wilful ignorance. Importantly, it is not actual ignorance and this is a confusion which mars some theoretical approaches. Wieland's focus on attitudes is interesting and contentious. Whether applied in the legal, social, or global context, there is much more to be said about wilful ignorance.

Self-deception is one phenomenon which demonstrates that in life, challenges are not always met with reflection and reason. Conversely, such phenomena will not be understood by explanations relying on doxastic issues or logical analysis. We live complex lives which require agential approaches to the present, future and sometimes the past, and in order to establish our self-concept and our close relationships. When this is threatened, we attempt to regain control by manipulating information. We then believe against the evidence and this belief arises exactly because we are aware of the conflicting, or even contradictory proposition. This is self-deception. It is protective and conservationist and it is often aimed at maintaining the status quo against the challenge or threat.

Related to self-deception, wilful ignorance is more outwardly focused. It involves our roles as law abiding, societal or global citizens. In wilful ignorance, we claim we don't know about something and it is difficult for others to believe this. Wilful ignorance, too, is conservative, allowing us to preserve our self-image, lifestyle and conscience. The effects can be devastating, for others as well as ourselves. Misapplied in a legal context, it can result in conviction and serious penalty.

It is clear that understanding these issues will not be achieved by a circumscribed rational, doxastic nor a normative approach. It seems that when the stakes are high, we will abandon rationality to maintain our self-concept and the life we have created for ourselves, even if it means tension, manipulation of evidence and others questioning the integrity of our first-person authority. However, recognising the conservationist purpose of these phenomena may be a first step to addressing some negative consequences for the self-deceived, the wilfully ignorant and those affected by these attitudes.

The End

Appendix A: Table showing selected paradigm examples

Self-Deception and Wilful Ignorance

Paradigm Example	Author	SD or WI
1. Emily believes husband Daniel is faithful and there is evidence he is unfaithful a. Sam and Sally b. Rachel – changes route she drives home to avoid evidence	Margaret Penhall-Jones (MPJ) Alfred R Mele Eric Funkhouser	SD
2. Ryan believes Amy is playing 'hard to get' but she is not interested in him a. Sid and Roz b. Bill and Kate	MPJ Mele Jordi Fernández	SD
3. Liam believes he will be a great writer and blames writer's block for lack of success	MPJ	SD
4. Evelyn believes heart symptoms are just getting older a. Jack b. Burke	MPJ Fernández Lynch	SD SD WI
5. Sarah believes she can't lose weight	MPJ	SD
6. Brenda is content with the restricted role as lifelong homemaker in 1958	MPJ	SD
7. Don is bald/not bald	Donald Davidson Funkhouser	SD
8. Carlos believes his charm will help him pass his driving test when he has not studied	Davidson	SD
9. Father believes his son is not abusing illicit drugs a. Mother as above	Mele Jan Willem Wieland	SD WI
10. Gordon's parents believe their son is innocent of treason charges made against him	Mele	SD
11. Don believes the reason his article was rejected was because the editors did not understand it.	Mele	SD
12. Jealous Jeff believes his wife is unfaithful when he desires that she is faithful	Mele David Pears	SD Twisted
13. Tom believes he trusts his wife but he stalks her	Fernández	SD
14. Waiter who over-identifies with his work role, choosing contingency and denying freedom.	Jean-Paul Sartre	SD
15. Flirt on a first date who avoids making a decision about the relationship,	Sartre	SD

choosing freedom and denying contingency.		
16. Near bankrupt who refuses to look at his finances	Bishop Joseph Butler	SD
17. 'Rakes' in the 17th century	Butler	SD
18. Racist Jean-Marie is mistaken or wrong but Jean-Marie* who makes a mistake about his own belief-forming standards is self-deceived	Richard Holton	SD
19. Albert Speer avoids visiting concentration camps.	MPJ Husak Lynch	WI
20. Hannah does not inquire into 'fast fashion'	MPJ	WI
21. Origin of consumer goods	Wieland	WI
22. Slavery ancient and modern	Moody-Adams Wieland	WI
23. Torturer masks her actions with euphemisms	Moody-Adams Wieland	WI
24. Banker /Corporate executive refuses information about wrongful behaviour	Moody-Adams Wieland Sarch Lynch	WI
25. University administrator avoids considering harassment allegations	Moody-Adams Wieland	WI
26. Do not look into car boot / suitcase which may contain drugs. ▪ *USA v Jewell* and appeal ▪ *USA v Heredia* and appeal	Sarch Husak Wieland	WI
27. Lawyer does not inquire into client's guilt	Deborah Hellman	WI
28. Doctor does not inquire into whether patient on-sells prescription medication	Hellman	WI

Appendix B Table showing necessary or sufficient conditions for self-deception

	Philosopher	Proposed Conditions
SD	Donald Davidson	An agent A is self-deceived with respect to a proposition p under the following conditions. A has evidence on the basis of which he believes that p is more apt to be true than its negation; the thought that p, or the thought that he ought rationally to believe p, motivates A to act in such a way as to cause himself to believe the negation of p [256]
SD	Alfred R Mele	1. The belief that p which S acquires is false. (This is a 'lexical' point) 2. S treats data relevant, or at least seemingly relevant, to the truth value of p in a motivationally biased way. 3. This biased treatment is a nondeviant cause of S's acquiring the belief that p. 4. The body of data possessed by S at the time provides greater warrant for ~p than p. [257]
SD	Richard Holton	Adds a necessary condition re subject-matter. S is self-deceived about a subject-matter a if: 1. S's body of beliefs about a contains fundamentally de se self-knowledge errors. (This is a necessary condition). 2. S treats data relevant, or seemingly relevant to the truth value of these mistaken beliefs in a motivationally biased way. 3. This biased treatment is a non-deviant cause of S's acquiring these mistaken beliefs. 4. The body of data available to S at the time provides greater warrant for rejecting these beliefs than for accepting them.[258]
SD	Dana Nelkin	*The Desire to Believe Account* (a) *p* is false. (b) S has the desire to believe that *p* is true, and this desire causes S to treat evidence concerning the truth value of *p* in a biased way. (c) S's treating the evidence in a biased way (because of S's desire to believe that *p*) results in S's belief that *p*. (d) The body of data possessed by S at the time provides greater warrant for *not-p* than for *p*, or, if it does not, then the explanation

[256] Davidson *Problems* 208
[257] Mele *SD Unmasked* Chapter 3
[258] Holton, Richard 'Role of the Self in SD' 6

		for that fact is selective data-gathering on the part of S.[259]
WI & SD (not distinguished)	Jan Willem Wieland	S's ignorance of p is willful if (i*) p implies that A, an action of S or another agent S*, is wrong; (ii) S should have considered p; (iii) S could have considered p; (iv) but S does not consider p; (v*) because this is inconvenient for S, e.g. ➢ [forward-looking self-interest] S wants to keep on doing A in the future, since A has certain advantages for her over alternatives, or ➢ [backward-looking self-interest] S did not prevent S* or herself from doing A in the past, and this affects the image that S and/or others have of S, or ➢ [other-interest] S cares about S*, and S does not want to know that S* commits wrongful actions' [260]
WI	Kevin Lynch	Identifies 'an analysis' of wilful ignorance, but by using 'if and only if' this is the indication that he is outlining necessary and sufficient conditions. A subject, S, is willfully ignorant of p <u>if and only if</u>: 1. p is true. 2. S has a warranted suspicion that p. 3. There are some actions, v, such that were S to do them, he would find out whether p, or there are some actions, u, such that were S not to do them, he would find out whether p, and S knows this. 4. Neither doing v nor not doing u would be exorbitantly demanding for S, and also, v and u are not instances of act types that it would be exorbitantly demanding for S to consistently do/not do 5. S avoids ving, or S does u, because he does not want to know that p. 6. S should know that p, or it is arguable that he should.[261]

[259] Nelkin, Dana 'Self-deception, motivation and the desire to believe' 394.
[260] Wieland. 'Willful Ignorance' 111 Note that Wieland does not distinguish between wilful ignorance and self-deception
[261] Lynch 'Willful Ignorance and Self-Deception' 512

BIBLIOGRAPHY

Arfini, Selene. *Ignorant Cognition.A Philosophical Investigation into the Cognitive Features of Not Knowing.* Cham: Springer International Publishing AG. 2019.

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5*. Arlington, VA: American Psychiatric Association.

Amoretti, Maria Christina, and Nicla Vassalo (eds). *Knowledge, language and interpretation: on the philosophy of Donald Davidson* Vol 14 Berlin: Walter De Gruyter. 2013.

Arnold-Baker, Claire. 'What is so bad about bad faith?' *Practical Philosophy* Autumn 2002.

Bach, Kent. 'Self-deception' *The Oxford Handbook of Philosophy of Mind.* Oxford:Oxford University Press. 2009.

Baghramian, Maria and Nicholson, Anna. 'The puzzle of self-deception' *Philosophy Compass* 8 No 11 2013: 1018–1029.

Bagrahmian, Maria (ed). *Donald Davidson Life and Words.* Oxford: Routledge. 2013.

Bar-On, Dorit and Douglas C Long, 'Avowals and first person privilege' *Philosophy and Phenomenological Research* March 62 No 2. 2001. 311 – 335.

Bayne, Tim and Jordi Fernandez. 'Delusion and self-deception: mapping the terrain.' *Self-deception: Affective, and motivational influences on belief formation.* Hove: Psychology Press. 2009 : 1-21.

Bermúdez, José Luis. 'Self-deception, intentions, and contradictory beliefs.' *Analysis.* 60, no. 4. 2000. pp309-19. https://www.jstor.org/stable/3329185

Bermúdez, José Luis. 'Defending Intentionalist accounts of self-deception.' *Behavioral and Brain Sciences* 20, no. 1 (1997): 107-08.

Bermúdez, José Luis. 'Personal and sub-personal; A difference without a distinction' *Philosophical Explorations*, 3 no 1 (2000) 63-82. https://doi.org/10.1080/13869790008520981

Blay, Samuel K N. *Torts Nutshell Books Series.* 6th Edition. Law Book Company Thompsons Reuters (Professional) Australia Ltd. Rozelle, NSW. 2010.

White D E and Butler, J. *The Works of Bishop Butler.* University of Rochester Press. 2006.

Camp, Elizabeth. 'Conventions' revenge: Davidson, derangement and dormativity' *Inquiry*, 2016 Vol. 59, No. 1. 2016 :113–138, 2016 http://dx.doi.org/10.1080/0020174X.2015.1115277

Casey, Scott. 'A cup of tea, a Bex and a good lie down' *Australian Pharmacist* June 12, 2018 https://www.australianpharmacist.com.au/cup-of-tea-bex-good-lie-down/

Catalano, Joseph. 'Successfully lying to oneself : a Sartrean perspective' *Philosophy and Phenomenological Research,* 50 no. 4. 1990 673 – 693. https://www.jstor.org/stable/2108229

Charlow, Robin. 'Wilful ignorance and criminal culpability' *Texas Law Review*. 70. 1991: 1351

Child, William. 'Wittgenstein on the first person' In *The Oxford Handbook of Wittgenstein* 2011. DOI: 10.1093/oxfordhb/9780199287505.003.0018

Corkery, Jim. 'Dangerous Sports and Obvious Risks - Anyone For Cricket?' *Sports Law eJournal.* 2011 Bond University. http://epublications.bond.edu.au/slej/12

Crisp, Roger. *Sacrifice regained: morality and self-Interest in British moral philosophy from Hobbes to Bentham.* Oxford: Oxford University Press, 2019.

Davidson, Donald. *Essays on Actions and Events. Philosophical Essays. Vol 1.* Oxford: Oxford University Press. 2001.

Davidson, Donald. *Inquiries into Truth and Interpretation. Philosophical Essays. Vol 2.*Oxford: Oxford University Press. 2001

Davidson, Donald. *Subjective, Intersubjective, Objective. Philosophical Essays. Vol 3*. Oxford: Oxford University Press. 2001.

Davidson, Donald. *Problems of Rationality.Philosophical Essays Vol. 4*. Oxford: Oxford University Press. 2004.

De Caro, Mario (ed). *Interpretations and causes : new perspectives on Donald Davidson's philosophy* Vol 285. Springer Science and Business Media. 2013. DOI 10.1007/978-94-015-9227-7

Dennett, Daniel. 'Intentional Systems Theory' Tufts University. https://ase.tufts.edu/cogstud/dennett/papers/intentionalsystems.pdf

Descartes, Rene. *Discourse on Method and Meditations* New York: Dover. 2003.

Deutscher, Max. *Genre and Void: Looking back at Sartre and Beauvoir* Oxford: Routledge. 2017

Deutscher, Max. 'Infallibility, Hypocricy, Artless Talk.' *Work-in-progress.* 2021.

Deutscher, Max. 'Some recollections of Ryle and remarks on his notion of negative action' *Australasian Journal of Philosophy.* 60 No 3 (1982) 254-264. DOI:10.1080/00048408212340671

Deweese-Boyd, Ian. 'Self-deception.' Stanford Internet Encyclopedia of Philosophy. 2006.
Dohmeyer, Douglas. 'The pervasive quality of self-deception' *Aletheia* Spring, 2018:49
Dowsett Johnson, Ann. 'Alcohol as an escape from perfectionism' *The Atlantic* October 15, 2013. https://www.theatlantic.com/health/archive/2013/10/alcohol-as-escape-from-perfectionism/280482/
Edwards, Sophie. 'Nondoxasticism about self-deception' *Dialectica* 67 No 3 2013: 265-282. https://www.jstor.org/stable/42971321
Elgat, Guy. 'Slave revolt, deflated self-deception' *British Journal for the History of Philosophy* 23 No 3 2015: 524-544. http://dx.doi.org/10.1080/09608788.2015.1013915
Ellis, Albert and Robert Allan Harper. *A New Guide to Rational Living* Engelwood Cliffs, NJ:Prentice Hall 1975
Fernández, Jordi. *Transparent minds: A study of self-knowledge.* Oxford:Oxford University Press. 2013.
Fernández, Jordi 'Self-deception and self-knowledge.' *Philosophical Studies.* 162, no. 2 2013: 379-400.
Fingarette, Herbert. 'Self-deception needs no explaining.' *The Philosophical Quarterly.* 48, no. 192 1998: 289-301.
Fodor, Jerry and Lepore, Ernest 1994. 'Is Radical Interpretation Possible?', in J. E. Tomberlin (ed). *Philosophica l Perspectives , Vol. 8,* 1994 pp. 101–19.
French, R.S. Justic., Federal Court of Australia 'Mental states in civil litigation' *Australian Institute of Judicial Administration.* 21 September 2003. http://www.austlii.edu.au/au/journals/FedJSchol/2003/15.html
Freud, Sigmund. The Unconscious. *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIV (1914-1916) On the History of the Psycho-Analytic Movement Papers on Metapsychology and Other Works,* 159-215.
Freud, Sigmund. Inhibitions, Symptoms and Anxiety. *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XX (1925-1926): An Autobiographical Study, Inhibitions, Symptoms and Anxiety, The Question of Lay Analysis and Other Works*, 75-176.
Freud, Sigmund. (1915). Repression. *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIV (1914-1916): On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works*, 141-158.
Freud, Sigmund. *On Psychopathology.* London: Penguin. 1979.
Freud, Sigmund. *The Wolfman and Other Cases* London: Penguin.2002
Freud, Sigmund. *The Psychopathology of Everyday Life.* London: Penguin. 2002.
Freud, Sigmund. *The Interpretation of Dreams.* London: Penguin. 1976.
Friedan, Betty. *The Feminine Mystique.* London: Penguin. 1963.
Funkhouser, Eric. Self-Deception. Oxford: Routledge. 2019. https://doi.org.proxy.library.adelaide.edu.au/10.4324/9781315146782
Funkhouser, Eric. 'Do the Self-Deceived Get What They Want?' *Pacific Philosophical Quarterly*. 86, no. 3. 2005: 295-312.
Funkhouser, Eric and David Barrett. 'Robust, unconscious self-deception: Strategic and flexible.' *Philosophical Psychology* 29 No 5, (2016): 682-696. DOI:10.1080/09515089.2015.1134769
Funkhouser, Eric. 'Self-deception and the Limits of Folk Psychology' *Social Theory and Practice* January 2009. 35 No 1 2009 1-13.
Gardner, Sebastian. 'Psychoanalysis and the personal/sub-personal distinction', *Philosophical Explorations*, 3 no. 1, 2000: 96-119. https://doi.org/10.1080/13869790008520983
Garrett, Aaron. 'Joseph Butler's Moral Philosophy.' 2012.
Gazziola,Patrizia, Enrica Pavione, Roberta Pezzetti and Daniele Grechi. 'Trends in the fashion industry. The perception of sustainability and circular economy. A gender/generation quantitative approach' *Sustainability* 2 April, 2020. MDPI. https://www.mdpi.com/journal/sustainability
Gertler, Brie. 'Self-Knowledge', *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2020/entries/self-knowledge
Gluer, Kathrin. *Donald Davidson: A Short Introduction* Oxford University Press USA. 2011.
Halen, Elinor. *A Different Kind of Ignorance: Self-deception as flight from self-knowledge* PhD dissertation, Uppsala University, 2011.
Harman, Gilbert. *Change in View. Principles of Reasoning* The MIT Press. 1986.
Harman, Gilbert. *Notes on practical reasoning* Princeton University, New Jersey. 2007. https://www.princeton.edu/~harman/Papers/SPAWN.pdf
Harman, Gilbert. 'Notes on practical reasoning' *Cogency:Journal of Reasoning and Argumentation.* 3 no. 1 2011: 127-145
Harmon-Jones, Eddie (ed). *Cognitive dissonance, second edition: reexamining a pivotal theory in psychology*, American Psychological Association. 2019.

Harre, Rom. 'Gilbert Ryle; The Concept of Mind' in Shand, John (editor) *Central Works of Philosophy Vol 4. The 20th Century: Moore to Popper* Acumen: Chesham U.K. 2006.
Harvard Law Review Editors. *United States v Heredia* Feb 1 2008 121 Harv L Rev 1245. https://harvardlawreview.org/wp-content/uploads/pdfs/us_v_heredia.pdf
Hellman, Deborah. 'Willfully blind for good reason.' *Criminal Law, and Philosophy* 3, no. 3 2009: 301-16.
Holton, Richard. 'What is the role of the self in self-deception?' *Proceedings of the Aristotelian Society.* 2001.
Husak, Douglas N. *The Philosophy of Criminal Law: Selected Essays.* Oxford: Oxford University Press. 2010.
Husserl, Edmund. *The crisis of the European sciences and transcendental phenomenology: An introduction to phenomenological philosophy* Evanston: Northwestern University Press. 1970.
James, Caryn. 'Is this TV's most toxic stereotype?' *BBC Culture.* 16 June, 2021. https://www.bbc.com/culture/article/20210615-why-tvs-most-toxic-stereotype-endures
Jackson, Frank. 'Psychological explanation and implicit theory', *Philosophical Explorations* 3, no. 1. 2000: 83-95. https://doi.org/10.1080/13869790008520982
Graff-Radford, Jonathon, Jennifer L. Whitwell, Yonas E. Geda and Keith A. Josephs. 'Clinical and imaging features of Othello syndrome' *European Journal of Neurology* 19 no. 1 2012: 38–46. DOI:10.1111/j.1468-1331.2011.03412.x
Kahneman, Daniel. *Thinking, Fast and Slow* New York: Farrar, Straus and Giroux. 2011.
Kirsch, Julie. *Self-Deception.* Stanford Internet Encyclopedia of Philosophy https://www.iep.utm.edu/eth-self/
Levy, Neil. 'Self-deception and moral responsibility.' *Ratio.* 17, no. 3 2004: 294-311.
Lewis, Michael, and Carolyn Saarni, (eds). *Lying and deception in everyday life.* Guilford Press, 1993.
Luban, David. 'Contrived ignorance' *Georgetown Law Journal* 1999. 87:957
Lynch, Kevin. 'An agentive non-intentionalist theory of self-deception' *Canadian Journal of Philosophy* 47 no. 6 2017: 779-798. https://doi.org/10.1080/00455091.2017.1321909
Kevin Lynch. 'On the 'tension' inherent in self-deception' *Philosophical Psychology*. 25 no. 3. 2012: 433-450. DOI: 10.1080/09515089.2011.622364
Lynch, Kevin. 'Self-deception and stubborn belief.' *Erkenntnis* 78. No. 6 2013: 1337-1345. DOI 10.1007/s 10670-012-9425-0
Lynch, Kevin. 'Willful ignorance and self-deception.' *Philosophical Studies.* 173, no. 2 2016: 505-23.
Malpas, Jeff. 'Donald Davidson', *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2019/entries/davidson
Manning, Richard R. 'Rationalism in the philosophy of Donald Davidson', in Nelson, Alan (ed). *A Companion to Rationalism* John Wiley and Sons. 2012.
Mele, Alfred R. 'Approaching self-deception: how Robert Audi and I part company.' *Conscious Cognition* 19, no. 3. 2010: 745-50. https://doi.org/10.1016/j.concog.2010.06.009.
Mele, Alfred R. *Irrationality : An essay on akrasia, self-deception, and self-control*. Oxford: Oxford University Press. 1992.
Mele, Alfred R. 'Have I unmasked self-deception or am I self-deceived?' Martin, Clancy (ed) *The Philosophy of Self-deception*. Oxford University Press. 2009. 260-267
Mele, Alfred R. 'Real self-deception.' *Behaviour Brain and Science* 20, no. 1. 1997: 91-102. Discussion 3-36. https://doi.org/10.1017/s0140525x97000034.
Mele, Alfred R. 'Recent work on self-deception' *American Philosophical Quarterly* 24 no. 1 1987: 1-17. https://www.jstor.org/stable/20014170
Mele, Alfred R. 'Self-deception and delusions' *European Journal of Analytic Philosophy* 2 No. 1. 2006.
Mele, Alfred R. 'Self-deception and selectivity' *Philosophical Studies* 177 no. 9. 2020: 2697-2711 https://doi.org/10.1007/s11098-019-01334-9
Mele, Alfred R. *Self-Deception Unmasked*. *(Princeton Monographs in Philosophy)* Princeton University Press, 2001. ProQuest Ebooks https://ebookcentral.proquest.com/lib/anu/detail.action?docID=797493.
Mele, Alfred R. 'Twisted self-deception.' *Philosophical Psychology.* 12, no. 2 (1999): 117-37. DOI: 10.1080/095150899105837
Milgram, Stanley. 'Behavioural study of obedience.' *Journal of Abnormal and Social Psychology* 67 no. 4. 1963: 371-378
Moody-Adams, Michelle. 'Culture, responsibility and affected ignorance.' *Ethics* 104. No. 2. 1994: 291-309.
Nelkin, Dana K. 'Responsibility and self-deception. A framework' *Humana Mente Journal of Philosophical Studies*, 5 no. 20. 2012: 117–139.
Nelkin, Dana K. 'Self-deception, motivation and the desire to believe'. *Pacific Philosophical Quarterly* 83 no.4. 2002:384–406.
Nietzsche, Friedrich. *On the Genealogy of Morals.* London:Penguin. 2014.

Nietzsche, Friedrich. *The Viking Portable Nietzsche.* London: Penguin. 1977.
Nuremberg Trials transcripts, Speer's testimony. The Avalon Project. Yale Law School Library. https://avalon.law.yale.edu/imt/06-20-46.asp
The Nuremberg Trials: Testimony of Albert Speer. Jewish Virtual Library. American-Israeli Cooperative Enterprise (AICE) website. 1998-2021. https://www.jewishvirtuallibrary.org/nuremberg-trial-testimony-of-albert-speer#N_2_
Pascal, Blaise. *Pascal's Pensees* New York: Dutton and Co. 1958 Kindle Edition from Project Gutenberg http://www.gutenberg.org/ebooks/18269
Pears, David F, and David Pugmire. 'Motivated Irrationality.' *Proceedings of the Aristotelian Society, Supplementary Volumes.* 56 1982: 157-96.
Peden, Knox. 'Donald Davidson's 'Spinozistic extravagance'.' *Critical Horizons*, 18 no. 4, 2017:347-358. DOI: 10.1080/14409917.2017.1374916
Phelan, Jessica. 'Made in North Korea; $300 ski jackets and a whole lot more' *The World* February 26, 2016.
Preyer, Gerhard, Frank Siebelt, and Alexander Ulfig (eds). *Language Mind and Epistemology:On Donald Davidson's Philosophy* Synthese Library. Dordecht: Springer Science and Business Media. 1994.
Quattrone, George A and Tversky, Amos. *Causal versus Diagnostic Contingencies: On Self-deception and the Voter's Illusion* Stanford University. 1983.
Reese, Hope. 'Anger boiled up and Betty Friedan was there. The Feminine Mystique at 50' *The Atlantic* January 28, 2013. https://www.theatlantic.com/sexes/archive/2013/01/anger-boiled-up-and-betty-friedan-was-there-feminine-mystique-at-50/272575/
Rorty, Amélie Oksenberg. 'User-friendly self-deception.' *Journal of Philosophy.* 69, no. 268. 1994: 211-28. https://doi.org/10.1017/S0031819100046854
Rowland, Robyn. 'Australian attitudes to the traditional cultural sex role stereotype', Doctor of Philosophy thesis, Department of Psychology, University of Wollongong, 1979. https://ro.uow.edu.au/theses/1649
Ryle, Gilbert. *The Concept of Mind.* Orig, published 1949. London: Penguin. 2000.
Sarch, Alexander. *Criminally Ignorant: Why the Law Pretends We Know What We Don't.* Oxford: Oxford University Press. 2019. Published to Oxford Scholarship Online: June 2019 DOI:10.1093/oso/9780190056575.001.0001
Sarch, Alexander. 'Willful ignorance, culpability, and the criminal law.' *John's Law Review.* 88. 2014: 1023.
Sarch, Alexander. 'Ignorance lost: a reply to Yaffe on the culpability of willful ignorance.' *Criminal Law and Philosophy.* 12, no. 1. 2018: 107-24.
Sarch, Alexander. 'Willful ignorance in law and morality.' *Philosophy Compass*. 13, no. 5. 2018: e12490.
Sartre, Jean-Paul. *Being and Nothingness: An Essay in Phenomenological Ontology.* Sarah Richmond (translator). Oxford: Routledge. 2018.
SBS (Australian Special Broadcast Service) Dateline. 'Covid-19. Fashion's Great Unravelling.' 2021. https://www.sbs.com.au/ondemand/video/1876811843844/dateline-s2021-ep5-covid-19-fashions-great-unravelling
Sereny, Gitta. *Speer's Battle with Truth* Independent (online) 9 September 1995. https://www.independent.co.uk/arts-entertainment/speer-s-battle-with-truth-1600342.html
Shea, Christopher. 'A brief history of Mindfulness in the USA and its impact on our lives' *Psychology Today* October 2018. https://psychcentral.com/lib/a-brief-history-of-mindfulness-in-the-usa-and-its-impact-on-our-lives/
Sinclair, Robert. 'The philosophical significance of Triangulation:locating Davidson's non-reductive Naturalism.' *Metaphilosophy* 36 no. 5. 2005:708 – 728.
Sinclair, Robert. 'What is Radical Interpretation? Davidson, Fodor and the naturalization of Philosophy.' Inquiry. 45 no. 2 2002:161-184. https://doi.org/10.1080/002017402760093252
Sorensen, Roy A. J 'Self-deception and scattered events.' *Mind.* 94, no. 373. 1985: 64-69.
Soyka, Michael, MD. Munich Psychiatric Hospital. 'Letter to Editor.' *Journal of Neuropsychiatry and Clinical Neurosciences* (online) Vol 10 94. 1998. https://neuro.psychiatryonline.org/doi/pdf/10.1176/jnp.10.4.472
Spade, Paul Vincent. *Jean-Paul Sartre's 'Being and Nothingness'.* Class lecture notes. Indiana University. 2015. Creative Commons licence. https://scholarworks.iu.edu/dspace/handle/2022/19119
Speer, Albert. *Inside the Third Reich*. New York: Macmillan. 1970.
Strandberg, H. *Self-Knowledge and Self-deception.* 2015.
Tanney, Julia. 'Gilbert Ryle', *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2015/entries/ryle/
Trivers, Robert. *Deceit and Self-Deception: fooling yourself the better to fool others.* London:Penguin. 2011.

Van den Bussche, Eva. *The mechanisms of subliminal stimuli: a meta-analysis and new experiments.* Thesis, Doctor of Psychology, Catholic University of Leuven, 2009.
Van Leeuwen, D S Neil. 'The spandrels of self-deception: prospects for a biological theory of a mental phenomenon.' *Philosophical Psychology*. 20:3. 2007: 329-348. DOI:10.1080/09515080701197148
Varga, Somogy 'Expression, self-knowledge and authority: Wittgenstein and subjectivity' *Filozofia* 9, no.63, 2008: 830-838.
Wieland, Jan Willem. 'Willful ignorance and bad motives.' *Erkenntnis.* 84, no. 6 2019: 1409-28.
Wieland, Jan Willem. 'Willful ignorance.' *Ethical Theory, and Moral Practice.* 20, no. 1. 2017: 105-19.
Wieland, Jan Willem. 'Responsibility for strategic ignorance.' *Synthese.* 194, no. 11. 2017: 4477-97.
Williams, Glanville. *Criminal Law: The General Part*, 2nd ed. 1961. 159.
Wittgenstein, Ludwig. *Philosophical Investigations* Wiley-Blackwell. 2009.
Wong, Sue-Lin and Wen, Philip. 'North Korea factories humming with 'Made in China' clothes, traders say.' Reuters, August 13 2017.
Wright, Crispin, Barry C. Smith and Cynthia Macdonald (eds.) *Knowing Our Own Minds* Clarendon Press. 1998
Yaffe, Gideon J. 'The point of mens rea: the case of willful ignorance.' *Criminal Law, and Philosophy.* 12, no. 1 2018: 19-44.
Zimmerman, Michael J. 'Recklessness, willful ignorance and exculpation.' *Criminal Law and Philosophy.* 12 no.2. 2018:327-329.

Cases cited:

Bolton and ors v Stone [1951] 1 All ER 1078
The Queen v John Edwin Crabbe (1985) 156 Commonwealth Law Reports (CLR) 464 26 March 1985.
United States v. Jewell, 532 F.2d 697 (9th Cir. 1976)
United States v Heredia 483 F.3d 913 (9th Cir. 2007)