A Different Kind of Ignorance

Self-Deception as Flight from Self-Knowledge

UPPSALA
UNIVERSITET

A Different Kind of Ignorance

Self-Deception as Flight from Self-Knowledge

Elinor Hållén

Dissertation presented at Uppsala University to be publicly examined in Geijersalen, building 6, Engelska Parken, Humanistiskt Centrum, Uppsala, Saturday, May 14, 2011 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in Swedish.

Abstract
Hållén, E. 2011. A Different Kind of Ignorance: Self-Deception as Flight from Self-Knowledge. Department of Philosophy. 198 pp. Uppsala. ISBN 978-91-506-2206-5

In this dissertation I direct critique at a conception of self-deception prevalent in analytical philosophy, where self-deception is seen as a rational form of irrationality in which the self-deceiver strategically deceives himself on the basis of having judged that this is the best thing to do or, in order to achieve something advantageous. In Chapter One, I criticize the conception of self-deception as analogous to deceiving someone else, the so-called "standard approach to self-deception". The account under investigation is Donald Davidson's. I criticize Davidson's outline of self-deception as involving contradictory beliefs, and his portrayal of self-deception as a rational and strategic action. I trace the assumptions involved in Davidson's account back to his account of radical interpretation and argue that the problems and paradoxes that Davidson discusses are not inherent in self-deception as such but are problems arising in and out of his account. In Chapter Two, I present Sebastian Gardner's account of self-deception. Gardner is concerned with distinguishing self-deception as a form of "ordinary" irrationality that shares the structure of normal, rational thinking and action in being manipulation of beliefs from forms of irrationality treated by psychoanalysis. I object to the way in which Gardner makes this distinction and further argue that Gardner is mistaken in finding support in Freud for his claim that self-deception involves preference. In Chapter Three, I present a different understanding of self-deception. I discuss self-deception in the context of Sigmund Freud's writings on illusion, delusion, different kinds of knowledge, etc., and propose a view of self-deception where it is not seen as a lie to oneself but rather as motivated lack of self-knowledge and as a flight from anxiety. In Chapter Four, I discuss some problems inherent in the three accounts under investigation, for example, problems arising because first-person awareness is conflated with knowledge of objects.

Keywords: self-deception, self-knowledge, anxiety, rationality, intentionality, psychoanalysis, Freud, Davidson, Gardner, Lear.

Elinor Hållén, Department of Philosophy, Box 627, Uppsala University,
SE-751 26 Uppsala, Sweden.

© Elinor Hållén 2011

ISBN 978-91-506-2206-5
urn:nbn:se:uu:diva-150701 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-150701)

Printed in Sweden by Edita Västra Aros, a climate neutral company, Västerås 2011

Contents

Acknowledgements

First and foremost, I want to thank those who planted the seed of philosophical interest and questioning in me in the first place. They came to be my supervisors, *Sharon Rider* and *Sören Stenlund*. As an undergraduate student I found their teachings greatly inspiring and I read and re-read Stenlunds essays on philosophy since reading them brought light to various philosophical problems with which I found myself confronted. To me, both are exemplary in only doing philosophy that really matters to them, in writing simply and clearly with the question in focus and, in so doing, avoiding to devote time to pseudo-problems. Through this work I have learned how difficult that is, but also how rewarding. I want to especially thank Rider for the early talks during which we worked out the topic for my thesis more carefully and for her hard work during these last months in helping me to articulate my thoughts better in correcting my English. Pär Segerdahl and Mats Persson also contributed to plant that seed and helping it thrive. I am very thankful to Niklas Forsberg for taking upon himself the task of acting as opponent at my final seminar. His careful comments have been very helpful in completing my thesis. I want to thank all the participants in the seminar in theoretical philosophy for inspiring discussions and support, special thanks to Tove Österman, Niklas Forsberg, Gisela Bengtsson, Simo Säätelä and Ulrika Björk, for philosophical camaraderie and friendship. I also want to thank the faculty and its staff for helpful assistance, in particular I would like to thank Rysiek Sliwinsky for so generously and hearthwarmingly helping us all with all kinds of matters. I am thankful for having had the chance to take part in several workshops, conferances and courses arranged by the Nordic Network for Wittgenstein Research.

I am fortunate to have had the opportunity to spend six months of my graduate studies at the Department of Philosophy at University of Chicago. I thank Professor James Conant and Professor Jonathan Lear for that opportunity. I am glad that I had the chance to take part in Lear's seminar and I am grateful to them for commenting on parts of my work. At University of Chicago I have experienced the most inspiring seminars, and great intellectual rigor and seriousness. I am especially grateful that I was given the opportunity to present my work at the Contemporary Philosophy Workshop and want to thank Associate Professsor David Finkelstein and the workshop participants for valuable and helpful comments. Last but not least, thank you friends for making my stay in Chicago so nice: Laura Werner, Tuomas Nevanlinna, Stina Bäckström, Joshua Connor and Anastasia Emmanouilidou.

Thanks to an invitation from Professor Christoph Menke I could visit Potsdam Universität. I thank PD Dr. Brigitte Hilmer for her inspiring colloquium at the Department of Philosophy and for generously offering me the chance to present my work. I also want to thank the members of the colloquium. Andy, I found a home when I stepped through your door. I could not possibly have found a better place to live in Berlin than in the WG with you, Leon, Pedro and Noemi. Herzlichen Dank.

Three years earlier I was in Berlin studying German at the Goethe Institut and philosophy in the library at Helmholtz Universität in the company of Filip Mattens. It is with special warmth that I thank you for inspiration and encouragement.

Although this last year has been a solitary one and a struggle to keep it thus so that I could find the peace of mind that I needed to finish this work and to enjoy the sweetness of fruitful work, it is a time at which it has been most evident to me to what great extents my spirit, confidence and well-being emanate from you, my loved ones. Rikard Ekholm, I am very happy that you have shared the pains and pleasures of completing this dissertation with me. I thank you for your loving support and for commenting upon my manuscript, but most of all I am happy about all the nice moments away from work that I share with you. I thank my family with special warmth, for their support and trust in me. I thank especially my parents, Kurt-Lennart Hållén and Britt-Marie Karlsson, my sisters Carina, Connie and Lotta, and my dear cousin Frida Hållén. I send a thought to my late grandparents Inez & Sture Hållén, as I have often done during these years.

I dedicate this dissertation to my dear friends. You have kept me on an even keel through these years and made life more pleasurable in general. I cannot mention you all by name here but I must mention three. Jacob Engstand, for enthusiastic encouragement and support, and for our existential talks. Frida Hållén, dear cousin and friend, for your love and encouragement. Karin Hedvall, for inspiring discussions and creative and therapeutic advise in times of distress, for delightful breaks from work (for that I thank Anna Karlsson too), for the warmth of your friendship.

This dissertation was made possible because of the generous grant Göransson-Sandviken Stipendium, awarded through Gästrike-Hälsninge Nation. It was my funding for the first two years. I have also received a Göransson-Sandviken travel grant for German studies at the Goethe Institut. Generuos funding from STINT (The Swedish Foundation for International Cooperation in Reasearch and Higher Education) made my visit to Chicago possible, and thanks to funding from Helge Ax:son Johnsons stiftelse I could visit Potsdam Universität. Jubelfeststipendium historiska-filosofiska fakulteten, awarded through Uppsala universitet, has helped me complete this dissertation in combination with a grant from Birgit och Gad Rausings Stiftelse.

Introduction
Views and Conceptions of Self-Deception

Introduction

The phenomenon of self-deception can appear perplexing to anyone, though perhaps most especially to the self-deceiver herself when she realizes that, in spite of many indications, she has been oblivious to something concerning herself for some time. For though the indications were there, she had not seen them *as* indications. She did not see any particular meaning in them. Self-deception can take many forms; here I simply want to present a couple of cases that seem to fit the description 'self-deception' as food for thought. Consider this example:

A woman has been abusing alcohol for many years without admitting to her abuse. Her children have tried to help her many times, once by sending her away for treatment from which the woman was soon banned, since she showed no insight into her abuse and, therefore, could not be treated. After attempting suicide, she is brought to the hospital again. After a horrifying delirium, she sleeps for 40 hours. She describes, with great sensitivity to details, how she wakes up, looks at the bright sky, pulls her bed up to look at the sun as it lights up the town below, and she asks herself: how did it come to this? She describes how her thoughts are suddenly clear. She recalls a series of failed relationships, how she has lived hand to mouth, her almost non-existent contact with her children and grandchildren. She now clearly sees all of this as a direct consequence of her drinking. She has never seen her drinking habits as abuse before, although she has been a slave to alcohol; it is at this moment that she admits that she is an alcoholic.

In this case, the person refuses to accept an obvious fact about herself. When she finally sees herself as an alcoholic, it opens up a world of lost chances, broken relationships with loved one's etc., but it also opens the possibility of taking control of her life, of redemption. Another case, which I would also call a case of self-deception, is when what the self-deceiver avoids acknowledging is not a fact about himself (at least not directly), but something that is unpleasant for him and a cause for concern.

A man has recently bought a flat. He paid more for it than he had planned, but so be it. It is in mint condition! He will be on a tight budget for a couple of years, but no repairs or renovation will be needed, and that's a relief. Strangely, he has had a runny nose ever since he moved in, and he sneezes often, especially when he is in the bathroom. He has noticed a patch next to the bathtub where the paint is coming off. He hasn't pulled out the bathtub to take a closer look. But yesterday, when the neighbor's dog was visiting, it kept sniffing at that patch. Now he is worried, and wonders what this might mean.

Self-deception can also seem baffling when one sees it manifested in someone else. Consider one last case, where a man notices that his friend seems to be oblivious to his own feelings and appears unwilling to reflect upon them.

John is eating his weekly brunch with his best friend Allan. As usual, Allan is talking about his colleague Sally. All the details about the lunch he had with her yesterday, about how she got a little angry with him the other day, but that it was just because of the stress she was under. A couple of months ago, John had asked Allan if he had a crush on Sally (it was obvious to John already at that point), but Allan denied it. On that occasion he was bothered, even a little upset, by John's question. "Well, he is not lying", John thought to himself now; "if he wanted to keep me in the dark he would not talk about her all the time. He doesn't seem to realize that he is obsessed with her."

Philosophers also find self-deception perplexing. One finds references to self-deception throughout the history of philosophy, displaying a wide variety of conceptions, explanations and positions. A quote ascribed to the ancient orator, Demosthenes reads: "Nothing is easier than self-deceit. For what each man wishes, that he also believes to be true."[1] In "Cratylus", Plato writes: "For the worst of all deceptions is self-deception. How can it help being terrible when the deceiver is always present and never stirs from the spot?"[2] And Friedrich Nietzsche writes in *Human, All Too Human*:

> I knowingly-willfully closed my eyes before Schopenhauer's blind will to morality at a time when I was already sufficiently clear-sighted about morality; likewise I had deceived myself over Richard Wagner's incurable romanticism, as though it were a beginning and not an end; likewise over the Greeks; likewise over the Germans and their future – and perhaps a whole long list could be made of such likewises? – Supposing, however that all this were true and that I were reproached with it with good reason, what do *you* know, what *could* you know about how much cunning in self-preservation, how much reason and higher safeguarding – is contained in such self-deception – or of how much falsity I shall *require* if I am to permit myself the luxury of *my* truthfulness?[3]

As a final example of philosophers' thoughts on self-deception, I cite Bertrand Russell: "No satisfaction based upon self-deception is solid, and however unpleasant the truth may be, it is better to face it once for all, to get used to it,

---

[1] Demostenes, *Olynthiacs, Phillippics, Minor Public Speeches*, trans. J. H. Vince, Loeb Classical Library (Harvard: Harvard University Press, 1930) Third Olynthiac, paragraph 19, p. 53.

[2] Plato, *Cratylus*, ed. G. P. Goold, transl. H.N. Fowler, Loeb Classical Library (London: Harvard University Press, 1977), 428d.

[3] Friedrich Nietzsche, *Human, All Too Human: A Book for Free Spirits*, transl. R. J. Hollingdale (Cambridge: Cambridge University Press, 2002), Preface, Section 1, p. 6. Or, as Dr. Relling says in Henrik Ibsen's play *The Wildduck*, "Deprive the average man of his life-illusion, and you rob him of his happiness at the same stroke." (Henrik Ibsen, *The Wildduck*, transl. Frances E. Archer, 2nd edition, (Massachusetts: Digireads.com, 2008), p. 97.

and to proceed to build your life in accordance with it."[4] If one tries to briefly summarize what these quotes say about self-deception, one arrives at something like this: nothing is easier than self-deception, and nothing is worse. On the one hand, self-deception may serve the purpose of self-preservation and even, ultimately, truth-seeking. On the other, it is an unstable ground to build on, and one is better off without it, even if facing the truth may be painful. The broad spectrum of viewpoints and attitudes reflects the many approaches one can take in trying to understand self-deception.

The discussion of self-deception within contemporary philosophy is lively and extensive, particularly within analytical philosophy. The increasing interest in self-deception seems partly due to the fact that self-deception is conceived of as a melting pot for many philosophical problems. As Brian P. McLaughlin and Amélie Oksenberg Rorty write in the introduction to the volume *Perspectives on Self-Deception*, "[w]e have used self-deception as a microcosmic case study that bears on a range of issues dividing contemporary philosophical psychology. The discussion of the issues surrounding self-deception gives us a red-dye tracer for tracking what is at stake in a variety of debates central to the philosophy of mind."[5] The phenomenon of self-deception raises a number of questions. Ontic questions, such as: What are the conditions for self-deception? What kind of self is capable of self-deception?[6] But it also raises psychological questions, such as: Is self-deception always motivated? How is self-deception related to repression and denial?[7] Further, it raises moral issues, such as: Can self-deceivers be held responsible for their self-deception? How does self-deception affect someone's capacity to act as a moral subject? Finally, it raises epistemic issues, such as: Is self-deception a failure of self-knowledge, or is it an intentional action of lying to oneself about something which one knows?

These aspects will all be present in the study of self-deception carried out in this book, but it is the last epistemic aspect that will open the investigation and be the most informative. Let us take a look at questions that arise in, and through, this perspective. If self-deception is a failure of self-knowledge, it sets in already at the level of forming knowledge about oneself. On the one hand, the idea of failing to know our own feelings, pains, beliefs, etc., can seem strange. Do we not have immediate knowledge of our own mental states, and are we not authorities when it comes to knowledge of our own feelings, desires, beliefs etc.? How is it then possible that we should be mistaken about our own mental states? (And mistaken in relation to what?) On the other hand, the

---

[4] Bertrand Russell, *The Conquest of Happiness* (London: The Unwin Brothers Ltd., 1930), p. 125.
[5] Brian P. McLaughlin and Amélie Oksenberg Rorty (eds.), *Perspectives on Self-Deception* (Berkeley and Los Angeles: University of California Press, 1988), "Introduction", p. 1.
Many influential papers in recent philosophical research on self-deception in the Anglo-American analytical literature have been collected in the volumes *Perspectives on Self-Deception*, edited by Brian P. McLaughlin and Amélie Oksenberg Rorty (1988) and Jean-Pierre Dupuy's *Self-Deception and Paradoxes of Rationality* (1998).
[6] McLaughlin, pp. 1.
[7] Ibid. p. 4.

history of philosophy teaches us self-knowledge as an *ideal*, that is, something we should strive for but which we do not have. Further, we ought to be *aware* of not having complete knowledge of ourselves. "Know thyself", the maxim ascribed to the Delphic Oracle, was an ideal upheld by the Greeks and it has survived, in some form, until today. It is what Socrates strives for, but before which he is humble; his wisdom consists in knowing that he does not know.[8] This suggests to us that there are things about ourselves that we do not know, and that there are things about ourselves that we think we know but which we do not know. Even though our mental states are *ours*, and are in that sense immediately known to us, in contrast to states of affairs outside of us requiring observation, inference and judgment – we may, nevertheless, be mistaken about our own feelings, capacities and values.

According to a different interpretation, self-deception ought not to be seen as a failure of self-knowledge but rather as a lie to oneself. On this account, self-deception is an intentional action of misleading oneself about something which one knows. The portrayal of self-deception as analogous to deceiving someone else is often referred to as the "standard approach to self-deception".[9] Just as the liar knows or truly believes that what he says to someone is false but says it with the intention of making the other person believe it, the self-deceiver knows or truly believes that something is false, but her aim is to make herself believe that it is true. Thus, in this view, the self-deceiver believes of something (a proposition) both that it is true and that it is false. Such a view gives rise to a paradox (or more precisely, a company of paradoxes): how is it possible for a rational person to believe both that the proposition is true and that it is false? The portrayal of self-deception as a paradoxical phenomenon is common, and it has been extensively discussed, for example, in the anthology *Self-Deception and Paradoxes of Rationality*. As I said above, in the standard approach, self-deception is conceived of as an act of intentionally lying to oneself. This division of accounts of self-deception into non-intentionalist accounts and intentionalist accounts serves as a watershed between philosophical accounts of self-deception. There are, however, other intentionalist accounts of self-deception that are not based on the model of deceiving others, but are still described as a strategy of misleading oneself about something in order to obtain a goal. In this dissertation I will first consider an example of the standard

---

[8] One of the most famous formulations is found in the *Phaedrus*: "But I have no leisure for them at all; and the reason, my friend, is this: I am not yet able, as the Delphic inscription has it, to know myself; so it seems to me ridiculous, when I do not yet know that, to investigate irrelevant things. And so I dismiss these matters and accepting the customary belief about them, as I was saying just now, I investigate not these things, but myself, to know whether I am a monster more complicated and more furious than Typhon or a gentler and simpler creature, to whom a divine and quiet lot is given by nature." (Plato, *Phaedrus*, in *Plato in Twelve Volumes*, Vol. 9 transl. Harold N. Fowler, London: William Heinemann Ltd., 1925), sections 229e -230a.

[9] For example, Ariela Lazar's paper, "Division and Deception: Davidson on Being Self-Deceived" in *Self-Deception and Paradoxes of Rationality*, ed. Jean-Pierre Dupuy (Stanford: CSLI Publications, 1998), p. 21.

approach and then another intentionalist account of self-deception, Donald Davidson's and Sebastian Gardner's accounts of self-deception, respectively. I would describe both accounts as intentionalist and rationalist. I will move on to present a conception of self-deception in which these problems do not arise, a conception which can be found, explicitly and implicitly, in Sigmund Freud's works.

The first chapter discusses Donald Davidson's account of self-deception, as presented in the four papers on irrationality in the collection *Problems of Rationality*. Davidson's account is considered a prime case of the standard approach to self-deception. According to Davidson, self-deception begins with the holding of a belief that is contrary to what one desires to believe, but which one knows that one has the best evidence for believing. Self-deception consists in intentionally misleading oneself by making oneself believe the opposite while still holding the first belief. Davidson's view gives rise to a number of paradoxes: how can a person, a rational subject, hold contradictory beliefs? As Jean-Pierre Dupey writes in his introduction to *Self-Deception and Paradoxes of Rationality*, the paradoxes to which self-deception is thought to give rise on this account "jeopardize the foundations of the rationalist paradigm."[10] This is Davidson's inducement for trying to solve them. Here I challenge Davidson's intentionalist and rationalist outline of self-deception, and try to tease out the assumptions and purposes at work in his account. I will question the knowledge which Davidson ascribes to the self-deceiver (that he initially holds a belief based on the best total evidence) as well as the rational ingenuity which Davidson takes self-deception to display. I will argue that what goes on in self-deception is not as transparent to the subject as Davidson's characterization suggests, and that the self-deceiver is not "in charge" to the extent that his interpretation suggests. Moreover, I will argue that although there are real problems involved in self-deception, the "problems of self-deception" that Davidson discusses are not; they are problems arising from his definition of self-deception, that is, strictly theoretical problems.

In Chapter Two, I discuss Sebastian Gardner's account of self-deception as presented in his book *Irrationality and the Philosophy of Psychoanalysis*. Gardner's account will serve as a bridge between Davidson's account and a Freudian conception of self-deception. While Gardner's account bears great similarity to Davidson's, it also relates the discussion of self-deception to psychoanalytic theory. In Gardner's and Davidson's interpretations alike, self-deception is described as a form of irrationality. Gardner, however, draws a sharp distinction between "ordinary irrationality", to which he argues that self-deception belongs, on the one hand, and "forms of irrationality accounted for by psychoanalysis", e.g. obsessional neurosis, on the other. I will question the way in which Gardner distinguishes between "ordinary irrationality" and irrationality requiring psychoanalytic terminology for its explication. Gardner

[10] Dupey, "Introduction", x.

argues, in line with Davidson, that self-deception is an intentional, goal-directed action in which beliefs are manipulated and, further, that the self-deceiver knows what she is up to in deceiving herself. In his view, self-deception is ultimately rational. Gardner even calls it hyper-rational. It can be fitted into the view of persons as rational unities, while the other form of irrationality cannot. Thus, in Gardner's conception, irrationality ought to be divided into two forms on the basis of qualitative differences. He claims that Freud's writings lend support to this distinction. I will argue that this claim is based on a misreading of Freud. I will also argue that Gardner, like Davidson, ascribes too much knowledge (awareness) and too much rationality (and strategy) to the self-deceiver. Though I am critical of Gardner's interpretation of self-deception, I will also make positive use of his writings by employing his careful analyses of the forms of irrationality accounted for by psychoanalysis to argue, against Gardner, that they help us to understand the phenomenon of self-deception.

In Chapter Three, I present a conception of self-deception that differs greatly from that of Davidson and Gardner, whose accounts, I will argue, share many characteristics. I open the chapter with some etymological remarks regarding the term 'self-deception', in particular, certain connotations that are missing in Davidson's and Gardner's accounts. Self-deception is here explored in the context of illusion and delusion rather than in analogy with deceiving someone else. I then turn to Freud. Whereas Davidson's and Gardner's accounts of self-deception treat it as an intentional action of misleading oneself about something which one knows, Freud's writings on different kinds of illusion, delusion, knowledge and ignorance, and – especially – on defensive reactions, show that self-deception is more basic. Avoidance of displeasure and anxiety influences even apprehension and affects the very formation of knowledge. In this conception, self-deception does not have the structure of a lie. The lie presupposes knowledge of that which one misleads someone (oneself) about, while self-deception is, to a large extent, avoidance of knowledge, or of the forming of beliefs. My intention in arguing for this conception of self-deception is not to claim that the word 'self-deception' *never* refers to something that is similar to a lie, but I do reject the view that this is the "core" of self-deception, the fundamental sense from which there can be only deviations. In Chapter Three, I continue discussing the theoretical problems arising in and through the conception of self-deception as a lie to oneself. I also raise the more important question of whether this description is true of cases we are inclined to call cases of self-deception. I study closely a few such cases and in the light of Freud's texts, I argue that self-deception is not best understood as a lie to oneself. If a general description of self-deception should be given at all, I argue it could be something like "a flight from something that provokes anxiety or displeasure".

I have argued that self-deception is not an intentional action directed at achieving something or promoting a wish. How are we then to understand the sections in which Freud speaks of unconscious intention? In Chapter Three, I

discuss Freud's use of unconscious intention with help of, among others, Elisabeth Anscombe and Jonathan Lear.

In Chapter Four, I reflect upon two problems that I find in varying degrees and forms, in all three accounts. First, I call attention to something that I have remarked upon earlier in this study: that a number of the problems that arise in these accounts arise because of certain unexamined assumptions and aims. I will discuss how the interpretive and theoretical perspective itself gives rise to problems. Second, I turn to Richard Moran's discussion of problems arising out of the conflation of self-knowledge, on the one hand, and knowledge about things and other people's mental states, on the other. Moran asserts that accounts of self-knowledge often fail to do justice to the peculiarity of self-knowledge, either by describing it as a third-person phenomenon or by transposing a third-person situation to some kind of mental exterior. I argue that Davidson is guilty of the latter in accounting for self-deception on the model of deceiving someone else. Gardner's account of self-deception and even parts of Freud's work are problematic in the former sense. Freud sometimes accounts for introspective awareness, for instance, as an "inner eye", as if self-knowledge were akin to perceptual knowledge. I give examples of how such conflation gives rise to certain problems in the interpretation of self-deception and related phenomena. When self-knowledge is understood on the model of observation of something inner, e.g. a feeling, the feeling is understood as an independent object which is not affected by if one is conscious of it or not. This conception cannot account for the moral importance of becoming aware of one's feelings, preconceptions, the meaning of our actions, etc.; of how we in reflection can alter our attitudes.

Donald Davidson

The Paradoxical Nature of Self-Deception

Self-Deception in The Context of Radical Interpretation

The volume *Problems of Rationality* contains several of Donald Davidson's essays that revolve around the topic of rationality. The last four essays deal with the topic of irrationality, and include discussions of self-deception. These essays are the material around which this chapter will evolve. In her introduction to *Problems of Rationality*, Marcia Cavell briefly sketches the background to Davidson's discussion of irrationality. She writes: "It has been argued that large-scale rationality on the part of the interpretant is an essential background of his interpretability, and therefore, in light of Davidson's argument […], of his having a mind."[11] She continues by formulating the problem that occupies Davidson in the four essays on irrationality: "Yet […] cases of irrationality, as judged by the interpretant's own standards, do exist. How can we explain them without falling into inconsistency ourselves, as we would, for example, in attributing to an agent both a belief and a disbelief in the same proposition?"[12] As we see, the problem is how to account for irrationality when it is assumed that people (language-users) are, on the whole, rational.

In the quotes above, Cavell is referring to Davidson's theory of meaning, that is, to his account of radical interpretation. Although I will not treat Davidson's account of radical interpretation here, I will present a summary of some basic assumptions which Davidson makes in order to get the project of radical interpretation off the ground. For this purpose I will refer to Simon Evnine's summary from the book *Donald Davidson*. Later in this chapter, I will argue that these assumptions influence Davidson's account of self-deception in a way that causes problems. I will argue that what Davidson takes to be fundamental problems inherent in the notion of self-deception are rather problems arising in and out of his account because of how he explicates self-deception. I will argue that the most central problems (or, apparent problems) originate in presumptions that Davidson makes in his theory of meaning, in outlining the Principle of Charity.

What is at issue in regards to radical interpretation is the possibility of interpreting the linguistic behavior of a speaker from scratch, without reliance on either knowledge of his beliefs or knowledge of the meaning of his utterances. Thus what Davidson needs to provide is both a theory of belief and a theory of meaning at the same time: radical interpretation must address the problem that one cannot assign meaning to a speaker's utterances without knowing what the speaker believes, while one cannot identify beliefs without

[11] Marcia Cavell's *Introduction* to Donald Davidson, *Problems of Rationality* (Oxford: Clarendon Press, 2004), xviii.
[12] Ibid. xvii-xix.

knowing what the speaker's utterances mean.[13] To get his project of radical interpretation off the ground, Davidson makes assumptions about how beliefs and utterances are related. Davidson says that we must assume that people believe the obvious. As the commentator Simon Evnine notes, the question then arises: obvious for whom? It could well be that although it is obvious for the interpreter that it is raining, it is not obvious for the interpretant; thus she may not hold the belief "It is raining", although the interpreter, by his own lights, holds that if she is rational then she should hold such a belief. Evnine continues: "The assumption that Davidson is requiring us to make, therefore, is that we take others (the interpretees) to find obvious what we (the interpreters) find obvious. […] we must assume that the people whom we are interpreting will believe what we think it is right to believe. This means that in radical interpretation we must assume that the objects of interpretation, by and large, believe what we think is true."[14]

Davidson calls this fundamental assumption of his the Principle of Charity, or, on some occasions, the policy of rational accomodation[15]. It serves to optimize agreement between interpreter and interpretant and "counsels us quite generally to prefer theories of interpretation that minimize disagreement."[16] Davidson describes the practice of radical interpretation briefly in "Truth and Meaning":

> It must be possible, of course, for the speaker of one language to construct a theory of meaning for the speaker of another […] the aim of theory will be an infinite correlation of sentences alike in truth. […] The linguist then will attempt to construct a characterization of truth-for-the-alien which yields, so far as possible, a mapping of sentences held true (or false) by the alien on to sentences held true (or false) by the linguist. […] just as we must maximize agreement, or risk not making sense of what the alien is talking about, so we must maximize the self-consistency we attribute to him, on pain of not understanding *him*.[17]

We thus see that the interpreter both needs to make sense of the speaker's words and of the pattern of his beliefs.[18]

---

[13] In "Truth and Meaning", Davidson says: "We do not know what someone means unless we know what he believes; we do not know what someone believes unless we know what he means. In radical interpretation we are able to break into this circle, if only incompletely, because we can sometimes tell that a person accedes to a sentence we do not understand." (Donald Davidson, "Truth and Meaning" in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 1984, p. 27.)

[14] Simon Evnine, *Donald Davidson* (Stanford: Stanford University Press, 1991), p. 103.

[15] Donald Davidson, "Expressing Evaluations", in *Problems of Rationality*, p. 36.

[16] Davidson, *Inquiries into Truth and Interpretation*, Introduction, xvii.

[17] Ibid. "Truth and Meaning", p. 27.

[18] Compare: "On Saying That" in *Inquiries into Truth and Interpretation*, p. 101.

It will be noticed that the Principle of Charity is actually composed of two related notions: coherence and correspondence. In "Three Varieties of Knowledge", Davidson writes:

> The Principle of Coherence prompts the interpreter to discover a degree of logical consistency in the thought of the speaker; the Principle of Correspondence prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances. Both principles can be (and have been) called principles of charity: one principle endows the speaker with a modicum of logic, the other endows him with a degree of what the interpreter takes to be true about the world.[19]

The first notion is a *holistic assumption of rationality in belief* (coherence). It says that interpretation must be guided by normative principles. These normative principles express necessary and *a priori* conditions for a belief or an intentional action, such as that a belief generally does not give rise to open contradictions, or, if someone acts intentionally, the action should be based on judgments of what is, all things considered, best to do.[20] The other is an assumption of causal relatedness between beliefs and the objects of belief ('correspondence'). Evnine writes on the Principle of Charity: "Rationality, in this context, is limited to the standards of logical reasoning (deductive and inductive inference), and to the ways in which mental states should relate to the world, through perception and belief on the one hand, and through action and desire on the other."[21] He continues: "It is part of the very concept of belief that, as one of the normative principles might say, beliefs should be determined (not that they always are determined) just by how the world is. Similarly, it is part of the concept of intentional action that actions should be based on judgements of what, all things considered, is best to do."[22] Thus, in radical interpretation it must be possible to make the assumptions that beliefs and other mental states correspond to the world, and that intentional action is motivated by judgments of what is reasonable to do.

We saw above that, according to the holistic notion that beliefs are rationally related and make up coherent wholes, the interpreter must maximize *self-consistency* in ascribing beliefs to the speaker and make sense of the pattern of his beliefs. Although the main source of evidence for telling which mental state someone is in is his behavior, this is not enough to attribute a mental state to that person. The attribution of a belief must be made against the background of attributions of other mental states.[23] In "Mental Events", Davidson writes:

---

[19] Donald Davidson, "Three Varieties of Knowledge" in *Subjective, Intersubjective, Objective*, (Oxford: Clarendon Press, 2001), p. 211.
[20] Evnine, p. 12.
[21] Ibid.
[22] Ibid.
[23] Ibid. p. 14.

"There is no assigning beliefs to a person one by one on the basis of his verbal behavior, his choices, or other local signs no matter how plain and evident, for we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations, and the rest."[24] Propositional attitudes, such as beliefs, preferences, intentions etc., can only be attributed to someone against the background of other propositional attitudes that he holds: there must be a context that rationalizes the belief attributed. One problem with the holistic conception is that it makes it difficult to account for inconsistency between a person's attitudes. In "Incoherence and Irrationality", Davidson says:

> The difficulty in describing inner inconsistency is created by the character of the so-called propositional attitudes: belief, desire, intention, and many of the emotions. Put briefly, the problem is this: one way in which propositions are identified and distinguished from another is by their logical properties, their place in a logical network. But then it would not seem possible to have a propositional attitude that is not rationally related to other propositional attitudes. For the propositional attitude itself, like the propositions to which it is directed, is in part identified by its logical relations to other propositional attitudes.[25]

As we will see, Davidson understands the irrationality that goes into self-deception as inconsistency and incoherence between beliefs; what he expresses above is a worry about how to account for irrationality of this sort. Against the backdrop of this sketch of the relevant features of Davidson's assumptions about the rationality of the language user in his theory of meaning, I will now turn to Davidson's discussion of irrationality.

The Paradox of Irrationality

Davidson conceives of irrationality as paradoxical. He opens his essay "Paradoxes of Irrationality" by saying:

> The idea of an irrational action, belief, intention, inference, or emotion is paradoxical. For the irrational is not merely the non-rational, which lies outside the ambit of the rational; irrationality is a failure within the house of reason. (When Hobbes says only man has 'the privilege of absurdity' he suggests that only a rational creature can be irrational.) Irrationality is a mental process or state – a rational process or state – gone wrong. How can this be?[26]

[24] Donald Davidson, "Mental Events" in *Actions and Events* (New York: Oxford University Press, 1980), p. 221.
[25] Davidson, "Incoherence and Irrationality", in *Problems of Rationality*, p. 189.
[26] Ibid. "Paradoxes of Irrationality", p. 169.

Irrational actions, beliefs etc. appear paradoxical to Davidson because they are the actions, beliefs etc. of a rational being. His interest in irrationality is to understand how a rational being can hold irrational beliefs. Davidson writes that irrationality, like rationality, is a normative concept, by which he means that the basis for regarding an action as irrational is not anyone's judgments; rather, when an agent acts irrationally, he deviates from the normal condition of consistency and coherence in his own behavior and thought. To be irrational is to act, think or feel counter to one's own conception of what is reasonable.[27] He writes: "The sort of irrationality that makes conceptual trouble is not the failure of someone else to believe or feel or do what we deem reasonable, but rather the failure, within a single person, of coherence or consistency in the pattern of beliefs, attitudes, emotions, intentions, and actions."[28] According to Davidson, irrationality implies not only that a rational creature is acting or reasoning in an irrational way, but also that the irrational element is part of an otherwise rational process. Davidson holds that all actions, including irrational ones, are rational *at the core*. He says: "all intentional actions, whether or not they are in some further sense irrational, have a rational element at the core: it is this that makes for one of the paradoxes of irrationality."[29] Davidson says that irrationality is a rational process *gone wrong*. In order to better understand irrationality as a deviation from the rational, let us look at one of Davidson's examples: Sigmund Freud's case study of a man, Mr. S, with a neurotic condition, who acts in an irrational way.

> A man walking in a park stumbles on a branch in the path. Thinking the branch may endanger others, he picks it up and throws it in a hedge beside the path. On his way home it occurs to him that the branch may be projecting from the hedge and so still be a threat to the unwary walkers. He gets off the tram he is on, returns to the park, and restores the branch to its original position.[30]

Davidson analyses this example as follows:

> Here everything the agent does (except stumble on the branch) is done for a reason, a reason in the light of which the corresponding action was reasonable. Given that the man believed the stick was a danger if left on the path, and had a desire to eliminate the danger, it was reasonable to remove the stick. Given that, on second thought, he believed the stick was a danger in the hedge, it was reasonable to extract the stick from the hedge and replace it on the path. Given that the man wanted to take the stick from the hedge, it was reasonable to dismount from the tram and return to the park. In each case the reasons for the action tell us what the agent saw in the action, they give the intention with which he acted, and thereby give an explanation of the action. Such an

[27] Ibid. "Incoherence and Irrationality", p 189.
[28] Ibid. "Paradoxes of Irrationality", p. 170.
[29] Ibid. pp. 173.
[30] Ibid. p. 172.

> explanation, as I have said, must exist if something a person does is to count as an action at all.[31]

Davidson argues that for something a person does to count as an action, there must be a reason for it; and this reason expresses or explains the agent's intention in performing the action. In analyzing what it is to be a reason for an action, Davidson brings out two components: a desire and a belief. For an agent to have a reason to do something, he must want to do it. He must also have the belief that by acting as he does he can get what he wants: "At the minimum, the explanation calls on two factors: a value, goal, want or attitude of the agent, and a belief that by acting in the way to be explained he can promote the relevant value or goal."[32]

Davidson draws two conclusions from his analysis of reason explanations. The first is that all intentional actions, even the ones that, in a further sense, are irrational, have a rational element at the core, since the beliefs and desires of which intentional actions consist have propositional content which bear appropriate logical relations to one another: "Beliefs and desires have a content, and these contents must be such as to imply that there is something valuable or desirable about the action."[33] The second conclusion is: "the reasons an agent has for acting must, if they are to explain the action, be the reasons on which he acted; the reasons must have played a causal role in the occurrence of the action."[34] The causal and the logical merge in the explanation of intentional action; since beliefs and desires are causes of the beliefs, desires and actions for which they are reasons, Davidson states: "there is no inherent conflict between reason explanations and causal explanations. Since beliefs and desires are causes of the actions for which they are reasons, reason explanations include an essential causal element."[35]

In Davidson's account, even an action such as that of Mr. S', which is irrational taken in its totality, consists of elements with logical structure, that is, of reasonable relations between pairs of beliefs-and-desires and actions. Let the example of Mr. S characterize the logical (and causal) relation. According to Davidson's analysis, Mr. S has a reason to go back and put the stick on the path again. This reason consists in his desire to eliminate danger and his belief that the stick is a danger in the hedge.[36] For this to be the reason for the action, it must have caused it. Davidson thus argues that an irrational action should be analyzed as consisting of a set of structures, which, in themselves, are

---

[31] Ibid. pp. 172
[32] Ibid. p. 173.
[33] Ibid.
[34] Ibid.
[35] Ibid. p. 174. Davidson holds that since all psychological events are also physical events, they can always be described in a physical vocabulary, even when they can be given a further psychological explanation, such as in terms of reasons for an action, for example.
[36] Recall "Given that the man *believed* the stick was a danger if left on the path, and had a *desire* to eliminate the danger, it was *reasonable* to remove the stick." My italics.

reasonable: they make up the *rational core* of the irrational action. Irrationality enters the picture at the point at which these elements are related, since what marks an irrational belief or action, according to Davidson, is that the logical relation between different belief structures is missing. In an irrational action, a belief is a cause of another belief without being a reason for it.

> In standard reason explanations […] not only do the propositional contents of various beliefs and desires bear appropriate logical relations to one another and to the contents of the belief, attitude, or the intention they help explain; the actual states of belief and desire cause the explained state or event. In the case of irrationality, the causal relation remains, while the logical relation is missing or distorted.[37]

In an intentional action of the irrational kind, the relation *between* the different belief-structures that make up the full action is not logical. In irrational action, one action causes another, such as when the action of moving the branch from the path to the hedge causes the second action of replacing the branch back on the path. There is, however, no logical relation between these elements of Mr. S' obsessive behavior; there can't be, since the belief that goes into the second action is inconsistent with the belief that resulted in the first action.[38] So there is a point at which actions that are irrational, in the sense of being incoherent and inconsistent, can only be accounted for in terms of a causal relation: that is, insofar as there exists no rational or logical relation between beliefs and/or desires, on the one hand, and the resultant action, on the other. Although the former causes the latter, there is no logical connection between them. Differently described, irrational acts are acts in which the otherwise rational subject breaks with the standards of rationality. I will look more closely at what this means in the section on self-deception, but let us first cast a glance on what it means to break with the standards of rationality thus far.

In Davidson's view, irrationality is the failure to act in accordance with the rational principles to which one normally adheres: it is a condition for having thoughts and intentions, as well as making judgments, that the basic standards of rationality are applied. To behave irrationally is to go against one's own principles. The explanation of an irrational act must retain what is typical of rational thinking and action to some extent: "the man who returns to the park to replace the branch has a reason: to remove the danger. But in doing this he ignores his principle of acting on what he thinks is best, all things considered."[39] Davidson explains:

> there is no denying he has a reason for ignoring his principle, namely that he wants, perhaps very strongly, to return the branch to its original position. This is

[37] Davidson, "Paradoxes of Irrationality" in *Problems of Rationality*, p. 179.
[38] Davidson seems to use 'logical relation' to stand for a consistent and coherent relation only so that in his vocabulary, inconsistency, for example, is not a logical relation.
[39] Davidson, "Paradoxes of Irrationality" in *Paradoxes of Rationality*, p. 178.

> the point at which irrationality enters. For the desire to replace the branch has entered into the decision to do it twice over. First it was a consideration in favor of replacing the branch, a consideration that, in the agent's opinion, was less important than the reasons against returning to the park. Given his principle that he ought to act on such a conclusion, the rational thing for him to do was, of course, not to return to the park. Irrationality entered when his desire to return made him ignore or override his principle. For though his motive for ignoring his principle was a reason for ignoring the principle, it was not a reason against the principle itself, and so when it entered in this second way it was irrelevant as a reason, to the principle and to the action.[40]

In Davidson's understanding Mr. S' action was irrational because, although he had judged that he had better reasons not to return to the park, he acted on his strong wish to return the branch to its original position. Although his wish was a reason for his action to return, this reason is not rational, since he had judged that it was, all things considered, best not to return, and, in being a rational being, he subscribes to the rational principle saying that one ought to act as one has, on the basis of the available evidence, judged best.[41] According to Davidson, the irrational subject apprehends the full situation, and finds reasons for performing the act he performs and judges (correctly) that the alternative action would be better; yet he acts on the former and thereby acts against his own principles.

> It is easy to imagine that the man who returned to the park to restore the branch to its original position in the path realizes that his action is not sensible. He has a motive for removing the stick, namely that it may endanger a passer-by. But he also has a motive for not returning; which is the time and trouble it costs. In his own judgement, the latter consideration outweighs the former; yet he acts on the former. In short, he goes against his own best judgement.[42]

In "Incoherence and Irrationality", Davidson writes that irrationality is to go against one's own best judgments; it is to act, think or feel counter to one's own conception of what is reasonable. By this he means that the basis for regarding an action as irrational is not anyone's judgment; rather, the standard from which the agent deviates is the normal condition of consistency and coherence in the agent's own behavior and thoughts.[43] The situation is transparent to the irrational actor and he judges correctly what to do, yet acts against his judgment: he willingly acts counter to what he knows is best.[44]

---

[40] Ibid.
[41] See, for example, "Deception and Division" in *Problems of Rationality*, p. 201.
[42] Ibid. "Paradoxes of Irrationality", p. 174
[43] Ibid. "Incoherence and Irrationality", p 189.
[44] Davidson ascribes the thought that no one willingly acts counter to what he knows to be best, and that only ignorance can explain foolish acts, to Socrates, and calls it the Plato Principle. Davidson objects to this Principle. Davidson says, in a quote that points ahead: "we must assume that the mind can be partitioned into quasi-independent structures, that interact in ways the

What is needed in order to explain irrationality is something that belongs to the mental but which is not a reason. Davidson uses the term *mental cause* to refer to something mental which causes an effect without being a reason. He writes:

> The difficulty in explaining irrationality is in finding a mechanism that can be accepted as appropriate to mental processes and yet does not rationalize what is to be explained. What makes trouble is that our normal way of explaining the formation of propositional attitudes, including intentions and intentional acts, is to state the reasons that caused the attitude or act. Thus many of Freud's explanations of apparently irrational thoughts and acts are intended to show that from the agent's point of view (enlarged to embrace unconscious elements) there were good reasons for his thinking or acting. The paradoxical consequence is that explaining irrationality necessarily employs a form of explanation which rationalizes what it explains; without the element of rationality, we refuse to accept the account as appropriate to mental phenomena. We look, or tend to look, not merely for causes or forces, but for causes that are reasons. To explain irrationality we must find a way to keep what is essential to the character of the mental – which requires preserving a background of rationality – while allowing forms of causality that depart from the norms of rationality. What is needed to explain irrationality as a mental cause of an attitude, but where the cause is not a reason for the attitude it explains.[45]

In other words, the rational core of the intentional act must be preserved in the explanation of the irrational act too, but the notion of a mental cause makes it possible to account for the deviation from the rational in a mental process: it should allow for an explanation that does not rationalize the irrational action beyond the rational elements that are there. Davidson made another attempt at explaining what a mental cause might be in "Paradoxes of Irrationality".

> In the cases of irrationality we have been discussing, there is a mental cause that is not a reason for what it causes. So in wishful thinking, a desire causes a belief. But a judgement that a state of affairs is, or would be, desirable, is not a reason to believe that it exists.
>
> It is clear that the cause must be mental in this sense: it is a state or event with a propositional content. If a bird flying by causes a belief that a bird is flying by (or that an airplane is flying by) the issue of rationality does not arise; these are causes that are not reasons for what they cause, but the cause has no logical properties, and so cannot of itself explain or engender irrationality (of the kind I have described).[46]

The cause must be given a mental description, says Davidson, since in a non-mental description we lose touch with what is needed to explain elements of irrationality. Davidson's dilemma is this: "if we think of the cause in a neutral

---

Plato Principle cannot accept or explain." (Davidson, "Paradoxes of Irrationality" in *Problems of Rationality*, p. 181)

[45] Davidson, "Incoherence and Irrationality" in *Problems of Rationality,* p. 190.

[46] Ibid. "Paradoxes of Irrationality", p. 179.

mode, disregarding its mental status as a belief or other attitude – if we think of it merely as a force that works on the mind without being identified as part of it – then we fail to explain, or even describe irrationality. Blind forces are in the category of the non-rational, not the irrational."[47] For something to be a mental cause, it must be a propositional attitude: a belief, desire etc. with a propositional content. In trying to account for what a mental cause can be in order both to explain irrationality and still be a part of the mind, Davidson asks us to consider a case of social interaction about which it is unproblematic, he claims, to say that one mental event causes another mental event without being a reason for it.

> There is, however, a way one mental event can cause another mental event without being a reason for it, and where it is no puzzle and not necessarily any irrationality. This can happen when cause and effect occur in different minds. Wishing to have you enter my garden, I grow a beautiful flower there. You crave a look at my flower and enter the garden. My desire caused your craving and action, but my desire was not a reason for you craving, nor a reason on which you acted.[48]

Davidson suggests here that there is a causal relation between a mental event in A's mind, A's desire, and a mental event in B's mind, B's craving: the latter is an effect of the former (or, an effect of the resultant action of A's desire). But it is clear that A's desire cannot be a reason for B's craving, since B need not even know of A's desire. Davidson concludes: "Mental phenomena may cause other mental phenomena without being reasons for them, then, and still keep their character as mental, provided cause and effect are adequately segregated."[49] Davidson continues: "I suggest that the idea can be applied to a single mind and person. Indeed, if we are going to explain irrationality at all, it seems we must assume that the mind can be partitioned into quasi-independent structures that interact in ways the Plato Principle cannot accept or explain."[50]

Davidson claims that in order to account for how a rational person can hold beliefs and other propositional attitudes that are incoherent, it must be assumed that the mind can be divided. I will return to this claim later. At this juncture, it will suffice to note that Davidson holds that irrational actions, like all actions, are intentional and therefore can be explained in terms of reasons. These consist of a rational core of elements of belief-desire-pairs and actions, and the irrationality consists in that the relation between the rational elements of the action can be given a causal description but not a logical one.

I have presented parts of Davidson's conception of irrationality in general that are of relevance to his account of self-deception. I will now turn to that

---

[47] Ibid. p. 180.
[48] Ibid. p. 181.
[49] Ibid.
[50] Ibid.

account. My discussion presupposes that Davidson's theory of radical interpretation, as well as his analysis of irrationality, are borne in mind. The central texts for what follows are, in particular, Donald Davidson's papers "Deception and Division" (1986), where the problem of self-deception is at issue; "Who is Fooled" (1997), where Davidson further expounds on the discussion of self-deception, and "Incoherence and Irrationality". I will also refer to the paper "Paradoxes of Irrationality" (1982). In this chapter, I will investigate some central assumptions that Davidson makes in his outline of self-deception. My aim is to show why Davidson's account of self-deception takes the form it takes, i.e., why he makes the assumptions that he makes and how these assumptions are related.

Self-Deception as Paradox

Self-deception, like irrationality in general, seems paradoxical to Davidson. His aim in the paper "Deception and Division" is to explain how self-deception is possible. According to Davidson, self-deception requires that one hold contradictory beliefs. In laying out the conditions for self-deception, he says: "In the sort of self-deception that I shall discuss, a belief like that reported in (1) 'D believes that he is bald' is a causal condition for a belief which contradicts it, such as (2) 'D believes that he is not bald'."[51] The problem is to explain what separates this from believing (3) "D believes that (he is bald and he is not bald)." A belief like (3) cannot be accepted because nothing a person could say or do would count as good enough grounds for the attribution of a straightforwardly and obviously contradictory belief."[52] This would be to break with the principle of non-contradiction. Instead, Davidson says: "it is possible to believe each of two statements without believing the conjunction of the two."[53] He holds this to be the case regarding self-deception.

In outlining the conditions for his discussion of self-deception, he says: "We have the task, then, of explaining how someone can have beliefs like (1) and (2) without putting (1) and (2) together, even though he believes (2) *because* he believes (1)."[54] In "Who is Fooled?" Davidson expands on this: "What is important is that to be self-deceived one must at some time have known the truth, or, to be more accurate, have believed something contrary to the belief engendered by the deception. […] This original knowledge must, of course, have played a causal role in self-deception."[55]

Davidson's account of self-deception thus proceeds from the assumption that the self-deceiver holds two incoherent and inconsistent beliefs true at the

[51] Ibid. "Deception and Division", p. 199.
[52] Ibid. p. 200.
[53] Ibid.
[54] Ibid.
[55] Ibid. "Who is Fooled?", p. 216.

same time. This poses a threat to the understanding of persons as fundamentally rational: how can the mind – which Davidson describes as "the house of reason"[56] – harbor irrational thought? This is the paradox that Davidson tries to solve in his paper "Deception and Division". In his account, the former belief causes the second. It is not the self-deceiver's experiences that interest Davidson, nor is he particularly concerned with how a person in a state of self-deception behaves. He does not wish to describe self-deception, but rather to explain how it is possible. In the essay "Who is Fooled?", he comments upon his own work on self-deception thus: "This highly abstract account of the logical structure of self-deception never was intended as a psychologically revealing explanation of the nature of self-deception. Its modest purpose was to remove, at least mitigate, the features that at first make self-deception seem inconceivable."[57] In "Deception and Division", he compares the focal point in his own account with the accounts of two other analytical philosophers, David Pears and Kent Bach, noting: "*To me it seems important to identify an incoherence or inconsistency in the thought of the self-deceiver*; Pears and Bach are more concerned to examine the conditions of success in deceiving oneself."[58] The incoherence, or inconsistency, in the thought of the self-deceiver is the fundamental problem, as Davidson sees it: it is what distinguishes irrational behavior and thought, of which self-deception is an instance, from normal rational behavior and thought. The question is not how a rational person can think and act irrationally, but how a rational person can think irrationally, *assuming that this means to hold incoherent and inconsistent beliefs*. Davidson is not interested in the psychological *per se*. His problem is to explain how a rational person's thoughts can harbor inconsistencies.

We have seen Davidson delineating the structure of self-deception: it is to hold two incoherent beliefs true at the same time. I want to begin by asking why self-deception should be understood in this way at all; what motivates Davidson to ascribe this structure to self-deception? While it is standard to conceive of self-deception as paradoxical, it is especially common in the analytical tradition. This is exemplified in Jean-Pierre Dupuy's anthology on self-deception, *Self-deception and Paradoxes of Rationality*, which includes Davidson's article, "Who is Fooled?" The articles bear titles such as "Two Paradoxes of Self-Deception" and "(Apparent) Paradoxes of Self-Deception". As the titles suggests, self-deception is assumed to be constituted by one or more paradoxes. To start with, it is common that the initial outline or definition of self-deception includes a paradox. Alfred Mele, for example, provides the following formulation of the philosophical problem posed by self-deception:

[56] Ibid. "Paradoxes of Irrationality", p. 169.
[57] Ibid. "Who is Fooled?", p. 221.
[58] Ibid. "Deception and Division", p. 210. My italics.

> If ever a person *A* deceives a person *B* into believing that something, *p*, is true, *A* knows or truly believes that *p* is false while causing *B* to believe that *p* is true. So when *A* deceives *A* (i.e. himself) into believing that *p* is true, he knows or truly believes that *p* is false while causing himself to believe that *p* is true. Thus, *A* must simultaneously believe that *p* is false and believe that *p* is true. But how is this possible?[59]

Self-deception here is explicitly understood on the model of deceiving someone else. As *A*, in deceiving *B*, causes *B* to believe what *A* himself knows is false, the self-deceiver knows that something is false but causes himself to believe that it is true. Davidson's account too follows this model. It is easy to see why such expressions as 'to deceive oneself' and 'to lie to oneself' invite the thought that to deceive oneself (to lie to oneself) is the same as to deceive (lie to) someone else, with the crucial difference that the former takes place within a single mind. The surface grammar of self-deception invites this analogy. Davidson draws on the analogy between lying to another and lying to oneself in "Who is Fooled?" when he asks: "Is it possible to lie to oneself? In trying to answer this question, we must first ask what is involved in telling a lie to anyone. Telling a lie pretty clearly requires a speech act performed with the intention that someone be deceived, that is, mislead with the respect to the truth of some proposition."[60] This is also the structure that Davidson ascribes to self-deception, with the exception that self-deception is not a speech act: as we will see, Davidson holds that self-deception is indeed an act performed with the intention to deceive oneself. What it means to deceive oneself then, is that one misleads oneself with respect to the truth of a proposition.

In "Deception and Division" Davidson provides an example of a typical case of self-deception and how it may come about. Because his outline of self-deception here reveals the features that he takes to be essential to self-deception, I will return to this example throughout this chapter and in my discussion of Davidson throughout the book.

> Carlos has good reason to believe he will not pass the test for a driving licence. He has failed the test twice before and his instructor has said discouraging things. On the other hand, he knows the examiner personally, and he has faith in his own charm. He is aware that the totality of the evidence points to failure. Like the rest of us he normally reasons in accordance with the requirement of total evidence. But the thought of failing the test once again is painful to Carlos (in fact the thought of failing at anything is particularly galling to Carlos). So he has a perfectly natural motive for believing he will not fail the test, that is, he has a motive for making it the case that he is a person who believes he will (probably) pass the test. His practical reasoning is straightforward. Other things being equal, it is better to avoid pain; believing he will fail the test is painful; therefore (other things being equal) it is better to avoid believing he will fail the

[59] Alfred Mele, *Irrationality: An essay on Akrasia, Self-Deception, and Self-Control* (New York: Oxford University Press, 1987), p. 121.
[60] Davidson, "Who is Fooled?" in *Problems of Rationality*, p. 214.

> test. Since it is a condition of his problem that he takes the test, this means it would be better to believe he will pass. He does things to promote his belief, perhaps obtaining new evidence in favour of believing he will pass. It may simply be a matter of pushing the negative evidence into the background or accentuating the positive. But whatever the devices (and of course there are many), core cases of self-deception demand that Carlos remain aware that his evidence favours the belief that he will fail, for it is awareness of this fact that motivates his efforts to rid himself of the fear that he will fail.[61]

In accordance with Davidson's example of Carlos, which he takes to be a typical or "core" case of self-deception, the self-deceiver wants to pass the test, but he knows that he has better reasons to believe he will not pass. As Davidson would say, "he is aware that the totality of the evidence points to failure." Carlos has a desire to pass the test, but he knows his chances are slim and holds the belief that he will not pass the test. Despite the fact that he is aware that the total evidence point to failure, he induces a belief in himself that he will pass, a belief which is in line with his desire and which allows him to avoid the pain of thinking that he will fail. It is the former belief, the belief that he will fail, which causes the latter, irrational, belief that he will pass; and the former belief must remain throughout the deception, since what motivates Carlos' self-deception is his desire to avoid the pain inflicted by the original belief. Davidson means that Carlos' fear of failing is a perfectly natural motive for him to embark upon a process of practical reasoning directed at coming to hold the belief that he will pass, but inducing the belief that he will pass in himself stands in conflict with the general principles of rational reasoning. This is what makes it irrational.

Davidson asserts: "Like the rest of us he normally reasons in accordance with the requirement of total evidence." What does it mean to reason in accordance with the requirement of total evidence for inductive reasoning? Briefly, it means to believe that which one knows that one has best reasons to believe: "when we are deciding among a set of mutually exclusive hypotheses, this requirement enjoins us to give credence to the hypothesis most highly supported by all the available relevant evidence."[62] The person who has evidence both for and against a hypothesis and judges that, relative to all the evidence available to him, the hypothesis is more probable than not, can, however, still fail to accept the hypothesis which in his own judgment is the most probable.[63] Davidson calls this *weakness of the warrant*, and holds that self-deception manifests weakness of the warrant.[64]

In "Incoherence and Irrationality", Davidson expands on what he means by "subscribing to a principle". He says that one should not understand these principles as something to which one *ought* to subscribe. He writes: "For I think

[61] Ibid. "Deception and Division", p. 209.
[62] Ibid. p. 201.
[63] Ibid.
[64] Ibid. p. 209.

everybody does subscribe to those principles, whether he knows it or not. This does not imply, of course, that no one ever reasons, believes, chooses, or acts contrary to those principles, but only that if someone does go against those principles, he goes against his own principles."[65] Rather than being prescriptive, these principles express how we rational persons normally reason. He continues: "These are principles shared by all creatures that have propositional attitudes or act intentionally; and since I am (I hope) one of those creatures, I can put it this way: all thinking creatures subscribe to my basic standards or norms of rationality. […] it is a condition of having thoughts, judgements, and intentions that the basic standards of rationality have application."[66] According to Davidson, if one is a creature who has thoughts and intentions and who judges, one implicitly holds rational principles such as the requirement of total evidence for inductive reasoning. What one says or does, therefore, ought to meet the standards of rationality. (Let us recall the assumption that grounds the project of radical interpretation: the interpreter must assume that the speaker's utterances and thoughts are guided by normative principles, and this must mean that he reasons as is correct by *our* lights.)

According to Davidson's analysis, Carlos has a motive for his action of making himself believe that he will pass: to avoid the pain of believing that he will fail. Carlos acts on a motive (which Davidson holds to be a reason, as we saw in his analysis of irrational action). Still, Carlos is not acting rationally. Another reason overrules the reason to act so as to avoid pain. While he normally reasons in accordance with the requirement of total evidence for inductive reasoning, Carlos now makes himself believe that he will pass, and takes himself to have a reason for doing so, that it is better to avoid pain, although he knows that the evidence points to his eventual failure. The reason why Carlos' self-deception is irrational, according to Davidson, is that he doesn't hold the belief that he knows that he has best evidence to hold, and this is in conflict with the requirement of total evidence for inductive reasoning which requires him to do so. It is typical of core cases of self-deception that the self-deceiver holds the rational belief to be the most probable. He also holds that he should accept what in his own judgment is most probable. But he doesn't accept this. Instead he induces in himself a belief of what he wishes to be the case. Thus the self-deceiver holds both the rational belief and the irrational one, which is incoherent and inconsistent with the first.

We have seen that Davidson's account fits in with Mele's definition of self-deception as, first, analogous to deception, and second, paradoxical. According to Davidson, the self-deceiver holds a belief and its contradiction true at the same time and that the initial, rational belief causes the second, irrational one. We are now in a position to more clearly see exactly why self-deception is construed by Davidson and others as a paradox: if one defines self-deception *as*

[65] Ibid, "Incoherence and Irrationality", p. 195.
[66] Ibid.

a paradox (i.e. the holding of two contradictory beliefs as true), then it will appear paradoxical. But why should we define it thus?

Self-Deception, Intention and Evidence

In this section, I will continue examining the supposed analogy between lying to someone else and lying to, or deceiving, oneself. According to Davidson, they have in common that they are both intentional actions. In what follows, I will work out the theoretical underpinnings of this definition, as well as problems arising out of it. In "Deception and Division", where Davidson is describing what he takes self-deception to be by distinguishing it from wishful thinking, he writes:

> it is not self-deception simply to do something intentionally with the consequence that one is deceived, for then a person would be self-deceived if he read and believed a false report in a newspaper. *The self-deceiver must intend the deception.*
>
> To this extent, at least, self-deception is like lying; there is intentional behavior which aims to produce a belief the agent does not, when he institutes the behavior, share. *The suggestion is that the liar aims to deceive another person, while the self-deceiver aims to deceive himself.* The suggestion is not far wrong. I deceive myself as to how bald I am by choosing views and lighting that favour a hirsute appearance; a lying flatterer might try for the same effect by telling me I am not all that bald.[67]

As the liar intends to deceive the one to whom he lies, the self-deceiver intends to deceive himself, according to Davidson, and to intend to deceive oneself is to aim to produce a belief in oneself which one does not initially hold. As we have seen, this belief is a belief that the self-deceiver knows he is not justified in holding. The requirement that the self-deceiver must intend the deception thus seems to imply that for the action to be self-deception, it must be intentional under the description 'self-deception' (for something to be an action, it must be intentional under *some* description). The fact that the self-deceiver intends the deception makes it similar to lying: just as the deceiver aims to produce a belief in another person that is the opposite of that which the liar himself holds true, the self-deceiver aims to produce a belief in himself that is the opposite of what he knows he has best evidence to hold true. Davidson says: "*All that self-deception demands of the action is that the motive originates in a belief that p is true (or recognition that the evidence makes it more likely to be true than not), and that the action be done with an intention of producing a belief in the negation of p.*"[68]

In Davidson's account, the self-deceiver chooses to form a belief that he knows that he is not justified in holding. In other words, he holds a belief that

[67] Ibid. "Deception and Division", p. 207. My italics.
[68] Ibid. p. 208. My italics.

he knows to be irrational because he wants to believe it. The philosopher Ariela Lazar questions this picture of the self-deceptive subject as someone who chooses. She suggests that there are strong reasons for supposing that most cases of self-deception are not best explained in this way. One way in which she does this is by showing the great disadvantages associated with the holding of an irrational belief. I think she is on to something important here, and I will therefore recapitulate the most salient points in her paper "Division and Deception: Davidson on Being Self-Deceived".

Assume that the evidence at my disposal indicates that I will fail a certain test, yet I believe that I will pass. Lazar says: "According to Davidson's account, the formation of the irrational belief is due to my forming an intention to form the belief in question. I choose to form the belief because I think it will enhance some goals of mine – relieve anxiety, boost self-confidence, etc. The portrayal of the self-deceptive subject as *choosing* to form the belief is an essential element of this account."[69] Lazar asks if it is reasonable to portray the subject of self-deception in this way, as choosing to form the irrational belief. In Carlos' case, holding the belief that he will pass is likely to *decrease* his chances of actually passing the test, since this belief might induce, for example, a sense of complacency which would make him less inclined to prepare for the test. Rendering self-deception as a matter of choice is problematic because choosing to hold an irrational belief reduces the subject's chances of satisfying the initial desire. In Davidson's portrayal, the situation is transparent to the subject: he knows what he has best reasons to believe, and he knows that it is not what he wants to believe. How then can he hold that it is better to believe what he judges to be false? Rendering self-deception as a matter of choosing poses a problem for Davidson's account, according to Lazar: "Rather than wonder why (and how) the irrational belief is formed, we end up wondering why (and how) the irrational decision is made".[70] Davidson's account provides an explanation for how Carlos comes to believe that he will pass in spite of evidence to the contrary, but it leaves us wondering: why, if it is so important for him to pass the test, does he choose to fulfill his desire to avoid pain instead of admitting to himself that he will have to work very hard if he is to have any chance of passing the test? Lazar notes that there are two desires at play in what she calls "the standard approach to self-deception", which she takes Davidson's account to exemplify. This approach, she says, "emphasizes the centrality and the intensity of the desire that caused self-deception. This account cannot, without losing much of its appeal, view the self-deceptive subject as choosing to satisfy another desire at the cost of decreasing her chances of satisfying the initial desire which caused self-deception."[71]

---

[69] Ariela Lazar, "Division and Deception: Davidson on Being Self-Deceived" in *Self-Deception and Paradoxes of Rationality*, ed. Jean-Pierre Dupuy (Stanford: CSLI Publications, 1998), p. 23.
[70] Ibid. p. 24.
[71] Ibid.

What troubles Lazar is the rendering of the self-deceptive subject as choosing to satisfy the desire to avoid pain rather than aiming to satisfy the desire to pass the test. Lazar's point is that if we accept Davidson's explanation of how the irrational belief is formed – that we choose to hold the irrational belief rather than the belief we know that we have best reasons to believe – we still wonder *why* and *how* the self-deceiver makes the irrational decision. She says: "Many cases of self-deception are such that, if understood as involving an intention to form a belief, they must be viewed as involving a crazy choice."[72]

I agree with Lazar's critique of the character of self-deception as an intentional act in which the self-deceiver chooses to make herself believe the irrational belief, but I want to show that there is another, equally serious, problem with Davidson's account, one which arises before Davidson brings in intention. "Core cases of self-deception", Davidson writes, "demand that Carlos remains aware that his evidence favours the belief that he will fail." In another passage cited earlier, Davidson says: "A has evidence on the basis of which he believes that *p* is more apt to be true than its negation." I want to take issue with the idea that the self-deceiver has evidence for the rational belief, especially the notion that intentional action rests on such evidence.[73]

Davidson says that the person in a state of self-deception must recognize that evidence speaks in favor of a belief contrary to the one he holds. In the example of Carlos, we saw that, according to the evidence he accepts, he knows that he is more likely to fail the test. Still Carlos holds the irrational and self-deceptive belief that he will pass. The "error" here is not that he holds a false, distorted belief, but that, though he rightly judges that relative to the evidence available to him he will fail the test, he does not accept this. Instead, his judgment that the proposition "I will not pass the test" is true causes him to seek, favor and emphasize evidence that point to the falsity of that proposition.

Davidson makes a sharp distinction between weakness of the will and self-deception. Weakness of the will consists in irrational intentions (and perhaps action) in the face of conflicting values, while weakness of the warrant involves an irrational belief in the face of conflicting evidence. Davidson claims that weakness of the warrant is not a matter of simply overlooking evidence one has, nor a matter of not appreciating the fact that things one knows or believes constitute evidence for or against a hypothesis. Weakness of the warrant is rather a matter of departing from a custom or habit of accepting the requirement of total evidence (to give credence to the hypothesis most highly supported by all the available relevant evidence).[74] In the case of self-deception,

---

[72] Ibid.

[73] Lazar doesn't discuss evidence in detail, but when she mentions it, she uses the term slightly differently than Davidson does. In developing a parallel example to the one of Carlos, she says: "The evidence at my disposal indicates that I will fail, yet I believe I will pass." (Lazar, "Division and Deception", p. 23) I will return to the difference between having evidence on the basis of which one acts, judges etc. and having evidence at one's disposal.

[74] Davidson, "Deception and Division" in *Problems of Rationality*, p. 201-204.

Davidson claims that the person recognizes the weakness of the warrant; he knows he has better reasons for accepting the negation of the proposition he accepts. To this Davidson adds:

> Weakness of the warrant always has a *cause*, but in the case of self-deception weakness of the warrant is self-induced. It is no part of the analysis of weakness of the warrant or weakness of the will that the falling off from the agent's standards is motivated (though no doubt it often is), but this is integral to the analysis of self-deception.[75]

Self-deception runs deeper than mere weakness of the warrant. The person who is self-deceived must have a *reason* for weakness of the warrant; he must have a reason for not accepting the evidence, and he must have played a part in bringing weakness of the warrant about. Since it is self-induced, it must be something the self-deceiver does intentionally. The motivation for Carlos' falling away from his own rational standards in believing something although he knows that he has better evidence to believe the opposite, as we have seen, is that having the thought that he will fail the test is painful to him. According to Davidson, this is the reason for judging that it is best to believe that he will pass and to go on inducing this belief in himself. Here Davidson sees a close connection between self-deception and lying:

> To this extent, at least, self-deception is like lying; there is intentional behavior which aims to produce a belief the agent does not, when he institutes the behavior, share. The suggestion is that the liar aims to deceive another person, while the self-deceiver aims to deceive himself.[76]

I want to dissect Davidson's claim that the self-deceiver holds the belief about which he aims to deceive himself. I will use the example in which Davidson deceives himself about his baldness to do this. Davidson says that he first held the belief that he was bald, and that the self-deception consisted in producing the belief that he was not bald, a belief which he did not hold when he initiated his self-deception. But why should we assume that he held the belief that he was bald at the beginning of self-deception, and, further, that he knew that evidence favored this belief? The example doesn't suggest this. "I deceive myself as to how bald I am by choosing views and lighting that favor a hirsute appearance." Rather than assuming that self-deception starts with the belief that he is bald (together with knowledge that this is what the total evidence points to), I mean that the example suggests rather that the person who deceives himself is clinging to a picture or idea of (or a belief about) himself as someone with hair – as someone who is not bald. I suggest that self-deception consists in *avoiding* the formation of the belief with which Davidson suggests that self-deception begins.

---

[75] Ibid. pp. 204.
[76] Ibid. p. 207.

In Davidson's view, the self-deceiver must first know that evidence favors the belief *p*, then form an intention to change his belief through a process of practical reasoning for his behavior to count as self-deceptive. But is it really typical of self-deception that the self-deceiver first has evidence for a certain hypothesis, and then "makes a deceptive move"? Isn't self-deception rather the term we use to describe a distorted view of the situation one is in in the first place?

I will now return to Lazar's formulation, "the evidence at my disposal", and compare it with Davidson's, "A has evidence on the basis of which…". The difference in their formulations reflects a difference in how they interpret the role of evidence. Davidson's formulation suggests that the self-deceiver is aware of the evidence, and that it plays a significant role for his actions, while Lazar's formulation invites the reading that the evidence lies open to the self-deceiver's view, although she is not aware of it, or doesn't see it as evidence for or against a belief (or proposition). Although I take Lazar's formulation to point in the right direction by suggesting that the self-deceiver is unaware that she would be more justified in believing the contrary to what she believes, I have hesitations about the use of 'evidence' in this case. For what is evidence, if it is not evidence *for someone*? If the self-deceiver is not aware of there even existing any "evidence" of something, why use the term 'evidence' at all? In this respect, I argue that it is confused and confusing to say that self-deception is a state in which one deceives oneself about a state of affairs for which one has evidence. To the contrary, when we say of someone that she is deceiving herself, what we want to express is rather that she seems blind to what everyone else can see, as if she's turned her head away from a reflection of herself *so as to avoid* acknowledging something that she somehow senses that she doesn't want to see. In short, she doesn't want *to begin* to know. For something to count as evidence, it must be taken into account *by someone* who sees it as evidence, i.e. who is prepared to judge a situation, and wishes his judgment to be well-founded. The self-deceiver, by contrast, is not prepared to assess his situation at all. Her self-deception, one might say, consists in her instinctive unwillingness to reason at all, to see something as requiring reflection or judgment. For her, evidence is entirely irrelevant.

Self-Deception as Incoherency of Beliefs

As we saw in the last section, according to Davidson, self-deception begins with the belief *p* that motivates the person to form an intention to make himself believe not-*p*. But *p* doesn't simply give way to not-*p*. Davidson states in the opening paragraph: "Self-deception is notoriously troublesome, since in some of its manifestations it seems to require us not only to say that someone believes

both a certain proposition and its negation, but also to hold that one belief sustains the other."[77] In discussing the example of Carlos, Davidson provides a characterization of what this means: "Core cases of self-deception demand that Carlos remains aware that his evidence favours the belief that he will fail, for it is awareness of this fact that motivates his efforts to rid himself of the fear that he will fail."[78] The belief not-*p* coexists with the belief *p* in self-deception. Davidson explains the requirement that the rational belief be maintained to sustain the irrational belief by saying that it is needed to push the negative evidence into the background or accentuate the positive.[79]

In the last section, I criticized the idea that the person in a state of self-deception initially holds the rational belief and is aware that evidence favors the truth of this proposition. Here I will focus on this other condition, which Davidson takes as typical for self-deception: that the rational belief accompanies and sustains the irrational belief throughout the self-deception. If we reject the idea that self-deception starts with the self-deceiver holding the rational belief at the outset (while being aware that evidence supports this belief), as I suggest that we should do, we see that what Davidson takes to be "notoriously troublesome" with self-deception – that the self-deceiver holds contradictory beliefs – is dissolved as a problem because the presupposition on which it rests is rejected. If we do not affirm, with Davidson, that the self-deceiver is aware of *p* (or at least aware of *p* to the degree that he has evidence that *p* is more likely than not-*p*), Carlos' self-deception cannot be described as ridding himself of awareness of his belief that evidence points to his failing. In rejecting the idea that Carlos initially holds the belief that he will fail, the idea that self-deception involves two contradictory beliefs, where the first, rational belief sustains the latter, irrational belief, is also rejected.

In Davidson's account, it is *essential* for an action to count as self-deception that it originates in holding a proposition true which the self-deceiver comes to deny, and that one recognizes that the evidence points to the truth of that proposition. But *why* does Davidson place this requirement on "core cases" of self-deception? Why must the self-deceiver first hold a belief true and then go on to mislead himself about its truth in an intentional action? In order to understand why this assumption is central to Davidson's account, we must examine Davidson's appeal to intention in explaining self-deception.

Two questions arise here: Is self-deception an intentional action? If it is an intentional action, what does that imply? Does it imply that the self-deceiver must have a reason for the action of deceiving himself? What can self-deception be, if it is not an action? Is it an instinctive reaction or is it perhaps behavior? However one wishes to explain its genesis, it would seem to be almost definitive of self-deception that it consists in *avoidance* of some kind, for example, of

[77] Ibid. p. 199.
[78] Ibid. p. 209.
[79] Ibid.

being confronted with some unpleasant or disturbing truth and the anxiety that it provokes. Instead of thinking of such avoidance as an action supported by a reason explanation, as Davidson does, we can think of it as similar to how an animal can react by taking flight when it sees a larger animal or the light on an oncoming car, i.e., as something which scares it. The animal does not act for a "reason", but it is strongly motivated to take flight. "Anxious behavior" or an "anxious reaction" may nevertheless be described as intentional in the sense that it is directed at something (or from something: the anxiety-provoking animal or object). It is not simply a reaction on par with the physical reaction of separation of the elements oil and water when they are poured together. Neither is it a behavior as behavior can be used to describe inanimate objects, such as the behavior of a ship in a storm. It is the reaction of a creature who directs its attention at something in perception and action. If all that is required of an action is that it is intentional in *this* sense, I believe that it is unproblematic to see self-deception as an action. I will suggest that it is typical of self-deception that the attention is directed away from that which makes the person anxious, and is in this sense intentional. Intention is normally construed as the direction of attention at or toward something, a kind of focusing or delimiting of scope. But we could expand the notion to include a specific directing *away from* or *avoidance of* something that a self-conscious being could perform as a kind of higher-level instinctual behavior in which he is not merely reacting physically. On such a definition, an "instinctive reaction" or "intentional flight" would be a characteristic behavior of beings capable of thought (reason) without itself being "rational", i.e. a consequence or result of thinking; it is a behavior rather than an intentional action, but a quintessentially human one.

Mark Johnston puts forth a similar view in his paper, "Self-Deception and the Nature of Mind", where he criticizes the approach to self-deception that he lets Davidson's account exemplify, and introduces his own understanding of self-deception. Johnston suggests that we view self-deception as *a purposeful reaction* and not as an intentional and deliberate act. Although the mental processes that go into self-deception serve some interest of the self-deceiver, Johnston argues that they are not necessarily initiated by the self-deceiver for the sake of those interests or for any other reason. He says: "If we call mental processes that are purposive but not initiated for and from a reason *subintentional* processes then we can say that *our over-rationalization of self-deception consists in assimilating subintentional processes to intentional acts*, where an intentional act is a process initiated and directed by an agent *because he* recognizes that it serves specific interests of his."[80] Self-deception is not an act performed *so as to achieve* anything. Still it serves a purpose (what this purpose is, in Johnston's view, I will return to at the end of this chapter). Johnston holds that over-rationalization is a problem in many accounts of self-deception.

[80] Mark Johnston, "Self-Deception and the Nature of Mind", in *Perspectives on Self-Deception*, , p. 65. My italics.

He says: "The suggestion I wish to explore is that the surface paradox and deeper paradoxes of self-deception (i.e. those developed by Bernard Williams, by Sartre in a different passage, and by Donald Davidson) arise because as theorists of self-deception *we tend to over-rationalize mental processes that are purposive but not intentional*."[81]

I agree with Johnston here: in explaining self-deception, Davidson over-rationalizes it. According to Davidson, self-deception is an intentional action and, as such, it must have a rational core: there must always be a rationalizing explanation to be given for why the agent acts as he does, since the agent acts on reasons throughout the irrational action. In Davidson's conception, intentional actions *are* rational actions.

I will now turn to Elisabeth Anscombe's *Intention*, one of the most important texts on the topic of intention written in the 20th century and one which has informed Davidson's account – in order to address two questions. First, must it be possible to rationalize an intentional action? Second, is self-deception an intentional action? I will start with the first question.

Anscombe and Intentional Action

Anscombe lays down a condition that anything that is an intentional action must fulfill: intentional actions are actions to which a certain sense of the question "Why?" has application. Where it is legitimate to ask, perhaps in retrospect: "Why did you do x?", we are dealing with an intentional action. She says:

> What distinguishes actions which are intentional from those which are not? The answer that I shall suggest is that they are the actions to which a certain sense of the question "Why?" is given application; the sense is of course that in which the answer, if positive, gives the reason for acting. But this is not a sufficient statement, because the question "What is the relevant sense of the question 'Why?'" and "What is meant by 'reason for acting'"? are one and the same.[82]

Thus, Anscombe too holds that an intentional action is an action for which there is a reason, but it is not clear what a reason for acting would be. Anscombe goes on to discuss an example of an intentional action and what it means for an action to be intentional under one description but not under another.

> if you saw a man sawing a plank and asked 'Why are you sawing that plank?', and he replied 'I didn't know I was sawing a plank', you would have to cast about for what he might mean. […] Since a single action can have many

[81] Ibid.
[82] Elisabeth Anscombe, *Intention*, (Oxford: Basil Blackwell, 1976). §5, p. 9.

> different descriptions, e.g. 'sawing a plank', 'sawing oak', 'sawing one of Smith's planks', 'making a squeaky noise with the saw', 'making a great deal of sawdust', and so on and so on, it is important to notice that the man may know that he is doing a thing under one description, and not under another. […] He may know that he is sawing a plank, but not that he is sawing an oak plank or Smith's plank; but sawing an oak plank or Smith's plank is not something else that he is doing besides just sawing the plank that he is sawing. For this reason, the statement that a man knows he is doing X does not imply the statement that, concerning anything which is also his doing X, he knows that he is doing that thing.[83]

Anscombe's intention points to a central difficulty in Davidson's analysis of Carlos' self-deception. According to Davidson, Carlos is intentionally inducing in himself a belief which he knows he is not justified in holding. According to Davidson, Carlos also intends to deceive himself. Anscombe's analysis above, however, shows that the self-deceiver can be aware of doing something without being aware of what he is doing as falling under the heading (or description) of "self-deception". Even if he is aware of intentionally inducing in himself a belief which he knows is not supported by the evidence, he is not necessarily aware of any intention to deceive himself. I will argue, however, that the self-deceiver is typically unaware of what he is doing under either of these two descriptions. To do so, I will return to Anscombe.

As well as discussing examples of when the question "Why?" has application, Anscombe describes cases in which the question "Why?" does *not* have application. One such case is when the question "Why did you do x?" is answered by: "I was not aware I was doing that." Anscombe writes: "Such an answer is, not indeed a proof (since it may be a lie), but a claim, that the question 'Why did you do it (are you doing it)?', in the required sense, has no application."[84] If the question "Why?" in the relevant sense does not have application, the action was not performed with an intention. Anscombe later adds something that complicates this claim. The answer to the question "Why?" can be very weak and provide little reason.

> Now of course a possible answer to the question "Why?" is one like "I just thought I would" or "It was an impulse" or "For no particular reason" or "It was an idle action – I was just doodling". *I do not call an answer of this kind a rejection of the question.* The question is not refused application because the answer to it says that there is *no* reason, any more than the question how much money I have in my pocket is refused application by the answer "None".[85]

What Anscombe is saying is that there need not be a particular reason for an intentional action; still, it is typical of an intentional action that one can (that it is appropriate to) ask for a reason for it. While in the former case the person

[83] Ibid. §6, pp. 11.
[84] Ibid. §6, p. 11.
[85] Ibid. §17, p. 25. My italics.

was not aware of doing that for which she was asked a reason, in this case she is aware of the action but has no straightforward reason for it, and the question "Why?" has application.

I will consider Davidson's understanding of self-deception as an intentional action in the light of Anscombe's analysis. According to Davidson, as we have seen, it must be possible to provide a reason explanation for an intentional action. In his understanding, this explanation includes a desire, value or goal and a belief that, by acting as one does, one can promote the relevant desire, value or goal. Such a reason explanation must be possible to give even in the cases where the action is, in a deeper sense, irrational. There is a difference between Anscombe's claim that intentional actions are actions for which one can ask for reasons and Davidson's outline of self-deception. In Davidson's account, it is not enough that one can ask for reason; rather, it must be assumed that the self-deceiver has a *reason* for the intentional action. In the case of Carlos, this action consists in actively inducing in himself a belief for which the evidence, by his own lights, is weak. The reason is, as we have seen, the desire to avoid pain together with the belief that this action can promote his desire. The reason is not something that merely can be ascribed to the self-deceiver; it must be a reason on the basis of which the self-deceiver acts: it is what causes him to induce the belief in himself. And it is not enough "reason" for the self-deceiver to induce the false belief in himself "for no particular reason". For Davidson, the self-deceiver must have a reason for his intentional action which rationalizes it, even when the action is, like self-deception, irrational. Anscombe, on the other hand, claims that it is perfectly conceivable that a person has "no particular reason at all" for her action and yet that it is, nonetheless, an intentional action. I suggest that *if* we are to see self-deception as an intentional action, we ought to understand it as an action for which the self-deceiver need not be aware of any reason. But I also want to pose the question as to whether self-deception is best seen as an intentional action at all. I will stay with Anscombe in my consideration of this question.

Just as it is makes sense to ask how much money someone has in his pocket, it makes sense to ask for a reason for an intentional action. Anscombe continues: "An answer of rather peculiar interest is: 'I don't know why I did it'."[86] She says about such an answer that it "is appropriate to actions in which some special reason seems to be demanded, and one has none. It suggests surprise at one's own actions."[87] "'I don't know why I did it' may be said by someone who does not *discover* that he did it; he is quite aware as he does it; but he comes out with the expression as if to say 'It is the sort of action in which a reason seems requisite'. As if there were a reason, if only he knew it."[88] In Anscombe's view, there is no reason here, at least not in the relevant sense,

[86] Ibid.
[87] Ibid.
[88]Ibid. §17, p. 26.

"even if psychoanalysis persuades him to accept something as his reason, or he finds a reason in a divine or diabolic plan or inspiration, or a causal explanation in his having been previously hypnotized."[89] Anscombe emphasizes the peculiar character of such cases: "I myself have never wished to use these words in this way, but that does not make me suppose them to be senseless. They are curious intermediary cases: the question 'Why?' has and yet has not application; it has application in the sense that it is admitted as an appropriate question; it lacks it in the sense that the answer is that there is no answer."[90] In Anscombe's view, cases where the answer is 'I don't know why I did it' are better described as voluntary actions than as intentional;[91] nonetheless, as we see, it makes some kind of sense to her that one would call them intentional.

To summarize, in Anscombe's view, if the answer to the question "Why did you do X?" is "I was not aware of doing X", then the action is not intentional under the description X. Thus, if someone is asked, "Why did you deceive yourself?", and he answers that he wasn't aware of doing that, we cannot claim that he was intentionally deceiving himself. It is, however, conceivable that he would be aware of misleading himself about the truth of the proposition, and if asked, "Why did you mislead yourself about the truth of a proposition?", he would reply, "I found it more pleasant to think that I would pass". We could call this self-deception, but, if the man weren't aware of his action under this description, we wouldn't be justified in saying that he had an intention to deceive himself. The case where the person asked is aware of what she has done, but answers, "I don't know why I did X", is trickier; should we call this act intentional? It is typical of intentional actions that one knows one's reason for doing something, or knows that one had no particular reason. Neither is the case here. Here the person suspects that there is a reason for his action, but he doesn't know what it is. Is this typical of self-deception? I will discuss the possibility that it is in my discussion of Sigmund Freud's writings on self-knowledge in Chapter Three.

Anscombe's remarks above here served the purpose of problematizing Davidson's assumption that a person has a reason for every one of his intentional actions that rationalizes them (including irrational ones). Davidson wants to offer an explanation of self-deception that preserves a rational core in this irrational action. Thus he is motivated to look for a reason for every action, even if inducing in oneself a belief that one knows is false, or not well supported by the total evidence, can never be rationally justified. When Davidson argues that Carlos had a reason for inducing in himself a belief that he would pass, it plays the role of being the reason on the basis of which the agent acted. According to Davidson's story, Carlos takes avoidance of pain to be a good

[89] Ibid.
[90] Ibid.
[91] Ibid. I will not discuss the difference that Anscombe makes between a voluntary action and an intentional action here.

reason for the action of inducing the false belief in himself; moreover, the practical reasoning that goes into this action has a rational structure, since the belief-desire pair is rationally related to the action.

We see the same structure of explanation in Davidson's account of self-deception as in his analysis of Mr. S' obsessional action: both actions are depicted as consisting of "rational cores" consisting of a desire, a belief, and an action which rationally follows upon the belief-desire pair. The conception of irrational intentional behavior as consisting of rational structures makes it possible to provide an account of irrationality where the agent is rational "at the core", since anything that the agent does or believes is justified by a rationalizing structure: there is a reason for an action even when, in the grand scheme of things, it is truth-violating or irrational. I am arguing that this understanding of self-deception as rational "at the core" is problematic. To conceive of obsessional neurotic acts in this way, such as the one's described in the case of Mr. S, is all the more problematic since they are more removed from normal, rational action. This will be discussed in the next chapter.

Self-Deception Imbued with Presumptions of Rationality

Why should we assume there to be a *reason* for every move in self-deception? Why should we assume that self-deception must be rational "at the core"? And why should we assume that the self-deceiver apprehends the evidence at his disposal? We need to return to Davidson's definition of self-deception and his account of Carlos' self-deception in order to see how the idea that self-deception is rational at the core imbues Davidson's account of self-deception. In Davidson's account, self-deception begins with the self-deceiver first holding the rational belief, i.e., the belief that the self-deceiver knows is favored by total evidence. Davidson thus assumes that even an irrational action such as self-deception starts with the agent initially perceiving the situation in a sober fashion and judging it correctly. In short, he assumes that the situation is transparent to the self-deceiver. Evidence favors that Carlos will fail, and this is the belief he holds at first. I question these assumptions of Davidson's. What I have tried to show here is simply that the case of the self-deceiver Carlos seems rather to indicate that he fails to apprehend the situation correctly in the first place; it were as if he could hear the sounds but not the sense of his teacher's negative appraisal of his skills and the unlikelihood of his success. If this is an accurate picture of Carlos' initial situation, then it would be wrong to say, as Davidson does, that he *knows* that he has best evidence to hold the belief that he will fail. To the contrary, the idea that a *judgment* leads the self-deceiver to hold a rational belief, which he then rejects in favor of a more agreeable one, is not suggested by the behavior of the self-deceiver, but is rather assumed in Davidson's account.

According to Davidson, the self-deceiver first holds the belief *p* and then induces in himself the belief not-*p*. As we have seen, Davidson thinks that the self-deceiver must hold the former belief true throughout the deception, leading Davidson to draw the paradoxical conclusion that the self-deceiver holds the belief *p* and the belief not-*p*. Since Davidson wants to provide an explanation in which even the self-deceiver's action is rational "at the core", he needs to provide a rationalizing structure for holding the belief *p* and for holding the belief not-*p*. Davidson characterizes even the irrational move from *p* to not-*p* as an intentional action for which there is a reason explanation: the intentional action of avoiding pain is rationally supported by the belief that it is better to avoid pain. The only point at which the rational structure breaks down is when the self-deceiver takes himself to have a reason for his intentional action of deceiving himself (to avoid pain), which includes misleading himself about the truth of the proposition that he will pass the test; for he is not justified in misleading himself in this way. The move from *p* to not-*p* cannot be rationalized, i.e. it cannot be accounted for as a logical relation, but only as a causal relation: the belief that he will fail the test *causes* Carlos to induce in himself the belief that he will pass, but the former belief is not a *reason* for the latter. Still, the self-deceiver's reasons for inducing the false belief in himself remain untouched: he did so in order to avoid pain.

In Davidson's account of self-deception, elements central to rational reasoning are preserved. First, the self-deceiver perceives the situation as it is and makes a correct judgment. In Carlos' case, the judgment that he is more likely to fail the test. Second, there is a reason explanation to be given for each move in his action of deceiving himself. Recall Marcia Cavell's summary description of Davidson's account of rational action cited earlier in this chapter: "It has been argued that large-scale rationality on the part of the interpretant is an essential background of his interpretability." Cavell is referring here, in particular, to Davidson's account of meaning and interpretability, the so-called theory of radical interpretation. I now want to return to Davidson's account of radical interpretation to look at how well the two assumptions that he makes about the structure of self-deception correspond to the assumption that he makes to get his account of radical interpretation off the ground.

The Principle of Charity, as we have seen, consists of two assumptions: i) that the speaker reasons rationally, according to normative principles, and his attitudes make up a network of rationally related beliefs, desires, hopes etc. (coherence, holism); and ii) that beliefs should be determined by how the world is, in other words, that there is a correspondence between world and belief. I have questioned Davidson's assumption that self-deception begins with the self-deceiver holding the rational belief to be true. It is this claim that reveals self-deception to be so paradoxical for Davidson: the self-deceiver holds both the rational and the irrational belief true at one and the same time. I believe that by considering the assumptions Davidson makes in his theory of meaning, the

Principle of Charity in particular, we can better understand Davidson's motivation for assuming that the self-deceiver initially holds the rational belief.

We have seen that Davidson's project of radical interpretation requires that the interpreter can assume that the interpretant is rational (that our rational principles apply to him) and holds beliefs that corresponded to reality. If this cannot be assumed, it seems impossible that one can come to learn about the interpretant's beliefs and to understand the meaning of his utterances. Davidson's account of self-deception preserves the agent as, first and foremost, rational: self-deception is "rational at the core". The self-deceiver is rational in the sense that he perceives the world as it is, or, more precisely, as we do. He apprehends correctly that he is more likely to fail the test and forms that belief. Thus, his initial belief corresponds with reality. Davidson assumes that Carlos apprehends evidence both in favor of and against his passing the test, and judges correctly that the total evidence point to that he is most likely to fail. As understood by Davidson, self-deception poses no threat to the view of persons as first and foremost and on the whole rational beings, an assumption which, as we have seen, is a prerequisite for his project of radical interpretation. When self-deception is seen as misleading oneself about the truth of a proposition in this way, as analogous to lying, the self-deceptive subject is presented as someone who perceives the world as we do and who judges as we do: Davidson thus "preserves" the self-deceiver as a rational being.

I have discussed how the assumption that a subject must act according to rational principles and hold beliefs that correspond to how things are in the world leaves its mark on Davidson's account of self-deception. We have also touched upon the assumption of holism, i.e. the assumption that the interpreter must be able to presuppose that the attitudes of the interpretant are rationally connected, or connected to each other, more or less, as they are in us. Recall Davidson's remark from "Mental Events": "There is no assigning beliefs to a person one by one on the basis of his verbal behavior, his choices, or other local signs no matter how plain and evident, for we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations, and the rest." A requirement of radical interpretation is that the beliefs, intentions, fears etc. that the interpreter ascribes to the interpretant should fit in with other attitudes of his. The requirement of holism aids interpretation, since the beliefs, wishes, fears etc., which one has already mapped with utterances, stake out how other attitudes and expressions can and cannot be understood. Or, as Marcia Cavell writes in her introduction to *Problems of Rationality*: "Rationality comes with the propositional attitudes, since any one attitude means what it does, makes sense, only given its place in a network of other propositional attitudes, and only as they can more or less be mapped onto the interpreter's own norms of rationality."[92]

[92] Cavell in Davidson, *Problems of Rationality*, Introduction, xviii.

The requirement of holism in radical interpretation also leaves its mark on Davidson's account of self-deception in that the self-deceiver is assumed to have reasons for each and every move that he makes in deceiving himself. Indeed, as we have seen, in Davidson's account, all actions are intentional and accompanied by reasons that justify them (for example, even Mr. S' obsessive behavior/action is intentional and rational in this sense). The reasons are coherent structures – rational wholes – consisting of propositional attitudes and their logical relations. Even an irrational action has a rational core, insofar as it can be explained as consisting of different elements, or different actions for each of which there is a reason explanation consisting of a desire and a belief which justifies the action. Even the "irrational move" in Carlos' self-deception can be given a reason explanation: he comes to hold an irrational belief when he intentionally induces this belief in himself *because* it allows him to fulfill his *desire* to avoid pain, and he holds the *belief* that it is better to avoid pain. Thus Carlos has a *reason* even for inducing in himself a belief that he believes to be false. The reason explanation for the action consists of a desire and a belief, which together form a coherent and consistent whole. An irrational action such as self-deception or obsessive behavior cannot be given an explanation in which it is *one* rational whole, but it can and should be understood as consisting of rational wholes that are causally related, according to Davidson.

Jonathan Lear discusses Davidson's rational rendering of the case of Mr. S' obsessional neurosis (quoted on pages 25 and 26). As we saw, Davidson holds that Mr. S has a reason for putting the branch in the hedge as well as for going back and replacing it on the path; in both cases, he wants to prevent passers-by from getting hurt. Lear replies: "This does not seem the right description of Mr. S If he had decided the stick in the hedge was a danger, he would have had a reason to go back and remove it, but what reason could there be to *replace it on the path*? There seem to be a compulsion, not merely to remove the danger of the stick poking out of the hedge, but to restore it to its original position. Where is the reason in that?"[93] What sense is there in replacing the stick *on the path*? Why see this as a rational notion? There seems to be something arbitrary about the insistence upon the rationality of the action in this case. Lear objects that Davidson is rationalizing an act which is not rational, but compulsive. I agree with Lear. This *is* an over-rationalization. Davidson's penchant for over-rationalization comes out most clearly in his account of cases of obsessional neuroses, since obsessive behavior is so different from normal, rational action. But, I hold, over-rationalization is a problem also in Davidson's account of self-deception. Why then does Davidson account for it as a rational act? In analyzing the case of Mr. R, Lear notes that it poses a problem for Davidson's holistic view of the mental; "for there doesn't seem to be any perspective from which this behavior looks reasonable. We seem to have, rather, a particular form of irrationality, 'the failure, within a single person, of coherence or

[93] Jonathan Lear, *Freud. An Introduction* (New York: Routledge, 2006), p. 30. Later, *Freud*.

consistency in the patterns of beliefs, attitudes, emotions, intentions and actions'."[94] It is only by rationalizing each behavior in isolation that it is possible for Davidson to stay true, to some extent, to the idea of holism.[95]

Davidson's assumption of holism drives him to seek the rational core behind all actions and behavior. We have seen that he holds that even irrational actions, such as self-deception, have a rational core: they can be construed as rational wholes in that the desires, beliefs and the action to which they give rise are logically related. According to Davidson's holism, which is a necessary presupposition for radical interpretation, a propositional attitude makes sense only in a network of other propositional attitudes: we make sense of particular beliefs only as they cohere with other beliefs, preferences, intentions, expectations, and so on. In analyzing Carlos' self-deception, Davidson tries to make sense of how Carlos can believe that he will pass the test when it ought to be clear to him that he is very unlikely to pass. How can we understand why a rational person like Carlos believes something that is irrational (in that context) to believe? Davidson is trying to understand how it is that Carlos can hold the irrational belief "I will pass the drivers test". Davidson's first sense-making move is to understand the belief as induced from the contrary, rational belief, "I will not pass (am not likely to pass) the drivers test." I have suggested that Davidson makes this assumption because it preserves the understanding of a person's (initial) belief as corresponding to reality. Thus it also preserves the mind as, first and foremost, "a house of reason": irrationality enters only at a later stage.

Davidson's second move is to find a rationalizing structure even for the irrational belief, that is, a context in which it is reasonable to hold the irrational belief. The rationalizing structure consists of the desire to avoid pain and the belief that it is better to avoid pain, which, together with the action, form a coherent whole. Thus, we can makes sense of the irrational action (here, self-deception) such that we can assume that the irrational agent in fact apprehends the world as we do and, further, that his judgments about it share the same (rational) structure as our own. Moreover, even the irrational agent's attitudes are parts of a coherent whole. In perfectly rational action and thought, whatever they may look like, all the attitudes of the person are logically connected, while the attitudes of a self-deceiver are logically connected into rational clusters, but cannot combine to form a coherent whole.

Thus, Davidson's account of self-deception preserves the rationality of the agent by showing: a) that his beliefs correspond to reality; and b) that he acts intentionally and has reasons for his actions. Nonetheless, a problem remains. The self-deceiver's beliefs are inconsistent; they cannot be combined into a coherent, rational whole. The paradox remains paradoxical.

---

[94] Ibid. p. 26. The quote is to be found in "Paradoxes of Irrationality" in *Problems of Rationality*, p. 170.
[95] Lear, *Freud*, p. 26.

The question that remains is: if a person holds inconsistent and incompatible beliefs, how is it that he fails to put them together? Davidson explains:

> It would be a mistake to try to give a detailed answer to this question here. The point is that people can and do sometimes keep closely related but opposed beliefs apart. To this extent we must accept the idea that there can be boundaries between parts of the mind; I postulate such a boundary somewhere between any (obviously) conflicting beliefs. Such boundaries are not discovered by introspection; they are *conceptual aids to the coherent description of genuine irrationalities*.[96]

He continues:

> It is now possible to suggest as an answer to the question where in the sequence of steps that end in self-deception there is an irrational *step*. The irrationality of the resulting state consists in the fact that it contains inconsistent beliefs; the irrational step is therefore the step that makes this possible, the drawing of the boundary that keeps the inconsistent beliefs apart.[97]

The first thing to note is that Davidson proceeds from what he takes for granted, that is, his initial description of self-deception as consisting in maintaining closely related but opposed beliefs. As I have argued, this follows from the basic assumption in his account, i.e., that the self-deceiver initially holds a rational belief. There is a point at which Carlos believes that he will fail the test and knows that this is what he is justified by the available evidence to believe. It is this assumption that makes self-deception appear paradoxical, since it seems that the self-deceiver believes a contradiction. This would violate the most fundamental principle of rationality, the principle of non-contradiction, and thus this assumption would seem to conflict with the idea that human beings, even when seemingly irrational, are "rational at the core". Therefore, says Davidson, we must accept the idea that there can be boundaries between different parts of the mind. Davidson holds that it is possible to pick out an irrational step that constitutes the self-deception, that step being "the drawing of the boundary" that makes possible a state where inconsistent beliefs coexist.

Davidson seems to say that the self-deceiver draws this boundary. But a formulation in the earlier quote suggests that this is not how we should understand it. When Davidson writes that "such boundaries are not discovered by introspection; they are *conceptual aids* to the coherent description of genuine irrationalities", it seems we should see the boundaries as postulated in *accounting* for self-deception. In short, it is not the self-deceiver but Davidson who draws, or rather *postulates*, boundaries which divide the mind into parts, in

[96] Davidson, "Deception and Division" in *Problems of Rationality*, p. 211. My italics.
[97] Ibid.

order to be able to give what he sees as a coherent description of irrational actions. This interpretation fits well with Johnston's claim that the proponents of the interpretive view are only interested in providing an explanation that fits well, or reasonably well, with the holistic constraint. As we have seen in his analysis of Carlos, as well as in his analysis of Mr. S, Davidson accounts for irrational action as consisting of rational structures, which in themselves form coherent wholes. Davidson can retain this structure of explanation for irrational actions if he allows for a rupture in the rational. We have seen him account for this rupture by saying that only a causal relation can hold between the rational cores: the logical relation is missing. Here the rupture is accounted for as a boundary between the rational structures. While the former way of speaking – of logical (and causal) relations within the rational structures and only causal relations between them – clearly shows that Davidson is referring to how we are to understand, or account for, self-deception, the talk of boundaries seems to point in a different direction. Does the self-deceiver draw the boundary or does the concept 'boundary' come in as a conceptual aid to account for self-deception and how it is possible (in this case, performing a function analogous to the notion of 'causal relation')? In the paper "Who is fooled?", Davidson aims to clarify what he means by boundaries and semi-autonomous parts of the mind. He writes:

> I do not think of the boundaries, however permanent or temporary, as separating autonomous territories. The territories overlap: there is a central core of mostly ordinary truths which the territories share. […] The image I wished to invite was not, then, that of two minds each somehow able to act as an independent agent; the image is rather that of a single mind not wholly integrated; *a brain suffering from a perhaps temporary self-inflicted lobotomy.*
>
> This highly abstract account of the logical structure of self-deception is not, and never was, intended as a psychologically revealing explanation of the nature or etiology of self-deception. Its modest purpose was to remove, or at least mitigate, the features that at first make self-deception seem inconceivable. [98]

Is Davidson using the boundary merely as a "conceptual tool" in his explanation of self-deception in order to save the subject from holding a contradiction true? He understands "a mind not wholly integrated" as "a brain suffering from a perhaps self-induced lobotomy", which, together with the expression "drawing of the boundary", suggests that the boundary is actually drawn by the self-deceiver, who actively *does something* to keep the different beliefs apart. I believe that Davidson's use of notion of a 'boundary' is ambivalent. He uses it as a *theoretical notion*, on the one hand, and as a *descriptive notion*, on the other. Davidson's account requires that we understand it not as describing something that exists within the mind, and not only as a theoretical tool, but as an action performed by the self-deceiver. In this sense,

[98] Ibid. "Who is Fooled?", pp. 220. My italics.

the expressions "drawing the boundary" and "self-induced lobotomy" have metaphorical uses: they describe something that the self-deceiver does in order to hold these contradictory beliefs apart. Because it is a crucial step in Davidson's explanation, the drawing of the boundary must say something about self-deception itself. Were it merely a theoretical tool, it wouldn't solve the problem (that the self-deceiver seems to believe a contradiction). It cannot be accepted that the self-deceiver beliefs a contradiction; therefore, in order to do the work required, the boundary cannot come in only in accounting for self-deception. The drawing of the boundary (which can be read as a metaphor) must be a real move preformed by the self-deceiver in order to keep the incoherent and inconsistent beliefs apart.

Davidson says that his account was never intended as a psychologically revealing explanation of the nature of self-deception, but the drawing of the boundary plays a crucial role and, therefore, it is a problem that Davidson doesn't explain what the drawing the boundary consists in, or how it is possible. Davidson's talk of "boundaries", "self-inflicted lobotomy" etc. seems to have the character of a mythology, rather than an explanation. When turning to irrational action, Davidson's problem was the following: how is it possible for a rational person – a person whose mind is "a house of reason" – to act irrationally? The question can be reformulated thus: how can an intentional and rational system function in an illogical and irrational way? How can it harbor incoherence and inconsistency? Davidson's answer is that it cannot; rather, we should think in terms of different systems, albeit with some shared content. When there is incoherence between a person's beliefs, as there is in self-deception, according to Davidson, we must understand these beliefs as belonging to different systems, i.e. to different structures of supporting propositional attitudes.

I have claimed that Davidson's outline of self-deception is heavily influenced by the presuppositions that he makes in order to get the project of radical interpretation off the ground. But how does the division into different parts of the mind help to preserve the presuppositions necessary for radical interpretation? What role does the compartmentalization play? I will turn to Mark Johnston, with whom we are already acquainted, and to Lisa Bortolotti's paper "Intentionality without Rationality" in order to study these questions in greater depth. Johnston and Bortolotti are both critical of Davidson's rationalist rendition of irrationality.

Bortolotti cites a passage from Davidson's paper "Incoherence and Irrationality", in which he discusses the principles of rationality. "It is only by interpreting a creature as largely in accord with these principles that we can intelligibly attribute propositional attitudes to it, or that we can raise the question whether it is in some respect irrational."[99] Davidson refers to principles such as the principle of consistency and the principle of total evidence.

[99] Davidson, "Incoherence and Irrationality" in *Problems of Rationality*, p. 196.

According to Davidson, when an intentional system violates a basic standard of rationality, it departs from its own norms and has no good reason for doing so. Bortolotti summarizes Davidson's "solution" to the problem of irrationality thus:

> For Davidson, the only explanation of why an intentional system could suffer from, say, an inner inconsistency, lies in its compartmentalization in sub-systems. The conflicting beliefs are not found in the same sub-system, otherwise the inconsistency would be detected, but are in different ones. When the boundaries of the compartments break down, the inconsistency becomes explicit and there is no further excuse for the system that fails to recover from it. In general, if the system has the capacity to revise its beliefs as to restore conformity to the standards of rationality where these were violated, it does show that the breaches of rationality were not an open rejection of the standards themselves, but an effect of compartmentalization. The acceptance of the principle in question on behalf of the system is necessary for intentional description and is manifested if the system is disposed in the appropriate circumstances to conform to it.[100]

For Davidson, the mind, as rational, *cannot* harbor inconsistencies without detecting them. The division into sub-systems is thus a way to preserve the idea of a mental life that harbors no inconsistency and does not break with the principle of total evidence etc. *unless* something comes in which prevents irrationality and inconsistency from being detected. Compartmentalization is meant to guarantee that the standards of rationality are not openly rejected. Still, as we have seen, Davidson holds that it cannot be true that only ignorance can explain irrational action (the Plato Principle); rather, a part of the mind knows what the other part doesn't. The capacity to restore rationality plays a central role in Davidson's account; it is a condition of intentional systems that they can revise their beliefs so as to conform to norms of rationality. If the mind were one system, the beliefs would be revised so as to conform to the norms of rationality: they would form a coherent whole. Since self-deception involves inconsistency and incoherence between beliefs in Davidson's account, the only way out seems to be to speak of parts of the mind – each one a logically coherent whole – and boundaries between these parts. Compartmentalization thus makes it possible to retain the idea that intentional systems conform to norms of rationality by allowing for more than one intentional system within the mind. The suggestion is thus that breaches of rationality are something alien to the mind in that they violate the rational principles that govern the mental. In "Paradoxes of Irrationality", Davidson writes:

---

[100] Lisa Bortolotti, "Intentionality without Rationality", published in the *Proceedings of the Aristotelian Society*, vol. 105, 2005, p. 369-376 (Blackwell Publishing), p. 370. My italics. Bortolotti refers to a passage in "Deception and Division" in *Problems of Rationality*, p. 203.

> The idea is that if parts of the mind are to some degree independent, we can understand how they are able to harbour inconsistencies, and to interact on a causal level. Recall the analysis of akrasia. There I mentioned no partitioning of the mind because the analysis was at that point more descriptive than explanatory. But the way could be cleared for explanation if we were to suppose two semi-autonomous departments of the mind, one that finds a certain course of action to be, all things considered, best, and another that prompts another course of action. On each side, the side of sober judgement and the side of incontinent intent and action, there is a supporting structure of reasons, interlocking beliefs, expectations, assumptions, attitudes, and desires.[101]

By explaining an irrational action as stemming from beliefs that belong to different parts of mind, the requirement of holism is met, since the propositional attitudes within these structures are logically related: the different parts are coherent wholes of reason explanations for every action. This explanation can account for irrationality since it allows for conflict (i.e. breakdown of the logical relation) between these rational structures. Thus, someone's beliefs, wishes, intentions and actions can conflict with his other beliefs, wishes, intentions and actions. Like Bortolotti, Johnston is critical of Davidson's rationalist account of irrationality (Johnston discusses self-deception in particular) and his "homuncularist" solution, which assumes a division of the mind. In his paper "Self-Deception and the Nature of Mind", Johnston takes up the view, common among philosophers, that self-deception is paradoxical. What makes self-deception appear paradoxical, Johnston argues, is that it is portrayed as intentional. He writes:

> A dubious presupposition does lurk behind the familiar construal of the paradox. To be deceived is sometimes just to be *mislead* without being *intentionally* misled or lied to. The self-deceiver is a self-misleader. As a result of his own activity he gets into a state in which he is mislead, at least at the level of conscious belief. But the presupposition that generates the paradox is that this activity must be thought of as the intentional act of lying to oneself so that self-deception is just the reflexive case of lying. Evidently, *as the surface paradox shows*, nothing could be *that*. The homuncularist holds to the presupposition that the intentional act of lying is involved but drops the strict reflexive condition. If self-misleading is to be lying then the best one can do is to have parts of the self play the roles of liar and liar's victim.[102]

He argues that self-deception comes to appear paradoxical in accounts such as Davidson's due to a certain theory of interpretation. According to Johnston, its typical features are the following:

> there is nothing more to being in a mental state or undergoing a mental change than *being apt to have that state or charge attributed to one within an adequate interpretive theory*, i.e., a theory that takes one's behavior (including speech

[101] Davidson, "Paradoxes of Irrationality" in *Problems of Rationality*, p. 181.
[102] Johnston. p. 65.

> behavior) as evidence and develops under the holistic constraint of construing much of that behavior as intentional action caused by rationalizing beliefs and desires that it is reasonable to suppose that the subject has […].[103]

To deceive oneself is thus to *be interpreted* as deceiving oneself within a *theory* which portrays self-deception as an intentional action for which there are reasons. Davidson's account of self-deception has the structure it has because of the requirements of the interpretive view. This is my basic argument throughout this chapter in trying to show how Davidson tries to account for self-deception in a way which is in line with the principles guiding his account of radical interpretation. The quote from Johnston expresses more forcefully than I have so far that the interpreter's view, rather than the agent's, is authoritative in defining self-deception. All that matters to Davidson is that self-deception and other forms of irrationality *can conceivably* be mapped as intentional actions within an interpretive theory. He is not interested in understanding self-deception as such. In contrast, it can be helpful to recall Anscombe's discussion of intention here. According to Anscombe, if the answer to the question "Why?" is "I don't know why I did it", the person doesn't have a reason for what she did. Anscombe holds that "even if psychoanalysis persuades him to accept something as his reason", there is no reason, at least not in the relevant sense. Anscombe seems to be saying that there is no reason because, even if the agent accepts something as his reason, the agent cannot answer the question as to why he did it *by himself*. In Davidson's view, however, what matters in describing a behavior as an intentional act is only that it is *possible to provide a coherent explanation for it as an intentional act,* i.e. that it *can* be rationalized. In Johnston's words, it must be possible to "*construct* a plausible practical syllogism from beliefs and desires of the agent to the intention to carry out the act. This much follows from the assumption that the agent acted for and from a reason."[104] The question of whether or not it was actually the outcome of practical reasoning is irrelevant.

Self-deception poses a problem and a prima facie counterexample to Davidson's interpretive view of mental states and events since, as Johnston says, "they seem to be cases in which desire brings about belief in a characteristic way which on the face of it is not subject to rational explanation. The generated belief could not be the outcome of any practical syllogism with the desire as a premise."[105] This is reminiscent of Bortolotti's discussion of Davidson's concern with accounting for irrational phenomena in a way that does not contradict his assumption that intentional systems conform to the principles of rationality. It cannot be rationally justified that one believes that which one wishes to be the case; but Davidson attempts nonetheless to provide a rational explanation of self-deception in which he assumes that the self-deceiver believes that which the

[103] Ibid. p. 66. My italics.
[104] Ibid. p. 69.
[105] Ibid. p. 67.

evidence favors and that which he wishes to believe. Thus he holds two inconsistent beliefs, and Davidson is forced to assume a division of the mind.

Johnston questions the credibility of this picture in a number of ways. He questions the idea that a person's beliefs and desires must be coherent. Johnston holds that an important feature of self-deception is that there is a "characteristic, nonaccidental and nonrational connection between belief and desire"[106] and that it is a grave mistake not to recognize the important of this. In agreement with Bortolotti, he holds that "[r]ationality is not constitutive and exhaustive of the mental".[107] Johnston also questions the notion that the self-deceiver has evidence for believing the rational belief. He asks: how is it conceivable that the self-deceiver can believe something when he *knows* that the evidence points in the other direction? "What might seem odd is any conscious state that involves one's intentionally coming to believe some proposition *p* while recognizing that one neither presently possesses or will possess evidence for *p*, so that one has no evidential basis for thinking *p* true."[108] Johnston concludes that to the extent that the self-deceiver is to be understood as adopting the wishful belief *despite* his recognition at some level that the evidence is to the contrary, we have reason to regard the self-deceiver as divided, "[f]or it is hard to see how anxiety could be reduced by a wishful belief if the wishful belief is copresent in consciousness with the recognition that the evidence is strongly against it".[109] Thus it is a mistake to assume, as Davidson does, that the self-deceiver has evidence for that about which he goes on to mislead himself. This point knocks down the other leg on which Davidson's Principle of Charity stands: the assumption of correspondence between things in the world and a person's beliefs.

Johnston discusses how this view of self-deception invites the idea of separate parts of mind as *subagents*. Since recognition that evidence points to something which conflicts with what one wishes to believe must be avoided in self-deception, Johnston notes that some play must be given to the concept of *repression*: the subject must cease consciously acknowledging the evidence. Johnston further notes: "Where repressive strategies abound, it is plausible to postulate a repressive strategist. But the strategist cannot be the main system, in which the wishful belief allays anxiety. […] Consciousness of its reason for repression makes the main system's task of forgetting impossible. […] So we seem driven to recognize a subagency distinct from the main system."[110] Thus, in Davidson's account, a division of the mind is necessary in order to explain how the subject, who knows that the evidence favors that which she doesn't want to be the case, can avoid the anxiety that the belief provokes in Davidson's account. Further, it must also be assumed that the substructure is an agent who

[106] Ibid.
[107] Ibid. p. 90, note 9.
[108] Ibid. pp. 68.
[109] Ibid. p. 75.
[110] Ibid.

intentionally deceives the main system. Thus we end up with a perplexing picture of the mind as divided, each division being a (semi) separate agent. I will return to the role that the division of the mind plays in Davidson's account towards the end of this chapter. Here I will look closer at the subintentional account of self-deception that Johnston presents in his paper.

Johnston is not an interpretivist, and his account of self-deception differs from Davidsons. Most significantly, his account avoids the homuncular solution that is the natural continuation of the interpretive view. I present Johnston's view here because it points in the direction that I take in this book, that is, towards viewing self-deception as a flight from that which makes one anxious, rather than as an intentional action of misleading oneself about a belief which one knows that one has the best evidence for holding but which one doesn't want to believe.

Having argued that the paradox of self-deception can be seen as a parallel with the paradox of repression, since self-deception must be seen as including repression, Johnston says that instead of understanding repression as "an *intentional act of some subagency guided by its awareness of its desire to forget*",[111] we should understand it as "*subintentional, i.e., not guided by reasons* but *operating for the purpose of reducing anxiety.*"[112]

Johnston holds that the interpretive view fails to recognize what we really find problematic about self-deception because it portrays self-deception as a defect of *reasoning*. In Davidson's account, as we have seen, self-deception is a break with a rational principle to which the self-deceiver subscribes: the requirement of total evidence for inductive reasoning, which obliges one to hold the belief which one has best evidence for holding. But, argues Johnston, when I accuse someone else of self-deception or when I make self-accusations, I am not referring to mistakes of reasoning. Rather, I am disappointed with the other, or with myself, for not facing whatever it is that provokes the anxiety. What I am accusing the other (or myself) of is rather a defect of *character* than a (temporary) defect of reasoning. While Davidson opens his paper "Deception and Division" by removing self-deception from the moral realm, Johnston brings it back. Davidson asserts: "Self-deception is usually no great problem for its practitioner; on the contrary, it typically relieves a person of some of the burden of painful thoughts, the cause of which are beyond his or her control."[113] Johnston, on the other hand, holds the core of self-deception to be avoidance of acknowledging undesirable traits in one's own character. Accusations of self-deception, therefore, are accusations that the person in question withdraws from facing that which she does not want to admit. Johnston claims that the

---

[111] Ibid. p. 76.

[112] Ibid. Johnston refers to *mental tropisms* in the outline of his account. I will not discuss what he takes these mental tropisms to be any further than briefly referring to the way in which he describes them: as sub-intentional, nonaccidental, purpose-serving mental regularities which are the causal bases of rational and irrational connections alike. (Johnston, pp. 66)

[113] Davidson, "Deception and Division" in *Problems of Rationality*, p. 199.

subintentionalist account does a better job of accounting for this. "Though mental flight, like physical flight, is typically subintentional, one can still be held responsible for lacking the ability to contain one's anxiety and face the anxiety-provoking or the terrible. The accusation of self-deception is a familiar case of being held responsible for an episode that evidences a defect of character […]."[114] Johnston refers to a passage from Augustine's *Confessions* to characterize such a flight.

> You, O Lord, turned me back upon myself. You took me from my own back, where I had placed myself because I did not wish to look upon myself. You stood me face to face with myself, so I might see how foul I was, how deformed and defiled, how covered with stains and sores. I looked, and I was filled with horror but there was no place for me to flee from myself. If I tried to turn my gaze from myself, he still went on with the story he was telling, and once again you placed me in front of myself and thrust me before my own eyes, so that I might find out my iniquity and hate it. I knew what it was, but I pretended not to; I refused to look at it and put it out of my memory.[115]

We see how Augustine escapes his own gaze and self-scrutinization over and over again in order not to face that which he does not want to see. And, when acknowledgment is inescapable and he is "thrust before his own eyes", he tries to deny what he has found out. This is not a matter of misleading oneself about evidence one has. It is much better described in Johnston's words as a purposive and cowardly flight from "the terrible": "Here the self-directed accusation of self-deception is an accusation of mental cowardice, of flight from anxiety (or angst), a failure to contain one's anxiety, a lack of courage in matters epistemic."[116] Johnston notes that the homuncularist approach cannot account for this cowardly avoidance of getting to know oneself better: "The homuncularist picture of the self-deceiver prevents us from rationally reconstructing a fitting subject for this sort of censure. The protective system is simply lying. The main system is simply the victim of a paternalistic liar. This does not add up to anything like mental cowardice."[117]

The problems that Davidson ascribes to self-deception are really problems that arise from his own account of self-deception, not problem arising in a human life. If self-deception does not begin with holding a belief that one knows that one has the best evidence for believing, it doesn't involve the paradox of believing *p* and not-*p*, a paradox calling for a homuncular solution. No explanation in terms of mental causes is needed to account for the coexistence of the belief *p* and the belief not-*p* in the mind of someone who deceives himself. Neither need it be assumed that self-deception involves a

---

[114] Johnston, p. 85.
[115] St. Augustine, *Confessions*, transl. John K. Ryan, (Image Book, 1960) VIII, 7-16, p.193. Johnston discusses this quote on p. 85.
[116] Johnston, p. 85.
[117] Ibid.

process of practical reasoning in which the self-deceiver takes himself to have a reason to induce the belief he wishes to hold in spite of knowing that he is not justified in holding it. Finally, there are characteristic traits of self-deception for which Davidson fails to account; such as that self-deception is a fundamental moral term. It is about willing to know, not knowing *per se*.

In the last pages of this chapter, I want to return to Davidson and ask: how does his division into different rational parts of the mind help preserve what he takes to be the necessary conditions for radical interpretation? Davidson's explanation of irrational action as consisting of rational structures doesn't necessarily make it any easier for a radical interpreter to interpret a speaker whose thought is largely irrational. No doubt it would be much more difficult to match beliefs with meanings in the way required by radical interpretation if there were a great deal of inconsistency and incoherence in the speaker's thought and utterances than it would be to interpret a speaker who reasons rationally. It could be of help to the interpreter to know that, even if all attitudes weren't rationally connected, at least every attitude belonged to some coherent whole of attitudes. But Davidson's aim is not to develop a well-functioning practice of interpretation; instead his main task in explaining irrational action is to prevent the assumptions that he makes about persons, thinking and action in his account of radical interpretation from being refuted on the grounds that they cannot be generalized to account also for irrational thought. This is his motivation for interpreting self-deception in a way that is, on the whole, in line with the requirement of holism and rationality. As Davidson admits, in speaking of recognizing semi-independent departments within the same mind: "We attribute beliefs, purposes, motives, and desires to people in an endeavor to organize, explain, and predict their behavior, verbal und otherwise. We describe their intentions, their actions, and their feelings in the light of the most unified and intelligible scheme we can contrive".[118]

Davidson's motivation is then, quite explicitly, to "contrive" an intelligible "scheme" for the purposes of "organizing, explaining and predicting" human behavior; it is a theory, or model, for the generation and construction of hypotheses, speculation and systematization. His aim, then, is to save his model or picture of the mind from the threat posed by self-deception, not to capture the phenomenon itself. When Davidson remarks, almost in passing, that we can conceive of self-deception as a kind of "self-induced lobotomy", it were as if he were suggesting that, while it would please him to find out that this model could be supported and supplemented by empirical studies in psychology, this would be of only secondary importance. After all, his account "is not, and never was, intended as a psychologically revealing explanation." But one might ask if it is a "philosophically revealing one". If one conceives of philosophy as primarily concerned with the contrivance of models, then, of course, it is. If one sees the task of philosophy as trying to get clear about the matter at hand, in

[118] Davidson, "Paradoxes of Irrationality" in *Problems of Rationality*, p. 182.

this case, the phenomenon of self-deception, then Davidson's account of self-deception can hardly be seen to be an honest attempt.[119] Davidson's purpose is to mitigate the threat that the phenomenon of self-deception poses to his own view of human behavior as displaying large-scale rationality. Thus, to get clear bout the nature of self-deception is not his business.[120]

[119] In the paper "Självbedrägeriet i den analytiska filosofin" ("Self-Deception in Analytical Philosophy") published in the *Festschrift* to Professor Sören Stenlund I discuss the problems with Davidson's distanced, theoretical and disinterested perspective on self-deception. I am arguing that the idea of a purely theoretical interest in self-deception in problematic in itself; if self-deception is not regarded as *someone's* – that is not investigated from a perspective that someone may have – it will remain a purely theoretical construction. The argument was aroused by Stenlund's book *Det Osägbara* and by his paper "Ethics, Philosophy and Language". In *Det Osägbara*, Stenlund discusses the possibility of identifying absolute properties in a proposition. He writes that the questioner must pay attention to the linguistic context in which the proposition belongs. If one investigates a proposition without paying attention to the context in which it belongs, one views it *only* as an outsider and stranger, as a spectator, a theoretically interested onlooker, a "professional" philosopher. I argue, in the paper as well as here, that this is a problem in Davidson's account of self-deception. It is only by assuming a "stranger's" perspective on self-deception that he can ignore that self-deception has psychological and moral grounds as well as implications. This ignorance, I argue, does not only make Davidson's account poor in *some* respects but he leaves out what is fundamental to self-deception; the *experience* of self-deception. (Elinor Hållén, "Självbedrägeriet i den analytiska filosofin" in *Tankar tillägnade Sören Stenlund*, eds. Forsberg, Rider, Segerdahl, Uppsala Philosophical Studies 54, Västerås: Edita Västra Aros: 2008; Sören Stenlund, *Det Osägbara*, Stockholm: Norstedts förlag, 1980; Sören Stenlund, "Ethics, Philosophy and Language" in *Commonality and Particularity in Ethics*, Lilli Allanen, Sara Heinämaa och Thomas Wallgren, Ipswich: Macmillian Press Ltd, 1997.)

[120] We recall, "This highly abstract account of the logical structure of self-deception is not, and never was, intended as a psychologically revealing explanation of the nature or etiology of self-deception. Its modest purpose was to remove, or at least mitigate, the features that at first make self-deception seem inconceivable." (Davidson, "Who is fooled?", p. 220.)

Sebastian Gardner

Self-Deception in Relation to Psychoanalysis

Gardner's Project in Short

How should we understand self-deception in relation to Freudian psychoanalysis and psychoanalytic theory? In the first part of his book, *Irrationality and the Philosophy of Psychoanalysis*, Sebastian Gardner provides a thorough study of a wide span of irrational forms of self-understanding and their relation to psychoanalytic theory.[121] Gardner opens the book with the following reflection:

> The recognition that one's thoughts, behaviour and feelings run contrary to reason, and contradicts one's identity as a rational being, naturally provokes self-interrogation, along the following lines: What causes irrationality? Is it the result of choice, or the effect of a mechanism? To what extent does irrationality involve self-awareness? Should it be said that the irrational mind is divided? What – if anything – is the purpose of irrationality?[122]

In Gardner's view, psychoanalytic theory provides "the most penetrating and satisfying explanation of irrationality"[123] and his book is intended to provide philosophical argumentation for psychoanalytic theory, which he thinks can bring clarity to problems of irrationality. While the greater part of his book is devoted to psychoanalytic explanations of forms of irrationality, a central aim of Gardner's project is to find a ground in what he calls "ordinary psychology" for understanding the forms of irrationality that psychoanalysis treats. Gardner says of his account that "[i]ts target is the view that irrationality is off limits to ordinary psychology , or that ordinary psychology deals only with rationality."[124] This is how he summarizes the structure of his work:

> Chapter 1 attempts to see how far ordinary psychology may be pushed when faced with the problem of explaining irrationality, and takes self-deception as the prime candidate for explanation in such terms. Self-deception, as well as being intrincically interesting, provides an important term of contrast for the psychoanalytic irrational phenomena to be considered in Part Two.[125]

---

[121] Sebastian Gardner, *Irrationality and the Philosophy of Psychoanalysis* (Cambridge: Cambridge University Press, 1993). Gardner extends his study to include Melanie Klein and other psychoanalysts. I treat only the parts in which he discusses Freudian psychoanalysis.
[122] Ibid. p. 1. Later references to Gardner are to *Irrationality and the Philosophy of Psychoanalysis*.
[123] Ibid.
[124] Ibid. p. 16.
[125] Ibid. p. 5

Gardner says that the so-called ordinary forms of irrationality "are to be sharply distinguished from those with which psychoanalysis is concerned"[126], and, as we see in the quote above, self-deception is taken to be the "prime candidate" for description in terms of ordinary irrationality. The distinction between ordinary forms of irrationality and those treated by psychoanalysis will be explicated throughout this chapter, but Gardner's summary of the claims of the book can serve as a starting point. He writes:

> Freud's theories are directed primarily at the explanation of irrational phenomena which are *distinct in kind* from others that we are more familiar with, such as self-deception and akrasia. Psychoanalytic theory is therefore not, pace Sartre and others, a theory of self-deception, and does not stand in competition with ordinary, non-psychoanalytic explanations of ordinary forms of irrationality. [...] Sartre's criticisms indicate another constraint that psychoanalytic theory must observe: unconscious processes must not be assimilated to ordinary mental processes. Psychoanalytic theory attempts to go beyond ordinary psychological explanation in specific directions, and these hinge on the concepts of *wish-fulfillment* and *phantasy*.[127]

Gardner holds that self-deception, akrasia and other forms of ordinary irrationality should be understood as involving ordinary mental processes, while the irrational phenomena which Freud's theories are meant to explain involve unconscious processes, which must be understood differently, and require concepts belonging to psychoanalytic theory for their explication.[128] The irrational phenomena that Freud treats are *different in kind* from the irrational phenomena with which we are more familiar, such as self-deception, wishful thinking and akrasia, argues Gardner. Failure to appreciate this difference, like Sartre's (and Davidson's, whose account on irrationality Gardner also discusses[129]), leads to the mistake of reading a Second Mind structure into

---

[126] Ibid. By discussing self-deception and other forms of "ordinary irrationality", Gardner "aims to prepare for the introduction of psychoanalytic material, and to help set up the philosophical framework necessary for examining psychoanalytic concepts" (Gardner, p. 5)

[127] Ibid. 10. My italics.

[128] In the paper "The Unconscious", Gardner argues for a nonintentional understanding of the unconscious and that this makes it different from a "second consciousness", or the preconscious; while the latter is strategic the former is not. Gardner argues against Sartre's critique of the unconscious in *Being and Nothingness*: "What ultimately decides the issue between Freud and Sartre is how much *rationality*, or capacity for *strategic thought*, is invested by Freud in his account of the unconscious. If rationality, marked by the capacity to formulate *intentions*, is involved in unconscious thought, then the unconscious approximates to a proto-person, but if it is not involved, then it need not be so conceived. [ ] Although it might perhaps be argued on Sartre's behalf that there is reason for thinking that, in the case histories, Freud ought to have conceived of the unconscious as capable of manipulative intent, there is conclusive evidence that he did not aim to do so." ("The Unconscious", in *The Cambridge Companion to Freud*, ed. Jerome Neu, Cambridge: Cambridge University Press, 1991, pp. 152-53). The topic of Freud's intentional account of the unconscious will be discussed in Chapter Three.

[129] In Part I, Chapter 3 ("Persons and Sub-Systems") of *Irrationality and the Philosophy of Psychoanalysis* Gardner critically discusses Davidson's theory of sub-systems in accounting for irrational phenomena.

psychoanalytic theory, since, as we have seen in the previous chapter, the contradictory propositions in which the irrationality is thought to consist must have their own structure of justificatory propositions. Gardner grants Sartre's view that the outward appearance uncovered in psychoanalytic interpretation is that of self-deception, but he wants to show that psychoanalytic explanada and explanations are at a significant remove from self-deception. Gardner holds that ordinary psychology has the power to account for self-deception, akrasia etc., but that psychoanalytic concepts are not reducible to the terms of ordinary psychology.[130]

This chapter is a critical reflection on Gardner's book *Irrationality and the Philosophy of Psychoanalysis*. The focal point of my study will be Gardner's account of self-deception as a form of ordinary irrationality. While my project will involve discussing other forms of ordinary irrationality and irrationality treated by psychoanalytic theory, my aim in this discussion is to introduce and analyze Gardner's view of self-deception. I will focus on the introduction and first chapter of *Irrationality and the Philosophy of Psychoanalysis*.

Gardner lists wishful thinking, self-deception and akrasia as the main forms of ordinary irrationality, and he claims that we have no difficulty in saying, in broad terms, in what each of them consist. He writes:

> These forms of ordinary irrationality […] – those recognized in ordinary psychology – are *propositionally transparent*, by which it is meant that they are *constituted and defined by a particular structure of propositional attitudes*.[131]

He continues by specifying the structure of the three forms of ordinary irrationality:

> Wishful thinking is a matter of believing something simply because you desire it to be so, *self-deception consists in getting yourself to believe one thing in order to avoid facing what you know to be the truth.* Akrasia consists in failing to do what you know it to be best to do.[132]

As we saw in the last chapter, propositional attitudes are attitudes towards propositions. Many different attitudes can attach to one proposition. The proposition "I will pass the test", for example, can be believed, desired, doubted etc. Thus this proposition is included in propositional attitudes such as "I believe that I will pass the test", "I hope that I will pass the test", "I doubt that I will pass the test", and so on. Gardner argues that wishful thinking, self-deception and akrasia should be seen as constituted by propositional attitudes. Self-deception, as we will see, consists in holding incoherent beliefs. The

[130] Ibid.
[131] Ibid. p. 16
[132] Ibid. My italics.

structure that Gardner ascribes to self-deception resembles Davidson's definition.

Thus, self-deception, as a form of ordinary irrationality, can be identified by the particular structure of propositional attitudes in which it consists. Other characteristic traits of self-deception are, according to Gardner, *intentionality* and *practical rationality*. As we see in the definition of self-deception above, self-deception is directed towards a goal: to get oneself to believe something in order to avoid facing what one knows to be true. (I will soon move on to show how Gardner fleshes out self-deception as an intentional action involving reasoning and self-knowledge.) The forms of irrationality treated by psychoanalytic theory, on the other hand, do not consist of a structure of propositional attitudes; they are not intentional, nor do they involve reasoning and self-knowledge. Specific psychoanalytic vocabulary, such as the concepts of *wish-fulfilment* and *phantasy*, cannot be reduced to the language of ordinary psychological explanations. Gardner writes: "These are not to be thought of as sub-classes of non-psychoanalytic states – they are not species of desires or beliefs, or combinations of such – for their associated way of mental processing differs fundamentally from that of propositional attitudes."[133] *Wish-fulfilment* and *phantasy* do not include attitudes to propositions; further, while motivational, *phantasy* is not intentional.[134]

I am sympathetic to Gardner's project of showing how psychoanalytic theory contributes to our understanding of those experiences and phenomena that philosophers typically collect under the heading of "the irrational". But Gardner draws a radical distinction between, on the one hand, forms of irrationality to be accounted for by ordinary psychology, i.e. those which share the structure of normal rational reasoning and action in being intentional, consisting of a structure of propositional attitudes involving self-knowledge etc., and, on the other, forms of irrationality, which we can only grasp through psychoanalytic theory. I will not concern myself with this distinction as such. Rather, I will focus on what I find problematic in Gardner's portrayal of self-deception as a form of ordinary irrationality. We should recall that Gardner's aim in the chapter on ordinary irrationality is to "see how far ordinary psychology may be pushed when faced with the problem of explaining irrationality", and that he takes self-deception to be "the prime candidate for explanation in such terms". Self-deception is thus central in this distinction, which puts pressure on the concept to conform to Gardner's initial definition and description of ordinary irrationality. There are great similarities between Davidson's and Gardner's accounts both with regard to their basic assumptions and to certain problems arising from them. Thus, this chapter can be seen as a deeper discussion of issues raised in the last chapter. An important element in Gardner's account of self-deception, however, is its relation to psychoanalytic

[133] Ibid. p. 116.
[134] Ibid. p. 117.

theory and psychoanalysis. Since psychoanalytic theory and practice is the context in which I will investigate self-deception in the next chapter, this chapter serves as a bridge between Chapter One and Chapter Three.

Self-Deception

In developing his account on self-deception, Gardner starts by briefly contrasting self-deception with conditions which he thinks border on self-deception, but from which self-deception should be distinguished. Gardner initially describes self-deception as involving the idea of a person suffering a failure of self-knowledge because her motivation interferes with her beliefs. He recognizes that this description is broad enough to include cases of exaggerated credibility,[135] delusion,[136] self-distraction,[137] self-manipulation etc., all of which he wants to distinguish from self-deception. Among these, the condition that Gardner considers most difficult to separate from self-deception is the last one: self-manipulation. Gardner takes the case of Ulysses to be an example of self-manipulation. Ulysses ties himself to the mast so as not to succumb to the spell cast by the siren's song. Ulysses self-manipulation is a case of "reasonable irrationality", according to Gardner, which does not represent the same psychological phenomenon as self-deception. "Ulysses' intention is bounded in a way that the self-deceiver's is not: he plans to be impaired in limited respects for a specific length of time, solely with the view to achieving some determinate and fully realistic end, after which he means to return a state of full epistemic capacity."[138] Self-deception is different from self-manipulation, says Gardner, in that "it is not an intention to hold a false belief for a specific length of time, after which to return to a condition of true belief. Consequently, the state that self-deception aims to produce is not simply the instrument of an ultimately truth-respecting intention."[139] The difference can also be expressed in terms of control, holds Gardner; "whereas Ulysses tied to the mast remains logically in control of himself, the individual in self-deception is not similarly under the aegis or control of herself as a subject of self-deceptive intent."[140] Gardner understands self-deception as similar to self-manipulation, but with the

[135] "When José says of Carmen, 'I wonder whether that girl ever spoke one word of truth in her life; but whenever she spoke, I believed her – I couldn't help it', what he admits to is not self-deception, but a condition of exaggerated credibility." (Gardner, p. 17.)
[136] Gardner defines delusion by saying that it implies some impairment of the very faculty of belief formation; it is a fault at the level of competence rather than performance. (Gardner, p. 17.)
[137] "Self-distraction consists in a redirection of awareness, and the goal is directed at a change in experience; whereas the goal in self-deception is something closer to a change in beliefs." (Gardner, p. 17)
[138] Gardner. p. 18
[139] Ibid.
[140] Ibid.

difference that the self-deceptive intention is not truth respecting, and the self-deceiver is not in control of her intent to deceive herself.

Gardner introduces the kind of self-deception that he will discuss first through what is a more inclusive understanding of self-deception. He calls this "weak self-deception", to be contrasted with the kind of self-deception that will be the main focus in his account: "strong self-deception". According to Gardner, weak self-deception is "any structure of *motivated self-misrepresentation that does not involve an intention*"[141] Gardner defines it thus:

> a structure in which a psychological state S prevents the formation of another state S', where (i) S involves a misrepresentation of the subject, (ii) this feature is necessary for S to prevent the formation of S', and (iii) this structure answers to the subject's motivation.[142]

Gardner says that motivated self-misrepresentation is "the basic feature of all cases covered by ordinary use of the term 'self-deception', which distinguishes self-deception from wishful thinking and akrasia."[143] He holds that even animal behavior can manifest weak self-deception. The example Gardner gives is of an animal that deceives itself by putting itself in a state in which it misrepresents its own power and so avoids registering its own fear. Gardner says that this structure leaves open *how* it may operate. In order to separate specifically human self-deception from motivated self-misrepresentation of this kind, as well as from categories such as delusion and self-distraction, Gardner adds to the structure of motivated self-misrepresentation that it operates through an *intention* of the subject, and that it is directed towards manipulating *beliefs of psychological states* rather than psychological states directly. He adds to the form of weak self-deception that which he claims characterizes strong self-deception:

> Self-deception is a structure of motivated self-misrepresentation in which S and S' are *beliefs* and the process occurs through an *intention* of the subject's.[144]

This is strong self-deception, which is the form that Gardner goes on to discuss in his book. Gardner holds that it is specific to self-deception that it "*involves an ability to form a true belief combined with an attitude of truth-violation* (or at least truth-indifference)."[145] This makes self-deception different from delusion, for instance, for which it is typical that it disrupts the formation of a belief. Self-deception has a target, namely belief, and its means is an intention, directed at one's own beliefs.[146] In short, "a subject is self-deceived when he *believes one*

[141] Ibid. p. 19. My italics.
[142] Ibid. p. 18
[143] Ibid.
[144] Ibid. p. 19
[145] Ibid. My italics.
[146] Ibid.

thing in order not to believe another."[147] While belief is directly vulnerable to desire in wishful thinking and forms of weak self-deception, "strong self-deception manifests […] *practical rationality*, and exercising practical rationality requires an intention."[148]

To summarize, weak self-deception is to be understood as a form of ordinary irrationality that is constituted and defined by a particular structure of psychological states, i.e. fear and feelings of power, and involves motivation. In strong self-deception, which is the form Gardner discusses, it is not psychological states that are manipulated but *beliefs*: strong self-deception involves *beliefs* and an *intentional process of practical reasoning* through which the belief initially held true is replaced by a belief which the self-deceiver wishes to hold true. Gardner's "strong self-deception" closely resembles Davidson's "core cases of self-deception": it is conceived as intentional, as consisting of a structure of propositional attitudes, as involving a process of practical reasoning and, as we will see later, as involving self-knowledge. An important difference, however, is that Gardner does not posit a division of the mind in his explanation of self-deception. Gardner says of strong self-deception that it is most probably less common than wishful thinking, akrasia and weak self-deception, but that it is interesting because it invites the strategy of positing a division in the mind of the irrational person.[149] Gardner directs critique against this strategy in general, and against Davidson's account in particular, but he nonetheless finds strong self-deception interesting precisely *because* it invites this strategy. Gardner observes that Freud's theory of the unconscious has been viewed as strong self-deception, for instance by Jean-Paul Sartre in his critique of psychoanalytic theory – as if Freud's ambition was to offer a solution to the problem of understanding how a person can lie to himself. Gardner thinks that it is a mistake to see the phenomenon treated by psychoanalytic theory as comparable to strong self-deception. Rather, "as the 'most rational' irrationality, strong self-deception demonstrates, dramatically, one irrational possibility of the mind's — located, […] at the opposite end of the spectrum from those irrational phenomena to whose explanation psychoanalysis is directed."[150] Gardner acknowledges that some have denied the existence of strong self-deception, usually on the grounds that it has been thought that nothing can, logically, satisfy such a description. Gardner's mission is to show that strong self-deception is indeed possible and to elucidate what it means to attribute strong self-deception to someone.[151]

We cannot understand Gardner's claims about strong self-deception without considering how his theory is *used*, i.e. when applied to analyze a specific case. What is a case of this "most rational irrationality located at the opposite end of

[147] Ibid.
[148] Ibid. p. 21.
[149] Ibid. p. 31.
[150] Ibid. pp. 31.
[151] Ibid. p. 20.

the spectrum from those irrational phenomena to whose explanation psychoanalysis is directed"? Let us look at one of the examples of self-deception that Gardner discusses.

Anna Karenina's Self-Deception

Gardner discusses the character of Anna Karenina in Tolstoy's novel *Anna Karenina* as manifesting strong self-deception. This is the extract from the novel where, he claims, Anna's strong self-deception is revealed:

> At first Anna had avoided the Princess Tverskoy's set as much as she could, because it meant living beyond her means and also because she really preferred the other; but since her visit to Moscow all this was reversed. She avoided her serious-minded friends and went into high society. There she saw Vronsky and experienced a tremendous joy every time she met him. She met him most frequently at Betsy's, who had been a Vronsky herself and was his cousin. Vronsky went whenever there was a chance of meeting Anna and whenever he could speak to her of his love. She gave him no encouragement, but every time they met her heart quickened with the same feeling of animation that had seized her in the train the day she first saw him. She knew that at the sight of him joy lit up in her eyes and drew her lips into a smile, and she could not quench the expression of that joy.
>
> At first Anna sincerely believed she was displeased with him for daring to pursue her; but soon after the return from Moscow, having gone to a party where she expected to meet him but to which he did not come, she distinctly realized, by the disappointment that overcame her, that she had been deceiving herself and that his pursuit was not only not distasteful to her, but was the whole interest of her life.[152]

As a background, we should know that Anna Karenina is married to Karenin with whom she has a beloved son. Gardner's interpretation of this passage is that Anna Karenina is in conflict since she wants something that can only be pursued at a serious cost to herself; to be together with Vronsky. In confrontation with the fact of her marriage, self-deception functions advantageously: Anna Karenina gets to meet Vronsky without recognizing the risk of her falling in love. But it is not just Anna Karenina's good fortune that things turn out to her advantage. In an attempt to present the form which he holds strong self-deception to have, Gardner says: The self-deceiving subject, i.e. Anna Karenina, exercises preference over the reason on which she acts, or over her belief about these: she prefers to believe that she is Q-ing for reason R rather than R', because R' is discrepant with, whereas R accords with, how she wants to think of herself. Her irrationality consists in the fact that R, which is not the real reason for her action, only has the role it does because of R', with

[152] Extract from Leo Tolstoy, *Anna Karenina* (Harmondsworth, 1978), p. 143, quoted in Gardner, p. 19.

which it is in fact inconsistent. In light of her preferences, Anna Karenina forms an intention to determine her belief accordingly. In sum, in deceiving herself, she secures her goal through a process of rational reasoning.[153]

Gardner does not explicitly state what exactly corresponds to the action Q and the reasons R and R', but I will propose an interpretation. I will refer to the context in which the quoted passage from *Anna Karenina* occurs in order to better make sense of the reasons and actions which Gardner takes Anna Karenina's self-deception to consist in. First, Anna Karenina believes that she has started to socialize with Princess Tverskoy's set (is Q:ing) for the reason that the her former social circle "now became unbearable to her. It seemed to her that both she and all the others were pretending, and she felt so bored and uncomfortable."[154](R) But this is not the real reason for her change of company. The real reason is that she desires Vronsky (R'), and she attends functions where she is likely to see him. The reason R, that the social circle with the serious-minded friends seemed false and boring, only serves the role of letting Anna Karenina think of herself in a way that is not upsetting to her, as it would be to think that her actions revolve around her desire to see Vronsky and thus to recognize that she in love with him.

Another stage at which Anna Karenina finds a spurious reason for her action is when, having waited impatiently for Vronsky at a party, she accuses him upon his arrival of courting her in a very unsuitable manner (Q). She castigates him for "not having any honor", for courting her, and making her feel "as if she was guilty of something".[155] The reason for her action seems to be *Vronsky's* inappropriate behaviour (R). As becomes evident in this passage, Anna Karenina's accusations of Vronsky are really — or *also* — self-accusations. Following Gardner's structure, one could say that the real reason for her outbursts and accusations is her guilty conscience (R'). She projects her self-accusations on to Vronsky as defense against admitting to herself that the feelings that she has for him are feelings which she herself finds dishonorable, and which would be regarded as dishonorable by others as well. In Gardner's words, the reason that she took herself to have for her action was not the real reason.

These ways of taking something to be a reason for her action when it is not the cause of the action — the real reason, as Gardner says — could be understood as a reaction to something disquieting, or as a reflex to preserve her dignity. However, Gardner holds that the process of self-misrepresentation in strong self-deception is intentional and manifests practical rationality and self-knowledge. But how does the process of intentionally misleading oneself, which

[153] Gardner. p. 20. Paraphrase.
[154] Leo Tolstoy, *Anna Karenina* (London: Penguin Books Ltd, 2006), p. 127. When I refer to Tolstoy in later footnotes, I will refer to this book, published by Penguin Books Ltd.
[155] Tolstoy, p. 139.

is characteristic of strong self-deception, get started, and what keeps it going? Gardner explains:

> Let us call the psychological states S and S' which are involved in strong self-deception the *promoted* and *buried* beliefs respectively. The two key causal sequences involved in strong self-deception run (i) from a first-order desire to a second-order desire (Anna Karenina's desire for Vronsky causes her to desire not-to-believe that she ought to renounce him); and (ii) from a second-order desire to bury a belief to a second-order desire to promote a belief (Anna Karenina's desire not-to-believe that she ought to renounce Vronsky causes he to desire to believe that she is displeased with his pursuit of her.[156]

If I understand him correctly, Gardner takes Anna Karenina's self-deception to have the following structure: Since Anna Karenina desires Vronsky, she desires not to believe that she ought to renounce him. But the belief that she desires not-to-believe that she ought to renounce him (S) is a belief which she doesn't want to accept (lest it reveals the true nature of her feelings for him), so she invokes a belief in herself which is in line with what she wants to believe, a belief that is incoherent with the former belief: that she is displeased with his pursuit of her (S'). In other words, in order to avoid facing what she knows is true, that she desires not-to-believe that she ought to renounce Vronsky, she intentionally makes herself believe that she is displeased with him for courting her. In light of her preferences – to meet Vronsky without recognizing the danger of falling in love with him – Anna Karenina forms intentions to determine her beliefs accordingly. In deceiving herself, she secures her goal through a process of rational reasoning.[157] Gardner adds to the structure of practical reasoning: "These are relations of instrumentality: the second-order desire to bury is instrumentally related to the first-order desire, and the desire to promote is instrumentally related to the desire to bury."[158]

In order to understand the role that this structure plays in Gardner's account of strong self-deception, we can recall the role that practical reasoning plays in Davidson's account of Carlos' self-deception: it is by practical reasoning that Carlos goes from holding the belief that he will fail to holding the belief that he will pass, and the assumption in this practical reasoning — that it is better to avoid pain — makes this transition appear "reasonable". Similarly, in Gardner's explanation of Anna Karenina's self-deception, a belief is buried when another belief is promoted in the process of practical reasoning. An important difference between Gardner's account and Davidson's is that, for Gardner, the belief that the self-deceiver initially holds is not the same belief as that which the self-

---

[156] Gardner, p. 21. Recall how Gardner spelled out the structure of strong self-deception: "A structure in which a belief S prevents the formation of another belief S', where (i) S involves a misrepresentation of the subject, (ii) this feature is necessary for S to prevent the formation of S', and (iii) this process comes about because of an intention of the subject."
[157] Ibid. p. 20.
[158] Ibid. p. 21.

deception is intended to conceal. Anna Karenina does not initially hold the belief that she desires Vronsky and is about to fall in love with him. Anna Karenina's self-deception begins with the *desire* for Vronsky, and includes the belief that she desires not-to-believe that she ought to renounce Vronsky. This desire and this belief are buried in self-deception, in a process of practical reasoning. In justifying his intentional account of self-deception, Gardner argues:

> What establishes that an intention is required for these connections is that rationality is required to account for the derivation of the instrumental term in each. There is simply no other way of explicating the relations of instrumentality: sequences of the kind exemplified by (i) and (ii) are just what it means for something to occur through a person's intention. Any sparser story, that seeks to dispense with an intention in favour of wishful thinking and non-rational dispositions, will entail a conception of mental processing in which a desire can avail itself of the services of the right means miraculously – i.e. without need of reasoning to determine which are the right instruments for a desire to make use of. It follows that self-deceptive intent in required by the nature of practical reason. [159]

While belief is directly vulnerable to desire in wishful thinking and many cases of weak self-deception, strong self-deception cannot be accounted for in this way, according to Gardner, "for the reason that strong self-deception manifests, as we have seen, *practical rationality*, and exercising practical rationality require an intention."[160] We see that the understanding that strong self-deception manifests practical rationality plays a key role in Gardner's account; the presupposition that self-deception is instrumental burial and promotion of desires and beliefs makes it difficult to see it as anything but intentional. What can be questioned, and what I *will* question, is that self-deception is best understood in this way, i.e. as a process of reasoning. Again, we should recall that the assumption which Gardner's project of finding an explanation for some forms of irrationality within ordinary psychology hinges on taking strong self-deception as his prime candidate, is that self-deception must be explained in rational terms, as sharing the structure of normal rational reasoning.

Ordinary Irrationality vs. Irrationality Treated by Psychoanalysis

I will now look more closely at how Gardner distinguishes between ordinary irrationality, on the one hand, and, on the other hand, the forms of irrationality that, he argues, cannot be accounted for within ordinary psychology but are in

[159] Ibid.
[160] Ibid. p. 21.

need of psychoanalytic theory for their explication. I will begin with a few critical remarks regarding Gardner's account of self-deception, and, in particular, his way of delimiting the object of inquiry. We have seen Gardner argue that self-deceptive self-misrepresentation must come about through an intention on the part of the subject. Gardner claims that this is a distinctive difference between ordinary forms of irrationality and those forms treated by psychoanalysis: neurotic structures of motivated self-misrepresentation do not come about through an intention of the subject, and hence do not qualify as self-deceptive. Rather, says Gardner: "Neurotic symptoms are structures of motivated self-misrepresentation that pervert the ways in which the world appears to the person, and in which they appear to themselves."[161] Gardner refers to Freud here, saying "Freud speaks of neurotics 'need for uncertainty in their lives', which draws them 'away from reality': 'it is only too obvious what efforts are made by the patients themselves in order to be able to avoid certainty and remain in doubt'."[162] Strong self-deception on the other hand, according to Gardner, "necessarily involves rationality, intention and self-knowledge."[163] Gardner says that although strong self-deception, like obsessional neurosis, falls under the description "irrational", it "demonstrates one irrational possibility of the mind which is located at the opposite end of the spectrum from those irrational phenomena to whose explanation psychoanalysis is directed."[164] The difference between neurotic behavior and strong self-deception lays in this:

> If the Ratman's symptoms issue from self-deceptive intentions, there must be beliefs correlative with those intentions. Such as: [...] that obeying the Captain is a way of atoning for hating his father — and so on! Now, whereas the beliefs required for self-deceptive intent usually present no difficulty, these so called beliefs are very different: they do not belong to and can not be derived from, the stock of 'core' beliefs that everyone may be assumed to share. Since there is no way in which they could have been rationally formed, their explanation must be sought elsewhere.[165]

In the previous chapter's discussion of Davidson's analysis of the case of Mr. S and Mr. R (The Rat Man), I tried to show that it is not plausible to explain Mr. S' replacing of the branch or Mr. R's replacing of the stone *as rational actions for which there are justifications*; thus, I argued, it is misleading to describe neurotic behavior as having a rational, intentional structure. Although Mr. S takes his reason for returning to the park to be to replace the branch on the path — his

---

[161] Ibid. p. 94.
[162] Ibid. The three quotations are from Sigmund Freud, "Notes upon a Case of Obsessional Neurosis" ("The Rat Man"), p. 232 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, *SE* (London: The Hogarth Press and The Institute of Psycho-Analysis), vol. 10.
[163] Gardner, p. 95.
[164] Ibid. p. 32.
[165] Ibid. p. 95. Gardner writes 'Ratman'. In *The Complete Psychological Works of Sigmund Freud* it is written 'Rat Man' and I will keep to this.

action being intentional in the sense of having a direction[166] — he doesn't have a justification for his action. He might think that he is trying to prevent passers-by from getting hurt, this being his rationalization, but this is not what *causes* his action (alternatively, it is not the *real* reason for his action). His rationalization is not reasonable; it cannot justify his act. A rational reconstruction of such behavior does not capture the complexity and "un-reason" which it, in fact, reveals. As far as I can tell, Gardner and I agree that neurotic behavior ought not to be explained in terms of rational justifications, and we arrive at this conclusion on similar grounds. According to Gardner, there is a line of demarcation between neurotic behavior (and other irrational behavior belonging to the kind of irrationality addressed by psychoanalytic theory) and strong self-deception such that the latter, but not the former, should be understood as sharing the structure of ordinary reasoning. In what follows, I will examine the way Gardner spells out this distinction, which, I will argue, is highly problematic.

In the quotation above, Gardner writes that the fact that there are self-deceptive intentions implies that there must be correlative beliefs. Gardner says that although neurosis and self-deception both involve motivated self-misrepresentation, the former is non-intentional and the latter intentional: "neurotic structures of motivated self-misrepresentation surely do not come about through an intention of the subject's, and hence do not qualify as self-deceptive. If this is right the self-knowledge required for self-deceptive intent is absent from neurotic structures, which must consequently operate through non-intentional processes, without knowledge of their operation."[167] Gardner calls the beliefs correlated with the self-deceptive intentions "core beliefs". What are these core beliefs? Gardner speaks of "a stock of core beliefs which everybody can be assumed to share", and, what does not belong to the group of core beliefs is left aside since there is no way of telling how such beliefs can be rationally formed. "Core beliefs" thus seems to mean rationally formed beliefs. As we have seen, Gardner defines strong self-deception as consisting of buried and promoted propositional attitudes; these propositional attitudes seem to fit the description "core beliefs". Gardner calls self-deception, akrasia and wishful thinking "propositionally transparent" irrationality as distinct from those forms of irrationality treated by psychoanalysis, implying that these forms of ordinary irrationality can be identified by a certain structure of beliefs, desires and intentions. He writes:

> there is no logical gap between recognizing that the concept of self-deception has application, and knowing the kind of psychological state of affairs it consists in. Just as identifying a case as one of self-deception and knowing what sort of beliefs, desires and intentions it consists in are but one move, so there is no logical gap between making a judgement of self-deception and knowing an

[166] In, for example, a Husserlian sense: as directedness of experience toward things in the world.
[167] Gardner, p. 95.

> explanation of the phenomenon. Such proximity of description and explanation is a general characteristic of ordinary, propositional psychology. Hence the earlier description of self-deception as "propositionally transparent" irrationality.[168]

The cases of irrationality treated by psychoanalytic theory on the other hand, such as obsessional neurosis, are really deviations from rationality in their very structure. Gardner claims that these cases are "propositionally *opaque* rather than transparent".[169] They do not include intentional manipulation of beliefs but are non-propositional, and should be seen as including reactions and mechanisms (like repression) rather than carrying out intentions.

In my earlier discussion of Davidson's interpretative view, we saw that what matters most for Davidson is that irrational phenomena can be *accounted for* by a structure of propositional attitudes. For Gardner, by contrast, it is not enough to bring propositional attitudes into the explanation of self-deception; self-deception includes or consists of propositional attitudes: "The discussion of ordinary irrationality suggests an argument for realism about ordinary, propositional psychology. We have seen that the phrase, 'propositional explanation of self-deception' is misleading, if it is taken to suggest that self-deception is a non-propositional explanandum and that propositional attitudes are its explanans."[170] Deeper forms of irrationality, on the other hand, do not consist of and cannot, without confusion, be accounted for in terms of propositional attitudes; rather, concepts belonging to psychoanalytic theory, such as *wish-fulfilment* and *phantasy*, are needed to capture neurotic (and other) behavior and what motivates it.[171] The structure of propositional attitudes that constitutes self-deception, unlike neurosis, is "ordinary", according to Gardner:

> Explanation in terms of self-deception consists in attributing a sequence of practical reasoning, one in which beliefs play the same role as they do in paradigm explanations of rational actions. Self-deceptive intent does involve confusion, but this is a feature of the content of the propositional states comprising the sequence, and does not make it a different kind of sequence.[172]

In Gardner's view, the confusion in self-deception lies in the *content* of the propositional states and not in the sequence itself. In line with this assumption

---

[168] Ibid. pp. 28.
[169] Ibid. p. 91.
[170] Ibid. p. 31.
[171] Gardner does not accept the division between a conscious and an unconscious part of the mind, an explanation that hinges on the assumption of a Second Mind. He says that he wants to "adduce the existence of a second *kind* of mental state, instead of a Second *Mind*"(Gardner, p. 116) The psychoanalytical concepts *phantasy* and *wish-fulfilment* characterize this second kind of mental states. As I quoted earlier, Gardner says: "These are not to be thought of as sub-classes of non-psychoanalytic states – they are not species of desires or beliefs, or combinations of such – for their associated way of mental processing differ fundamentally from that of propositional attitudes" (Gardner, p. 116)
[172] Gardner, p. 30.

that strong self-deception consists of a process of practical reasoning, the confusion in self-deception cannot lie in the form but must lie in the incoherence between propositional states: the buried and promoted desires and beliefs. In Gardner's account, the self-deceiving subject, Anna Karenina, has desires such as the desire not-to-believe that she ought to renounce Vronsky *and* the desire to believe that she is displeased with his pursuit of her. This seems to be the kind of confusion in the content of the propositional attitudes to which Gardner refers. The contents of the propositional states that make up the sequence are incoherent, but the sequence, as Gardner portrays it, is a rational sequence of practical reasoning involving propositional attitudes. In Anna Karenina's case, to get to meet Vronsky without recognizing the danger of falling in love so that they may continue to meet is the goal towards which Anna's self-deceptive reasoning is directed. Thus, Gardner holds that the "kind of sequence" that goes into self-deception is just the same as in "paradigm explanations of rational actions". This is what makes strong self-deception a paradigm case of ordinary irrationality.

Having treated Gardner's delimitation of what characterizes ordinary irrationality, and self-deception as one of its expressions, I will turn to the contrary case of irrationality treated by psychoanalysis. A distinction that Gardner makes between ordinary forms of irrationality and cases like the Rat Man's is that the Rat Man's irrational actions: "clearly do not exhibit the immediate intelligibility, and unity of identification and explanation, which we saw characterizes ordinary irrationality. The Rat Man's is a different and deeper grade of irrationality, which is propositionally *opaque*, rather than transparent."[173] Ordinary irrational phenomena can be identified and explained because, Gardner claims, self-deception, wishful thinking and akrasia can be "viewed as constituted by, and explained in terms of, configurations of propositional attitudes."[174]

What makes the Rat Man's irrationality opaque in relation to the self-deceiver's? I will repeat Gardner's description of the Rat Man's irrationality, because it captures rather well salient features of the latter's specific form of obsessional neurosis. The Rat Man's irrationality consists in his exhibiting symptoms such as:

> compulsive impulses (to cut his throat, and to undertake a near-fatal diet); groundless fears that terrible things will happen to the people he loves, and corresponding obsessive desires to protect them (he removes a stone from the road so that it will not bring harm to his beloved, who will later be passing in a carriage); chronic indecision (over his choice of marriage partners); absurd, ill-conceived projects (he undertakes a train journey in order to repay a trivial debt, knowing it to be erroneous, and he suffers a mental breakdown en route), and

[173] Ibid. p. 91
[174] Ibid.

> barely intelligible, violent and emotionally overwhelming trains of thought, that he finds foreign and repugnant, on themes of death and torture.[175]

Gardner emphasizes the importance of seeing that these phenomena present themselves as different in kind from the ordinary constituents of mental life. The latter, says Gardner, "strain the Rat Man's ordinary way of viewing himself, and create in him a corresponding need for self-explanation."[176] Gardner calls the Rat Man's symptoms "contra-rational", and adds that, although the Rat Man does not understand them, they should not, nevertheless, be understood as completely arbitrary psychological events that just 'befall' him. They manifest mental states that are *his*, but whose nature and content he is unable to grasp: "This puts him in the contradictory situation of knowing that his symptoms manifest mental states which are his own, but of not knowing what these states are. That he is unable to 'read' his own mind produces, in a more extreme form than ordinary irrationality, the self-contradiction constitutive of irrational phenomena."[177] Gardner adds that the Rat Man is "crying out to be made intelligible to himself, at whatever conceptual price is compatible with his continuing to view himself in intentional terms, that is, as a person."[178]

I find Gardner's description of the Rat Man as an obsessional neurotic adequate and helpful. I will leave this example without much discussion since it only serves as a contrast to self-deception in this investigation, but I want to emphasize that, although I object to some of the ways in which Gardner distinguishes self-deception from obsessional neurosis, and to the cardinal traits that he ascribes to self-deception, it doesn't mean that I don't take obsessional neurosis to be different from self-deception. I will mention some of the ways in which neurosis is distinctive from self-deception. As we see in Gardner's summary of the Rat Man's problems, he acts out compulsions. Furthermore, sometimes these compulsions are life-threatening. He gets stuck in repetitive behavior (such as removing and replacing the stone) because the battle of instincts within him "forces" him to do one thing and then the other. The Rat Man has intricate lines of thought about why it is of uttermost importance that he does a certain thing and exactly how this action must be carried out, although it ought be clear to him that he doesn't have a reason for performing the action, and, even if he had, it would not serve any purpose to perform it in minute detail. The Rat Man's actions are not sensible. Self-deception doesn't share these characteristics. It is not a compulsive action, it does not include the repetitive behavior that is typical of obsessional neurosis and it doesn't consist of intricately worked out — but crazy — plans for how to obtain something.

---

[175] Ibid. p. 90.
[176] Ibid.
[177] Ibid.
[178] Ibid. pp. 90.

Ambivalence and Preference

We have seen Gardner hold that self-deception, as a form of ordinary irrationality distinct from the irrationality treated by psychoanalysis, shares the rational, propositional and intentional form of ordinary rational reasoning. Another way in which Gardner distinguishes self-deception from the forms of irrationality with which psychoanalysis is concerned is by discussing the concepts of *ambivalence* and *preference*. Gardner claims that while a typical characteristic of self-deception is ambivalence between different attitudes, obsessional neurosis does not involve ambivalence. In his analysis of what distinguishes the Rat Man's hatred from ordinary hatred, Gardner discusses the moment in therapy when the Rat Man confronts these feelings of love and hate for his father. The Rat Man asks Freud how he *can* hate his father given that this is inconsistent with everything that he knows about his feelings towards his father, his father being the one he loves most in the world. Freud replies that it was precisely such intense love as his that was the necessary precondition of the repressed hatred. Gardner goes on to say:

> Freud contrasts the combination of love and hatred found in the Ratman with ordinary emotional ambivalence. […] The Ratman's hatred could only have become available for self-knowledge if it had been *of a kind* that would have allowed for combination with love in a composite attitude, one of ambivalence towards his father. Had it been of such a kind, it would not have become unconscious, and the outcome would not have been pathological – *it would have been something closer to the self-deception which characteristically accompanies ambivalence*, *where one of the contrary attitudes is simply buried* (my italics). But because the Ratman's hatred was not of such a kind, it had to be removed from the sphere of self-knowledge and control necessary to yield a non-pathological solution.[179]

According to Gardner (and Davidson), self-deception consists of two incoherent beliefs. The self-deceiver knows the one to be true but he misleads himself by way of rational reasoning to develop a contrary belief. Understood thus, self-deception originates as ambivalence between beliefs and as a transparent choice between two beliefs to hold. That is how Gardner characterizes self-deception: as "getting yourself to believe one thing in order to avoid facing what you know to be true". The Rat Man's hatred, on the other hand, as Gardner says, does not allow for combination with love in a composite attitude, and therefore it cannot become available for self-knowledge. I take Gardner to mean that while Anna Karenina is aware of both the belief that she desires not-to-believe that she ought to renounce Vronsky and the belief that she is displeased with him for pursuing her, and the burial of the first is performed through an intentional process, the Rat Man, on the other hand,

[179] Ibid. p. 96. My italics. Gardner discusses "Notes upon a Case of Obsessional Neurosis" ("The Rat Man"), SE, vol. 10, pp. 180.

cannot apprehend the animosity towards his father until he is confronted with it in analysis; his mind is not transparent to himself as the self-deceiver's is.

Having distinguished between self-deception and obsessional neurosis by arguing that self-deception characteristically includes *ambivalence*, Gardner goes on to argue that the concept of *preference* provides the key to what makes the Rat Man's hatred different from self-deception. In determining if something is a preference or not, the rule of thumb to follow is this, according to Gardner: "the more desires are satisfied through behavior, and the more we can imagine performing operations of thought that make the selected course of behavior seem attractive, the more it can be supposed to manifest a preference. Self-deception can be seen in such a light: Anna Karenina gains time to allow the mutual interest she shares with Vronsky to grow."[180] According to Gardner, the Rat Man does not exercise a preference for neurosis over awareness of conflict. Gardner relies on a quote from Freud to bring out the contrast between "the ordinary" and neurosis, Gardner says: "Freud is explicit that the 'chronic conflict' to which the Ratman is subject is different in kind from ordinary conflict: 'the pathological conflict in neurosis is not to be confused with a normal struggle between mental impulses both of which are on the same psychological footing.'"[181] Gardner continues: "As with the Ratman's hatred, we need to know what makes the difference from the ordinary. My suggestion is that the concept of preference provides the key. The Rat Man shares with the self-deceiver the motive for aversion to conflict, but the resemblance is limited in this crucial respect: the Ratman does not exercise a preference for neurosis over awareness of conflict."[182] Gardner holds that the self-deceiver chooses self-deception as a rational means towards a goal: Anna Karenina's self-deception is intentionally directed towards the goal of gaining time to allow the mutual interest she shares with Vronsky to grow, and the intention is carried out through a process of practical reasoning. Self-deception, holds Gardner, is an attractive means since it allows her desire to be satisfied and, therefore, can be seen as manifesting a preference.

It is worth spelling out what something has to be to count as someone's preference according to Gardner, and thus why he chooses to explain self-deception, but not neurosis, in terms of preference. Gardner discusses the possibility that the symptom formation that characterizes neurosis could be explained by preference. It could look like this: "neurosis is an alternative to madness or even greater suffering; so it may be explained by a preference for neurosis over madness or greater suffering."[183] But, Gardner continues: "the description of X, an actual state of affairs, as *preferable* to Y, a counterfactual state of affairs, does not show X to *have been preferred* to Y. If someone is

[180] Gardner, p. 99.
[181] Ibid. p. 98. The quote is from Freud's *Introductory Lectures on Psychoanalysis*, SE, vol. 16, p. 433.
[182] Ibid. p. 98.
[183] Ibid. p. 100.

credited with exercising a preference, they must have known how to execute it."[184] It seems we can spell out what Gardner is saying in terms of subjective and objective explanation. Though in imagining the choice between madness and forming symptoms, we can see forming symptoms as preferable, that doesn't mean that it is justified to ascribe this preference to the subject. According to Gardner, the critical point is this:

> The reason why neurosis can not be explained by preference is not, then, that this would have to be unconscious, for people do indeed exercise non-conscious preferences (e.g. to self-deceive). *It is because they cannot be credited with the appropriate* beliefs ("By forming symptoms, madness can be avoided"), *without overstraining the concept of belief by removing from it all connection with justification*; and because, if beliefs and desires sufficient to rationalize neurosis are attributed ("Let me form symptoms, lest I go mad"), it becomes incomprehensible that the rationality which is the conceptual concomitant of belief and desire does not lead the neurotic to exercise a better preference.[185]

The problem of imagining that neurosis is the result of exercising a preference is not that it seems difficult to imagine how anyone could intentionally induce symptoms in himself. (Gardner does not consider this.) Neither is it that the preference for neurosis would have to be a non-conscious preference, since Gardner allows for non-conscious preferences and holds that they are involved in self-deception. The problem is that it cannot be imagined of a being capable of executing preferences – a rational being – that he wouldn't execute a *better* preference.

I will try to flesh this out. If forming a symptom would be a preference, the neurotic must be credited with a belief like "[b]y forming symptoms, madness can be avoided". A belief must be justifiable and rational, and the Rat Man's purported belief does not live up to these requirements, according to Gardner. To paraphrase a quote I cited earlier, "it does not belong to and cannot be derived from, the stock of core beliefs that everyone can be assumed to share". I agree with Gardner that it is not appropriate to characterize the Rat Man's obsessive behavior as manifesting intention, preference and belief, but I also doubt that self-deception should be understood in such terms. Gardner argues that Anna Karenina can be credited with the belief "[b]y deceiving myself I gain time to allow Vronsky's and my interest to grow."[186] But it is not at all clear to me what makes the belief that goes into this preference justifiable while the purported neurotic belief is not. Taking into account what Gardner has said before, he must hold that the belief in self-deception but not the purported

---

[184] Ibid. pp. 100.
[185] Ibid. p. 101. My italics.
[186] "the more desires are satisfied through behavior, and the more we can imagine performing operations of thought that make the selected course of behavior seem attractive, the more it can be supposed to manifest a preference. Self-deception can be seen in such a light: Anna Karenina gains time to allow the mutual interest she shares with Vronsky to grow." (Gardner. p. 99)

belief in neurosis belongs to the "core beliefs that everybody can be assumed to share". Gardner holds that "[b]y forming symptoms, madness can be avoided" cannot be a belief because it cannot be justified: "it does not belong to and cannot be derived from, the stock of core beliefs that everyone can be assumed to share." To begin with, Gardner has not made clear what these "core beliefs" are. He does suggest that core beliefs are beliefs that have been rationally formed, but it is not clear what that means. The Rat Man's purported belief has the same intentional means-goal structure as the belief that Gardner ascribes to Anna Karenina but, Gardner holds, while this is the right structure in which to explain self-deception, the Rat Man's symptom-formation does not have this form.

It is difficult to see what justifies Gardner's distinction between neurosis and self-deception in terms of preference. It is evident that neurosis is not rational in the sense that it is a solution to a problem or a good response to a situation. The question is: is self-deception essentially different in this respect? I would say no. This is why they are both described as irrational. What is it that makes self-deception a case of ordinary irrationality to be distinguished from neurosis? What makes the purported belief "by deceiving myself I gain time to allow Vronsky's and my interest to grow" a more rational belief than the Rat Man's? Why doesn't the objection arise that self-deception cannot be a preference because it is incomprehensible that the self-deceiver doesn't exercise a better preference? Both self-deception and neurotic behaviour stand in the way of self-understanding and thus of responding responsibly and well to a situation. And as we see in Gardner's characterization, both the Rat Man case and the case of Anna Karenina *can* be characterized in terms of preference and intention. The Rat Man's situation can be understood as a case of seeing himself confronted with Scylla and Charybdis and choosing Scylla: he busies himself with repetative behavior so as to keep his mind on the routine and avoid his fears and anxieties to invade his mind. I am not suggesting that this is a *good* way in which to understand symptom-formation. I am merely suggesting that Gardner's clear-cut distinction between self-deception and neursis according to which the former, but not the latter, is characterized by preference, is doubtful.

Gardner argues that Anna Karenina's self-deception should be understood as intentional and as manifesting preference while the Rat Man's neurotic behavior should not. I argue, on the other hand, that this is not a suitable characterization of either the Rat Man's neurotic behavior or of Anna Karenina's self-deception. The same objection that is raised against ascribing a preference to the neurotic can also be raised against ascribing preference to the self-deceiver. Thus, the question is open whether or not self-deception should be seen as manifesting a preference. Gardner fails to show that the belief which he ascribes to the self-deceiver as expressing her self-deceptive intention is justifiable. Therefore, given his own definition of belief as something that must be justifiable, he fails to show that self-deception manifests belief and can be said to be the execution of a preference.

As we saw, Gardner relies on a quote from Freud's *Introductory Lectures* to bring out the contrast between ordinary forms of irrationality and neurosis in terms of preference: "Freud is explicit that the 'chronic conflict' to which the Ratman is subject is different in kind from ordinary conflict: 'the pathological conflict in neurosis is not to be confused with a normal struggle between mental impulses both of which are on the same psychological footing'".[187] Gardner continues by saying that preference provides the key: the Rat Man does not exercise a preference for neurosis over awareness of conflict. In arguing that self-deception manifests preference, Gardner portrays self-deception as a rational conflict in opposition to the conflict that goes into neurosis. I will now argue that Gardner is mistaken in finding support for the distinction that he makes in this quote; nothing in Freud's discussion supports Gardner's comparison of neurosis with *self-deception*. The cases of "normal struggle between mental impulses" that Freud discusses are not cases of self-deception.

The context in which Freud's remark arises is a passage in one of his introductory lectures, where he remarks upon the rumor that psychoanalysis encourages patient's to "live a full life", that is, to be promiscuous. The discussion moves on to the analyst's role, and to what extent he can or should take sides in a conflict. Freud says: "if an abstinent young man decides in favour of illicit sexual intercourse or if an unsatisfied wife seeks relief with another man, they have not as a rule waited for permission from a doctor or even from their analyst."[188] These are the examples to which Freud is referring when he speaks of the "normal struggle with mental impulses". It is clear that Freud characterizes the case of the abstinent man as one where a decision must be made: to have sex rather than abstain from having it. One might call this "exercising a preference". But nothing suggests that these cases exemplify self-deception. In Freud's description, the man seems to know exactly what he has decided to do and he seems to be aware of having made a decision. I cannot see how this reference to Freud can provide any justification for what Gardner sees as the key to what makes the Rat Man's hatred different from self-deception, since the examples Freud gives of "normal struggles between mental impulses both of which are on the same psychological footing" would seem to have nothing to do with self-deception. In Gardner's portrayal, the exercising of a preference in self-deception is *non-conscious*. What Gardner's interpretation shows, however, is how close his picture of self-deception as a form of ordinary irrationality comes to viewing it as a normal, conscious act of choosing where the options lie open to view. When one is in a conflict that one acknowledges, and then makes a choice, the two mental impulses are on the same psychological footing in the sense that both are conscious. This, I suggest, is not the case in self-deception, and this is why it is problematic to view self-deception as a choice between beliefs to hold, or as a preference. My point is

[187] See footnote 181, p. 82.
[188] Sigmund Freud, *Introductory Lectures on Psycho-Analysis*, SE, vol. 16, p. 433.

that, to the extent that it makes sense to speak of preference in self-deception, it is not *like* conscious preference. Gardner's choice of words in calling it "non-conscious" instead of "unconscious" suggests that he conceives of the preference in self-deception as the same as (conscious) preference with the exception that one is not aware of it.[189] I have argued that Gardner interpretation of self-deception as exercising of non-conscious preference does not have support in the quote from Freud that Gardner refers to.

Conflict and Awareness of Conflict

In discussing preference as a characteristic that distinguishes self-deception from neurosis, Gardner also describes traits that they have in common. Gardner describes both self-deception and neurosis as subjective conflicts. He says that an objective conflict is when a person's desires may conflict solely by virtue of their joint content being impossible to satisfy, whether or not the person knows this. A subjective conflict of desires, by contrast, involves appreciation of the objective conflict.[190] The difference between the neurotic and the self-deceiver, according to Gardner, is that, although the neurotic appreciates that there is a conflict, he doesn't know what it consists in; this is what makes it impossible to see him as exercising a preference. But it is questionable that the self-deceiver typically appreciates the conflict, i.e. that self-deception involves a subjective conflict in the way Gardner describes it. In fact, here I think that there *is*, in general, a difference between neurosis and self-deception. As Gardner writes, neurotics, for the most part, are aware of a conflict (or at least aware of that something bothers them), but they need the help to understand what this conflict consists in. This is why many choose to seek therapy. The self-deceiver, on the other hand, doesn't seek clarity, I suggest, not because she prefers to deceive herself (without being aware of her preference), but because she is not fully aware of the conflict, and, indeed, evades becoming aware of it. This evasion, however, is not a means towards a desired end, as Gardener suggests, but rather an end in itself. In self-deception, evasion is its own reward. I am proposing that self-deception be understood as an evasive reaction, i.e., a way of averting confrontation with, or reflection upon, something which bothers the self-deceiver; by deceiving herself she can be unaware both of feeling troubled and of what it is that troubles her, that is, unaware of the conflict and of what it consists in.

Gardner claims that Anna Karenina appreciates the conflict when she deceives herself. Does the extract from Tolstoy's *Anna Karenina* suggest this? We must pay attention to what Tolstoy ascribes to Anna Karenina regarding

[189] In psychoanalytic terms: as descriptively unconscious or preconscious rather than as dynamically unconscious.
[190] Gardner, pp. 98.

her thoughts, beliefs etc. and what he does not ascribe to her. She knows that seeing Vronsky makes her happy: “She knew that at the sight of him joy lit up in her eyes and drew her lips into a smile, and she could not quench the expression of that joy.” She is also aware of Vronsky’s passionate feelings for her, and she reacts with what she thinks is displeasure when he courts her. All this strongly suggests that there *is* a conflict in her attitudes. It does not mean that she is *aware* of a conflict between her own desires. She could experience herself as being very happy in Vronsky’s company and still dislike the kind of affection he forces upon her. I want to suggest that the passage, “soon after the return to Moscow, having gone to a party where she expected to meet him but to which he did not come, she distinctly realized, by the disappointment that overcame her, that she had been deceiving herself and that his pursuit was not only not distasteful to her, but the whole interest of her life” suggests that Anna had never realized the conflict between her desires before, since she had never realized what kind of feelings it was that she had for Vronsky. This dramatic realization suggests that she realized something that was manifested in her reactions to Vronsky and her behavior before, but which she hadn’t appreciated earlier.

Burial of Belief vs. Repression

As we have seen, Gardner distinguishes between strong self-deception and forms of irrationality treated by psychoanalysis by holding that strong self-deception is intentional, and that it includes practical reasoning and propositional attitudes, ambivalence between beliefs, preference and awareness of conflict. Another distinction that captures difference between self-deception and forms of irrationality treated by psychoanalysis is the distinction that Gardner makes between burial of belief and repression. To characterize strong self-deception as intentional and propositionally transparent implies that the way in which the self-deceiver hides some mental content to herself is not by repression but by "burial of belief", as Gardner calls it. Gardner says: "it would be a mistake to identify the burial of belief in self-deception with repression."[191] Gardner suggests that the burial of belief that takes place in strong self-deception can be understood as suppression, which he describes as “the commonsensical fact that persons are able, by redirecting their attention, to make themselves unaware of what they do not wish to be aware of.”[192] Repression, on the other hand, says Gardner, is the “*inability to come to a* realization, i.e. to form a self-ascribing belief that is effective in correcting their behaviour.”[193]

[191] Ibid. p. 196.
[192] Ibid. p. 102.
[193] Ibid. p. 103.

According to Gardner, in repression "the person is unable to form appropriate awareness of that thought, and that this is not due to self-deception etc., but comes about because *the thought itself can not be manifested in consciousness.*"[194] The Rat Man's fierce idealization of his father at the beginning of his treatment, for instance, makes it impossible for any antagonistic feelings towards his father to manifest themselves as beliefs. Repression affects the very *formation* of beliefs and not simply awareness of the beliefs that one already holds, as burial of belief does. Gardner describes how the avoidance in repression points to what evokes it: "The starting point for an attribution of repression is a situation in which a person's behavior manifests differential sensitivity to some object, and bears such marks as anxiety, flight or aggression. Such behavior *discloses a thought*, as, for example, of a particular person as hostile."[195] The Rat Man's inability to form the belief that he perceives his father as hostile leads us in the direction of a mechanism: "Repression is the mechanism which makes it impossible for the Ratman, at that stage, to co-realize his love and hatred."[196] Repression makes mental states inaccessible without imputing an intention and thus can be more accurately described as a *mechanism*. This gives repression a non-rational character: "no benefits accrue to the subject other than relief from anxiety; unlike the self-deceiver, nothing in his behavior shows that he draws on knowledge of the repressed."[197]

We see that repression, unlike self-deception, is mechanical and non-intentional: it is not directed at obtaining benefits but is simply an aversive reaction to anxiety. It doesn't include practical reasoning, but it is, to the contrary, non-rational. Moreover, it sets in before a belief (propositional attitude) has been formed. In his analysis of the example of Anna Karenina, Gardner argued that this example, as an example of self-deception, is different from cases of the forms of irrationality accounted for by psychoanalytic theory. In self-deception, the person makes herself unaware of what she knows but doesn't want to recognize through a re-direction of attention: she buries a belief. This is something she does intentionally and *in order to obtain a benefit other than simply avoiding anxiety*. The desire to meet Vronsky so that their mutual interest can grow is promoted through a process of practical reasoning in which the burial and promotion of desires and beliefs play an instrumental role: they help her remain unaware of her desire for Vronsky and her wish to be with him. By keeping herself unaware, she avoids her own self-reproaches and can promote her aim to be with Vronsky. Gardner holds that a person deceives herself in order to obtain something which she desires and which she could not obtain were it not for the self-deception. The characterization of what takes place in self-deception as the burial of beliefs hinges on the assumption that

[194] Ibid. My italics.
[195] Ibid.
[196] Ibid.
[197] Ibid.

self-deception is directed at obtaining benefits other than relief from anxiety, i.e. that self-deception is directed at achieving something, just as a rational, intentional action is. I disagree with Gardner's view of self-deception as strategic and directed at obtaining benefits (other than avoiding anxiety). In the next chapter I will delve more deeply into this theme. My aim is to show that the evasion associated with self-deception is more akin to repression, that is, an avoidance of self-conscious conceptualization and articulation, than an instrumental suppression of formed beliefs.

Self-Knowledge and Realization

Self-deception is "ordinary" not only in the sense that it is intentional and rational, but also in that it includes self-knowledge, according to Gardner. He says that "[s]elf-deception involves a plan as far as all intention does",[198] and that the self-deceiver, unlike the neurotic in repression, draws on knowledge of what she buries.[199] Gardner summarizes the distinction neatly by asserting that, "in so far as a person is self-deceived, she knows what she is up to, but in so far as she is neurotic, she does not."[200] Armed with so much knowledge, strategy and transparency, it is difficult to see how a person could succeed in deceiving herself. Exactly what is it that the self-deceiver knows, and what doesn't she know?

We have seen that the belief that Anna Karenina buries in self-deception only hints at that which she deceives herself about: she buries the belief that she desires not to renounce Vronsky, and that which she deceives herself about is her strong desire, or love, for Vronsky. Thus, to say that the self-deception draws on knowledge of that which she buries is not to say that the belief that she buries expresses exactly that which she realizes when she realizes that she has been deceiving herself all along. But if self-deception involves a plan, it must include that which she deceives herself about as the aim of the deception (in the case of Anna Karenina, as interpreted by Gardner, to create time together with Vronsky). How can this aim then be hidden to her as it must be for the self-deception to succeed? The burial and promotion of desires and beliefs in practical reasoning are means in the plan to deceive herself so as to obtain her goal. Gardner says that intention and rationality are required to explicate the relations of instrumentality needed to explain how a desire can make use of the right means, since reasoning must determine what the right means are for a desire to make use of.[201] But for reason to determine this, reason must know the

[198] Ibid. p. 18
[199] Ibid. p. 103.
[200] Ibid. p. 109.
[201] Ibid. p. 20.

desire it promotes. If one does not assume a divided mind, which Gardner does not, how then is self-deception possible?

Gardner says that the self-deceiver follows a plan, but that the plan is kept from view while she is deceiving herself. He discusses self-deceptive intent in relation to ordinary propositional attitudes and recognizes, as he says, a disanalogy. The intent is not stated or recognized before or during self-deception, but, says Gardner: "What we can however do instead is to postpone the evidence: at some later time Anna Karenina will think, or be able to think, 'so that's why I told myself…' And there is of course an explanation for why the intention does not show itself in the present tense: to do so would gain nothing for it, and risk its extinction."[202]

Gardner characterizes self-deception as an intentional act in which the self-deceiver is not cognizant of her intention or plan while she is deceiving herself. It is the same with preference. Although the self-deceiver is described as exercising a preference, she is not aware of choosing self-deception over awareness of conflict. As we have seen above, Gardner holds that intention is present in self-deception but only recognized by the self-deceiver in retrospect. He seems to take for granted that understanding in retrospect what made one act in a certain way must be seen as *a discovery of an intention that was there in the action all along*.

Gardner can maintain that a belief, preference, intention etc. is present in the self-deceiver's action without being acknowledged by her because of the sharp distinction that he makes between having a belief, preference, intention etc. and realizing it. Realization, says Gardner, is "distinct from the ordinary formation or 'onset' of belief."[203] He offers the following example: "I can be struck by, and hence realise, something that I already know (something I have said and acted on for a long time). The topic of realization — what one is struck by — again need not be a whole proposition: it may be only an 'aspect', as when one realises 'quite how dark' the sky is."[204] It is easy to recognize situations where one is struck by something which one already knew but hadn't quite realized the degree or depth of, such as, for example, realizing quite how dark the sky is, or when a certain aspect of something one already knows becomes salient.

Gardner says that realization "characteristically involves bringing together two matters, in such a way as to yield the kind of combination which is exemplified when one thing is visually seen as another: typically it is realised that one thing is *also* something else (that two descriptions apply to one and the same object)."[205] Anna Karenina's realization that she is head-over-heals in love

---

[202] Ibid. p. 22.
[203] Ibid. p. 26.
[204] Ibid.
[205] Ibid. Perhaps Marcel Duchamp's *Fountain* can serve as an example. It is a simple urinal exhibited as a work of art with the name of a famous artist tagged to it. In order to appreciate this work, one must see the simple urinal as a work of art and the work of art as a simple urinal. One

with Vronsky and that she has been deceiving herself about this can be seen as the coming together of different things of which she has already previously been aware, such as feeling very happy in his company, admiring him etc. We could say that she sees that these things, of which she has always been aware, are manifestations of her strong desire and love for Vronsky, but she had *not* previously been aware of these things *as manifestations of her love and desire*. In contrast, Gardner's quote above would seem to suggest that Anna Karenina already somehow *knew* of these feelings *as* feelings of love and desire. My point is that Anna first becomes aware *that* she is in love with Vronsky when she confronts and reflects upon her happiness in his company, her disappointment in not seeing him when he is expected and so forth. The belief "I am in love with Vronsky" was not there until she recognized that his pursuit of her "was the whole interest of her life".

According to Gardner's characterization of Anna Karenina's self-deception, she knows that she is in love with Vronsky, and she knows that she is a wife and a mother, but she experiences the world as divided. In order to move towards a realization that she is deceiving herself, she needs to undo this division, according to Gardner "[s]he would need to think, when with Karenin, that she loves Vronsky; and with Vronsky, that she is also a wife and a mother."[206] The problem, argues Gardner, is that Anna Karenina has different self-representations connected with Vronsky and Karenin, which she needs to bring together: "what is needed for realization is that she should form the thought 'I love Vronsky' with her *Karenin* self-representation, and form the thought 'I am a wife and a mother' with her *Vronsky* self-representation."[207] The passage from Tolstoy's book, however, doesn't suggest that Anna Karenina had formed the thought "I love Vronsky" before. Her realization doesn't consist in bringing two pieces of knowledge together, but rather in realizing and admitting to herself that she loves him in the first place. There is a scene in Tolstoy's book where Anna Karenina reacts to Vronsky's exclamations of love for her with anger and deprecation: "'What you were just talking about was a mistake, and not love.' 'Remember I forbade you to utter that word, that vile word', Anna said with a shudder." [208] There is a sense in which she knows that the feeling that they share is love, and that is why she so vehemently rejects Vronsky's advances. But we should also recognize that she is blatantly denying that what he feels is love. She will not accept 'love' as a description of his or her feelings at this juncture. Later she realizes that she cannot call it anything but love. This, I mean, shows that there is no Vronsky-representation "I love Vronsky" at this moment: not even

---

appreciates it, is provoked by it etc. because one realizes that such a common object, filled with simple, dirty, shameful associations, also serve as a work of art.

[206] Ibid.

[207] Ibid. p. 27.

[208] Tolstoy, p. 139.

when she is alone with Vronsky and he tells her of his love does she express or admit to hers, or even accept his avowal.

Thus, the problem is not that she cannot see one of her self-representations through the other, but that she cannot form the self-representation "I am in love with Vronsky". Anna Karenina happily speaks of herself as a mother and she doesn't avoid mentioning the fact that she is married to Karenin, but she doesn't speak the words "I love Vronsky". This shows, I suggest, that while being a mother and a wife is part of her self-representation, being Vronsky's beloved who loves him is not – yet. One could say that it is a self-representation that Anna Karenina is struggling to avoid recognizing. It is therefore misleading to say that Anna Karenina holds the belief "I am in love with Vronsky". It is not a belief that she recognizes only with her Vronsky-representation; rather she doesn't admit to it at all. Neither is it of the same as recognizing "quite how dark the sky is". When she eventually realizes that she is in love with Vronsky, it is not the same as simply realizing how very happy she is in his company and how very upset he can make her when he doesn't show up. The insight reveals an aspect to her that hadn't been clearly visible to her before. What makes Anna Karenina eventually realize, after matters are brought to head in her heated argument with Vronsky, are the things she already knew. The formation of the belief "I am in love with Vronsky" is inseparable from her recognition that she had been deceiving herself. One could say that once the belief was formed, self-deception was no longer possible. I will say more about this connection between conceptualization and insight in the next chapter.

As we have seen, Gardner argues that non-conscious preferences and intentions are constituent of self-deception, and that the self-deceiver holds the belief that she is deceiving herself about without realizing that what it means. Gardner's claims here derive from his understanding of self-deception as "ordinary irrationality", which he says is similar to ordinary rational reasoning in that it is intentional and includes propositional attitudes. This assumption leads him to presuppose that self-deception includes non-conscious preferences and non-conscious intentions, and that the self-deceiver's desires are represented to her as beliefs. Since self-deception is seen as part of the ordinary and thus as explicable in terms of the structure of ordinary rationality, Gardner needs to adopt ways in which to account for self-deception as irrational yet ordinary, for instance, that although the self-deceiver acts with intention, she doesn't know her intention immediately but will recognize it later. Further, the postponed acknowledgement serves a purpose, since the deception would not succeed if the self-deceiver knew of the intention in her self-deception at the moment of deceiving herself. But if the self-deceiver doesn't know of the beliefs, intentions, preferences etc., which Gardner ascribes to her, what ground is there to assume that they are there in her self-deceptive behavior? Why assume that there is intention in self-deception that the self-deceiver knows of only in retrospect, rather than admitting that the fact that the self-deceiver can understand her behavior and why it took the path it took in retrospect, need

not imply that there was a self-deceptive intention there all along? However clever and coherent an explanation we can provide for self-deceptive behavior may be, and however well we can account for the behavior by reference to prior patterns of thought or behavior in retrospect, none of this constitutes a proof that it is intentional, or that it is a means of promoting a desire through practical reasoning, or that there is some inchoate "plan" at work, of which the self-deceiver becomes aware only latter.

I have asked the following: when the self-deceiver is presented as having so much self-knowledge and her situation is portrayed as transparent to herself to such a great extent as it is in Gardner's account, can self-deception succeed at all? The self-deceiver has self-deceptive intentions, makes choices and buries and promotes beliefs in order to believe what she wishes to believe etc. Gardner seems to say that self-deception serves to uphold the confusion between how things are and how they are believed to be in order for the person to be able to go on living according to her desires, but I find it difficult to understand how the hyper-rational self-deceiver, who knows what she is up to in self-deception, could fail to notice the content that she is manipulating. Thus one might further ask, when self-deception is seen as sharing the same intentional and rational structure as normal rational action, how can the irrationality that it is supposed to manifest be accounted for? Gardner describes the irrationality of self-deception thus:

> Self-deceivers should be seen as mistakenly taking themselves to have solved their real problem in solving their psychological problem; or, put another way, as failing to make a proper distinction between psychological and real problems. This means that *self-deception is not fully rationalized at the level of meta-intention*: self-deception involves a plan in so far as all intention does, but there is confusion between how things are, and how they are believed to be.[209]

Gardner seems to indicate that self-deceivers are successful in escaping their worries, bad conscience etc. (solving their psychological problems), but that their "mistake" lies in not seeing that the self-deceptive plan only aims at avoiding to apprehend unpleasant feelings and not at solving the real problems which give rise to the discomfort. The self-deceiver confuses how she takes things to be with how they are. Self-deception, in so far as it is successful, makes worries, self-reproaches etc. go away, but this is only because the self-deceiver believes what she wishes was true; in reality nothing is changed and no problem is solved. When Gardner says that the self-deceiver but not the neurotic knows what she is up to, he seems to suggest that the self-deceiver's self-knowledge stretches so far as to knowing that she is manipulating desires and beliefs so as to believe what she wishes to believe, and that what she does *not* know is that believing what you wish to believe doesn't imply that what you believe is true. I am not sure this is what Gardner means; it seems very odd that a person who is

[209] Gardner, p. 18. My italics.

so clear-sighted and strategic as Gardner portrays the self-deceiver to be would not look past her desires to inquire into the truth.

Notice that Gardner moves irrationality up a level, to the level of meta-intention. Self-deception involves a plan consisting of the practical reasoning that plays an instrumental role in fulfilling the self-deceiver's desire. In this respect, self-deception is rationalized. But the self-deceiver makes herself believe something which is not true, and this is where irrationality comes in. Gardner states that the irrationality of self-deception lies in the content of the propositional states, and that it is not of a different kind than rational reasoning.[210] As we saw in the last chapter, Davidson holds that self-deception is irrational because the correspondence between the person's beliefs and the world, which we should assume of the beliefs of a rational being, is not there. This seems to be what Gardner refers to as the irrational point in the otherwise rational structure which he takes self-deception to be: it is not rational to make oneself believe something which is not true.

Human Beings as Rational Unities and Borderline Cases

Gardner's way of formulating his distinction between ordinary and extraordinary forms of irrationality tells us something about his view of human beings and normal action and behavior. He asks:

> How can the Ratman – as a rational believer and desirer, possessed of a single mind belonging to a psycho-physical unity – contain enough mental richness and disorder to generate the confusions and disintegrations involved in this deeper and distinct grade of psychoanalytic irrationality? What extension of the ordinary view of persons could account for this, without controverting the image of persons as rational unities? Nothing would appear to be further from the hyper-rationalism of the self-deceiver.[211]

The problem that Gardner raises here is how to understand the Rat Man's deeper grade of irrationality against the background of what he takes to be the ordinary. My concern, however, is what he takes to be the "ordinary view of persons". Formulations such as "the ordinary view of persons as rational unities" who hold "core beliefs which everybody can be assumed to share" provide the framework into which self-deception must fit, so it is important to look at how Gardner describes normal human behavior.

---

[210] "Explanations in terms of self-deception consist in attributing a sequence of *practical reasoning*, one in which beliefs play the same role as they do in paradigm explanations of rational actions. Self-deceptive intent does involve confusion, but this is a feature of the *content* of the propositional states comprising the sequence, and does not make it a different *kind* of sequence." (Gardner, p. 30)

[211] Gardner. p. 92.

Gardner describes us as "rational unities" who hold "core beliefs which everybody can be assumed to share". Further, he takes this description to be the common sense view. The first thing to note is that these descriptions, like many similar throughout his book, express a very particular view of human beings, that is, as rational through and through. First, I find this way of understanding ordinary human life and normal behavior deeply problematic. Second, when persons are perceived as "rational unities" whose beliefs, hopes, wishes etc. all fit together perfectly in being rationally related, the case of the neurotic is mystified. His way of acting is seen as completely alien to our normal way of acting, which is always intentional and rationally structured. The strong self-deceiver, although sharing the description "irrational" with the neurotic, is so far removed from the neurotic as to be placed on the other side of the rationality spectrum: she is referred to as "hyper rational". Her irrationality is strategic and directed at attaining a goal, while the neurotic is mechanically reacting to situations and dangers. In discussing Davidson's view of irrationality in the last chapter, I argued that it is misguided to account for the neurotic's action as consisting of elements of rational structures. I criticized the view of the neurotic's behavior as ordinary intentional and rational action with the exception that the rationality breaks down at some particular point. Now, Gardner does manage to avoid the pitfall of Davidson's rationalization of neurotic behavior, which I commend. What I find problematic in Gardner's account, however, is the radical distinction he makes between ordinary and extraordinary action and behavior, the latter, in his view, being the proper object of psychoanalytic theory. As we have seen, this distinction places self-deception on one side of the divide and neurotic action on the other, such that the "strong self-deceiver" is regarded as a "hyper-rational prime candidate" for "ordinary behavior".

I have considered the differentiations Gardner makes between ordinary irrationality and forms of irrationality treated by psychoanalysis. In passing, Gardner considers irrationality that lies in between these groups. He calls these borderline cases. In cases of self-deception, alleges Gardner, it is easy to identify the subject's motive for misrepresenting herself. In the case of Anna Karenina, for example, the motive for her self-deception is to gain time together with Vronsky for their love to grow. But there are other cases where the motive is obscure. "[P]eople often act from motives suggesting that they envisage the world in ways that it would be extremely difficult for them to articulate […] they are unable to recognise and identify the motives that move them."[212] Gardner continues by reflecting, "in such cases, the stretch of mental distance between a person's motive and her self-awareness is unusually great. These

[212] Ibid. p. 88.

examples cannot be given the explanation that self-deception can be given."[213] In such borderline cases

> there is a sort of detachment from reality: the person "imagines", "seems to think", "behaves as if", "appears to suppose", that something, which does not match reality, is the case, and this unreal envisagement of theirs is integral to their unavowed motive. *Because of the difficulty we think the subject would have in articulating her motive*, we would hold back from saying that they *believe* the relevant proposition. This suggests an attitude on the part of the irrational subject which is akin, but not straightforwardly equivalent to belief (an idea which will be made more definite later).[214]

The cases that Gardner describes above are motivated, but the subject does not articulate the motive, and only with great difficulty is it even possible to articulate. It has not yet taken form as a belief. It is, as we will see later, what Freud calls an 'idea' or an 'ideational representative'.[215] Gardner further develops his view of the relation between the irrational phenomena that stand on the border between strong self-deception (ordinary irrationality) and what belongs to the kind of irrationality treated by psychoanalysis.

> These conditions of first personal mental opacity, which are recognized in ordinary psychology, suggest a rudimentary sense in which mental states, without being self-deceptively buried, may nevertheless be inaccessible to their owner. They may be unlike those psychological states – paradigmatically, current perceptual beliefs – that we self-ascribe immediately, and their self-ascription may require their being *found out*. This can take one of the two distinctive routes: a sudden flash of realization, or interpretation.[216]

What makes these borderline cases different from self-deception as Gardner defines it, it seems, is that in this kind of mental opacity there is no such thing as hiding what one already knows. The subject has not become aware of her mental states, that is, she has not formed beliefs about them. Gardner says that self-deception, borderline cases, and the irrationality treated by psychoanalytic theory show different kinds of inaccessibility and degrees of self-knowledge

[213] There are cases that are mixtures of self-deception and borderline cases**.** Gardner refers to the example of Jane Austen's *Emma* as such a case. Emma realizes her self-deception, i.e. that her actions have been motivated by love for Knightley, but she still doesn't appreciate the deeper features of her motivation: her exaggerated attachment to her father. She is ignorant of the deeper motives for her action. (Gardner, p. 261, n. 6.)
[214] Gardner, pp. 88. The first occurrence of italics is mine.
[215] Gardner refers to 'idea', in Freud's use of the term, when he says: "this suggests an attitude on the part of the irrational subject which is akin, but not straightforwardly equivalent to belief (an idea which will be made more definite later)." (Gardner, p. 89)
[216] Gardner, p. 89.

since mental states differ in the facility with which they become topics of self-knowledge.[217]

My characterization of self-deception, which I contrast with Gardner's, has affinities with what Gardner calls borderline cases. I have suggested that what the self-deceiver deceives herself about is not a belief but something of which she has not yet formed a belief. She is affected by whatever it is that evokes her discomfort to the extent that she reacts by avoiding coming to grips with it. Anna Karenina might behave as if she could avoid the real problems that she is confronted with by avoiding her self-accusations, which is to say, deceiving herself. Although we could say with Gardner that she has a motive for behaving as if everything is in order to avoid confronting her anxiety, this motive is not a belief. It is manifested in her behavior, but it is not articulated. In the next chapter, I will develop the idea of self-deception as a way of actively avoiding discovery of one's real motives, feelings and desires, without assuming that what is avoided is something which one already knows.

Intention in Action

Gardner writes that self-deception could not be made sense of except as an intentional action where burial and promotion of beliefs are means for attaining the goal of fulfilling a desire, while keeping the same desire unacknowledged. This is the structure of self-deception, according to Gardner, although it is only in retrospect that the self-deceiver realizes what roles these beliefs played in the self-deception. I have suggested that it is unjustified to assume that self-deception *is* practical reasoning directed at promoting the desire of which one deceives oneself just because it can be interpreted in those terms by a third party, or by myself in retrospect. It could be argued that the ascription of intention to the self-deceiver (or to oneself in retrospect) is a rationalization of a behavior that is neither intentional nor rational.

Gardner argues that forms of irrationality such as the Rat Man's cannot be assumed to consist of intentions and beliefs, since the Rat Man's "beliefs" would be irrational, and this, according to Gardner, cannot be assumed of beliefs; they must be rationally formed. Gardner assumes an explanation of the Rat Man's behavior outside of rational reasoning, and in so doing avoids making the mistake that we see Davidson make when he describes every step in

[217] Gardner separates inaccessibility into *accidental* and *non-accidental inaccessibility* and takes self-deception as a case of accidental inaccessibility. In accidental inaccessibility, the mental state has an accidental property as, for example, that it simply has been forgotten, or buried in self-deception. In the case of the non-accidental inaccessibility of mental state, on the other hand, Gardner holds that the inaccessibility derives from its being a mental state of a certain *type*: states fundamental to psychoanalytic explanation are, as types, minimally accessible. (Gardner, p. 89)

the Rat Man's behavior as if it was a rational intentional action.[218] Gardner rightly points out that an explanation of the case of the Rat Man should not assume a structure of instrumentally related beliefs and self-deceptive intentions, but in this chapter I have raised critical questions concerning Gardner's own account: what makes a structure of beliefs, intentions and practical reasoning necessary and appropriate in accounting for self-deception? Gardner provides a characterization of self-deception which is similar to the characterization that Davidson gives of the Rat Man's behavior: a rational characterization in terms of beliefs, desires and intentions where irrationality comes in only at a meta-level, i.e. at the level of meta-intention. I criticized Davidson's rational reconstruction of the Rat Man's obsessive behavior on the grounds that it was a forced explanation made from the point of view of a rationalist observer. Here I argue that Gardner is tempted in the same direction as Davidson, although he doesn't go as far. Nonetheless, the fundamental assumptions in Gardner's characterization of self-deception are questionable. Just as Davidson is compelled to give an intentional explanation to the case of the Rat Man, Gardner is also compelled to provide an intentional explanation for self-deception. The guiding assumption, in both cases, is that to provide a philosophical account of human behavior means to explain it in terms of intentional actions displaying a propositional structure. As we have seen Gardner claim, one of the main ambitions of his book is to show that irrationality does not have to be given up to psychoanalysis, but that many forms of irrationality can be accounted for within the rational framework, which he thinks is provided by "ordinary psychology".

My point is not that theoretical concepts belonging to psychoanalysis, such as *phantasy* or *wish-fulfilment*, are necessary for an account of self-deception, but I do oppose the view that self-deception must have the propositional and rational structure that Gardner ascribes to it in separating it from forms of irrationality treated by psychoanalysis. I argue that this characterization is the result of a rationalization that Gardner makes in trying to account for self-

[218] As we have seen, Davidson makes this mistake in interpreting a passage from the case study of the Rat Man "Here everything the agent does (except stumble on the branch) is done for a reason, a reason in the light of which the corresponding action was reasonable. Given that the man believed the stick was a danger if left on the path, and had a desire to eliminate the danger, it was reasonable to remove the stick. Given that, on second thought, he believed the stick was a danger in the hedge, it was reasonable to extract the stick from the hedge and replace it on the path. Given that the man wanted to take the stick from the hedge, it was reasonable to dismount from the tram and return to the park. In each case the reasons for the action tell us what the agent saw in the action, they give the intention with which he acted, and thereby give an explanation of the action. Such an explanation, as I have said, must exist if something a person does is to count as an action at all." (Davidson, "Paradoxes of irrationality", p. 172) Jonathan Lear points out that even if we accept the doubtful move of dividing the Rat Man's action into clusters of actions, it is very hard to accept some of these reason explanations: the Rat Man's replacing of the stick *on the path* can hardly be seen as reasonable but must rather be seen as a compulsive act. (Lear, *Freud*, pp. 29.)

deception within ordinary psychology as a rational strategy directed towards promoting a desire. Gardner says that self-deception manifests intention but that the evidence is postponed since the self-deceiver can only recognize the intention in retrospect, and that she may or may not recognize it. To the extent that this account of what occurs in self-deception seems fitting, it is precisely as an *account*, a way of *re-describing* a course of events so that it fits together in a cohesive and structured whole. But Gardner, like Davidson, seems to vacillate between acknowledging that his account is some kind of rational reconstruction and claiming to have captured the phenomenon itself. The self-deceiver, in trying to make sense of her behavior afterwards, may indeed think "So that's why I did *x*". I suggest that she is not, in this case, ascribing an intention to her behavior but expressing that her behavior is beginning to make sense to her, viewed in the light that is cast upon it after she has realized that she has been confused about her feelings and reactions. This process can look much like the neurotic's, as he comes to understand that there is an answer to why he acted as he did. Freud's "Notes upon a Case of Obsessional Neurosis" can serve as an example. There is an answer to why the Rat Man obeyed the Captain's order. The answer is not that the Rat Man intended to obey his father, who the Captain represented, but he obeyed the Captain's order because there was an unconscious association between the Captain and his father which made the Rat Man *react* to the Captain as to his father.

The analogy between the cases of Anna Karenina's self-deception and the Rat Man's compulsive act should be taken as an analogy which, at best, can reveal some kinship between what it is like for Anna Karenina to understand what her self-deception consists in and what it is like for the Rat Man to understand what his obsessional behavior expresses. Self-deception and neurosis are different conditions and should not be assimilated. I believe we can understand the case of Anna Karenina in this way: Anna Karenina comes to understand that she reacted by thinking that she was displeased that Vronsky courted her so openly, that he was dishonorable etc., because the feelings of desire that Vronsky's presence and his exclamations of love arose in her were threatening to her. They evoked feelings and thoughts that were in conflict with who she took herself to be (who she thought she ought to be; who she wanted to be) and with other strong desires of hers (to be with her son and be a good mother to him). In reflecting upon her self-deception, Anna Karenina thus reaches a better understanding of her behavior, but this need not mean that she has discovered previously concealed intentions, beliefs and preferences. One easily falls into an interpretation in terms of intention, execution of preferences and manipulation of beliefs when one understands self-deception as a hyper-rational kind of irrationality. This, I have argued, is a misleading presupposition in Gardner's account.

In my view, Gardner fails to account for something central to self-deception because of how he defines self-deception. In accounting for self-deception as a rational action directed at promoting a desire, Gardner's account leaves out a

fundamental feature of self-deception, namely, the self. By focusing exclusivly on how "we" can describe self-deception, he neglects the element of *inner conflict* – a conflict between one's own desires, ambitions, responsibilities etc. – where one misleads oneself about, or is oblivious to, a desire (i.e.), not *in order to* satisfy it but because of *other* desires, ambitions, responsibilities etc. In Gardner's interpretation, Anna Karenina has the belief that she is displeased that Vronsky courts her in order to mask her desire for him to herself *so that* it can proceed towards fulfillment (her desire is partly revealed in her belief that she desires not-to-believe that she ought to renounce Vronsky). The self-deceiver's behavior and thought promote the desire that was buried at the onset of self-deception. This is how one can see that self-deception is a preference, albeit an non-conscious one.[219] The self-deceiver tricks herself into fulfilling her desires, much as the liar tricks the one to whom he lies. The liar often aims to benefit from the lie, just as the self-deceiver in Gardner's account is portrayed as fulfilling her desire through her self-deception. Rather than reducing the varying and perhaps incompatible thoughts of the self-deceiver to a means for achieving an end, we might see them as expressions of varying and perhaps incompatible desires. Anna Karenina's belief that she is displeased with Vronsky can be seen as an expression of another desire of hers, such as the desire to be a good mother to her beloved son. If seen in this way, her self-deception is not a strategy directed at promoting the desire to be with Vronsky, nor is it a preference to promote this one desire, but it is rather a *situation* of conflict between two (or more) desires, or other motivations, in which Anna Karenina actively but unreflectively refrains from realizing one of these motivations (the desire to be with Vronsky), and thus remains unaware of the conflict.

In this view, self-deception is not comparable to tricking someone else into believing something by lying. Gardner's characterization of self-deception as a promotion of a desire resembles a battle with something outside oneself, rather than a conflict with and within oneself regarding what one wants, or what one thinks best to do. Self-deception as an inner conflict — as a conflict between different things that one wants very strongly, or, perhaps, between what one wants and what one knows one ought to do — is neglected in Gardner's account.

[219] As we have seen, in Gardner's view, the loose criterion for taking a behavior to manifest a preference is that desires are satisfied by that behavior and that the person is justifying his behavior by operations of thought. A preference does not need to be recognized by the person whose preference it is, it can be non-conscious.

Sigmund Freud

Self-Deception as Flight from Anxiety

Introduction and Outline

In this chapter I will study self-deception on the basis of Sigmund Freud's work. Although Freud doesn't inquire into the concept of self-deception specifically, as Davidson and Gardner do, his discussions of illusion, repression, (psychiatric) delusion, different kinds of knowledge with respect to oneself and one's mental life, as well as other discussions that I will examine in this chapter, are highly relevant for my study. It is through an analysis of these writings that I will approach the notion of self-deception in this chapter. My portrayal of self-deception here should not be read as a claim that this is what I take to be *Freud's* account of self-deception. Freud does not offer an account of self-deception as such, but I aim to show that his writings are a fertile ground for understanding self-deception. In my view, Freud's writings reveal very important characteristics of self-deception that are not acknowledged in the two previous accounts. Freud's texts are also interesting in that he dwells on many of the topics that have been central in this study up to this point, such as intention and the status of the object and aim of self-deception. In this chapter I will broaden the context in which self-deception is to be understood and described.

A central topic for the investigation of self-deception is the question whether it can and should be sharply distinguished from, for instance, illusions and delusions about oneself. As we have seen, Gardner separates self-deception from these, which he calls forms of "motivated self-misrepresentation", and claims that self-deception, but not delusion, self-distraction, self-manipulation, etc. includes beliefs or intention.[220] In this chapter, I will introduce a perspective on self-deception which shows that self-deception cannot be sharply distinguished from these phenomena without serious loss of meaning. Rather, I will argue, the context of illusion, delusion etc. is essential for understanding self-deception. Another important question is whether self-deception is qualitatively different from psychological illnesses, or what Gardner refers to as "deeper forms of irrationality". In this chapter, I will discuss texts where Freud presents this distinction as being largely a difference of degree rather than kind. Third, I will consider what Freud writes about reactions to anxiety-provoking situations, and suggest that we should understand self-deception as a flight from anxiety – as, in itself, a defensive reaction to something that is perceived as a threat. Fourth, I will consider whether self-deception is an intentional action by considering passages where Freud presents the idea of unconscious intentions. Fifth, I will return to the perceived threat and investigate its status: does the self-deceiver know what it is that threatens him? Can he express it as a belief? I will suggest that self-deception often sets in prior to the formation of a belief. I also wish to show that the level of awareness varies in different cases of self-

[220] Chapter Two, pp. 69.

deception and that Freud's texts lend themselves to this interpretation. In Davidson's and Gardner's accounts, as we have seen, self-deception starts with the holding of a belief that gets buried, causing the development of another belief. Rather than self-deception being a burial of a belief, I will suggest ways in which forming a belief is avoided and provide examples of circumstances and feelings that give rise to self-deception.

Two Conceptions of Self-Deception

I will open this chapter by presenting two conceptions of self-deception that are reflected in a tension in the concept 'self-deception'. Through an examination of Davidson's and Gardner's accounts of self-deception, we have become familiar with self-deception understood as an intentional action involving manipulation of beliefs, an action which is analogous to deceiving someone else. The word 'self-deception' lends itself to this interpretation, as does the German word *Selbstbetrug*. Already in tracing the occurrences of 'self-deception' in *The Concordance to the Complete Psychological Works of Sigmund Freud* (abbreviated SE)[221] back to the original German, we see that the word that Freud uses most frequently is *Selbsttäuschung*.[222] *Selbstbetrug* and *Selbsttäuschung* are not perfectly synonymous: *Selbstbetrug* is usually translated to English as 'self-deceit', 'trickery' or 'deception', while *Selbsttäuschung* is translated alternatively as 'self-deceit' or 'delusion'. I will consider this other perspective on self-deception suggested by the word 'Selbsttäuschung' (delusion). Freud's texts provide a rich context for this, but before turning to Freud, let us first acquaint ourselves with the concepts *Täuschung* and *Selbsttäuschung*.

Historisches Wörterbuch der Philosophie describes *Täuschung* ('delusion') as synonomous with *Blendwerk* ('phantasmagoria') and *Illusion* ('illusion'), and as related to *Schein* ('appearance') on the one hand, and *Irrtum* ('mistake' or

[221] Samuel A. Guttman, *The Concordance to the Standard Edition of the Complete Psychological Works of Sigmund Freud*, 2:d ed. (London: International Univ. Press, 1984). The concordance is a helpful register of concepts discussed or mentioned in SE. Here one can see that the listing of occurrences of the concept 'deception' and other forms of that concept covers little more than half a page, and that the occurrence of the concept 'delusion', and other forms of that concept, cover more than two pages. Freud's discussion of delusion is much more extensive than his discussion of deception. Unfortunately, the concordance to Freud's works in the original language, German, is not as complete as the English, so one cannot as easily get an overview of the occurrences of the German concepts. Nevertheless, one can see that Freud's discussions of root terms, i.e. *Täuschung* (delusion), *Illusion* (illusion) and *Entlarvung* (exposure, unmasking) are extensive. (Sigmund Freud, *Gesamtregister* in *Gesammelte Werke* (*GW*) Ed. Anna Freud, Frankfurt am Main: Fischer Verlag, 1972, Band 18.)

[222] There are only three passages listed where the word 'self-deception' occurs. In comparing these three passages in the English translation with the German original, we see that Freud uses the word 'Selbsttäuschung' in all the three passages.

'error'), on the other.[223] *Täuschung* – 'delusion' in English – is thus a cognitive term, i.e. it has to do with the perception (or misperception), apprehension (or misapprehension) and/or understanding (or misunderstanding) of something. Someone can intentionally try to delude someone else, for example by misleading his sensual apprehension. A football player can move as if he were going to try to score in the upper right corner of the goal and then make a quick move to score to the left. The magician (illusionist) is also skilled in moving so as to direct the viewers' attention away from his action and towards what he makes it appear that he is doing. In the cases above, delusion involves the intention of misleading someone, but an intentional act with the purpose of misleading someone is not necessary for delusion to occur. I can also be deluded in mistaking something for something else; for example, I can be under the illusion that I see a person standing alongside the road when it is the fog that has accumulated into a shape that I apprehend as a person. *Historisches Wörterbuch* further differentiates between delusion of the senses (*Sinnestäuschung*) on the one hand, and delusion of the logical capacity or delusion of reason (*logische Täuschung* or *Versandestäuschung*)[224] on the other, referring to the cases in which what one apprehends sensually or what one apprehends intellectually is mere appearance.[225] An example of a delusion of reason would be to believe that the conclusion of a syllogism necessarily expresses something that is factually true while all that is assured, if the syllogism is correctly constructed, is logical truth: the conclusion follows logically upon the premises. As with delusion of the senses, delusions of the logical capacity/delusions of reason *can*, but need not, be the result of an intentional act of misleading. Thus, there need not be an intention to delude for delusion to occur. One can simply become deluded by how something appears to be and this delusion can be a perceptual illusion or a delusion of reason.

How shall we understand *Täuschung* ('delusion') as *Selbsttäuschung* – as "deluding oneself" or as "self-deception"? How can I delude myself about myself? Or the other way around: how can I be deluded by myself about myself? First, let us recall the examples of delusion above. We have seen that there need not be someone who intentionally deludes me for me to be deluded. I can be deluded by fog or by an argument that appears convincing. Analogously, it seems that I can delude myself without intending to do so. I can allow myself to be deluded when I uncritically take appearance at face value, or in the willing suspense of disbelief. These examples show that delusion can result from a passive, uncritical acceptance of appearances. For example, I may

---

[223] J. Ritter, K. Gründer, *Historisches Wörterbuch der Philosophie* (Darmstadt: Wissenschaftliche Buchgesellschaft). Band. 9 (Se/Sp), 1995; Band. 10 (St-T), 1998. Band. 10, s. 928.

[224] Ibid. ss. 927. Immanuel Kant calls these two forms of *Täuschung*, *empirischen Schein* and *logischen Schein*.

[225] Ibid. s. 927. Among the examples of delusions of reason are fallacies (*Trugschluss*) and prejudices (*Vorurteile*).

not consider the possibility that the fog can assemble into the form of a figure, and that the figure I see can be an aggregation of the fog and not a person. In this example I do not intentionally *make* myself believe in the appearance unreflectively; rather, there are shifts in attitude that makes one more or less prone to being deluded in different circumstances.

The article on *Selbsttäuschung* in the *Historisches Wörterbuch der Philosophie* opens with the assertion that: "The effort to liberate oneself from error, deception, illusion and bedazzlement belongs to the characteristics of the *Denkraum* from which Greek philosophy has grown."[226] That human thinking is held captive by its own delusions is a *fundamental experience* and the starting-point of philosophy since Socrates.[227] This is demonstrated, famously, by Socrates' great wisdom in not believing that he knows what he does not know. The insight that we are always caught up in delusion together with the experience that we can liberate ourselves from delusions forms the philosophical context for understanding delusion about oneself. While in Davidson's and Gardner's understanding self-deception is construed as an intentional action directed at convincing oneself of that which one wants to believe, according to the definition of *Täuschung* and *Selbsttäuschung* above, if there is an intentional act involved here, it lies not in deluding or deceiving oneself but rather in the conscious effort to *liberate* oneself *from* delusion. Rather than saying that self-deception serves the goal of fulfilling one's secret wishes through an active intentional act, one may instead hold that a passive attitude results in remaining under the sway of delusion. Another important difference is that in Davidson's and Gardner's understanding, self-deception is understood as a *process of reasoning*, while self-deception as *Selbsttäuschung* suggests that self-deception begins already with how one apprehends something which one perceives. Self-deception is here understood as something that affects both apprehension and intellectual understanding.[228]

---

[226] Ibid. Band. 9, s. 539. "Das Bemühen, sich von Irrtum, Trug, Wahn und Verblendung zu befreien, gehört zu den Charakteristika des Denkraums, in dem sich die griechische Philosophie von ihren Anfängen her ausbildet."

[227] Ibid.

[228] In her paper „Selbsttäuschung", Hilge Landweer nicely presents how self-deception differs from lying. The self-deceiver ignores that which does not fit in with her view, or interprets it so that it fits in with her self-understanding. Self-deception does not require any acting, while deceiving someone else, on the other hand, does require that one can put on a show. Self-deception is *not*, in her view, to make oneself believe something that is in conflict with what one knows, rather it is to make oneself believe what one wishes were true, or to feel what one wishes that one would feel, and to keep oneself from knowledge or insight into one's real feelings. (Hilge Landweer, „Selbsttäuschung", in *Deutsche Zeitschrift für Philosophie*, 2001, 49/2, s. 209-227.)

Illusion and Self-Understanding

If the claim that self-deception is an active intentional action directed at coming to hold the desired belief or fulfilling the forbidden wish is jettisoned in favor of holding that it is rather a matter of passivity or complacency with regard to the possibility of freeing oneself from delusion, it may be objected that such a passive attitude does not in itself qualify as *self-deception*. One might want to distinguish sharply between self-deception, on the one hand, and delusion or being under an illusion about oneself regarding one's abilities, feelings etc. on the other, as Davidson and Gardner argue.[229] They want to find ways in which to account for self-deception that clearly separate self-deception from something to which one is subjected; they want to see it rather as something that one does to oneself.[230] I will argue that an intentional account of self-deception is not necessary to characterize self-deception in a way that distinguishes it from something to which one is merely subjected, as if it were a force from without. Rather, I will argue, there is an element of motivated lack of self-knowledge, self-reflection and responsibility in delusion – an element of avoidance – that makes it accurate to say that someone actively *allows himself* to be deluded. I will turn to Freud's discussion of illusion in *The Future of an Illusion*, where he indicates that illusion is not something that just happens to someone. Freud writes:

> An illusion is not the same thing as an error; nor is it necessarily an error. […] it was an illusion of Columbus's that he had discovered a new sea-route to the Indies**.** The part played by his wish in this error is very clear. One may describe as an illusion the assertion made by certain nationalists that the Indo-Germanic race is the only one capable of civilization; or the belief, which was only destroyed by psycho-analysis, that children are creatures without sexuality. What is characteristic of illusions is that they are derived from human wishes. In this respect they come near to psychiatric delusions.[231]

---

[229] It will be recalled that for Davidson and Gardner, self-deception is intentional and involves propositional attitudes (beliefs), and that the self-deceiver knows what he is up to.

[230] For example, in Davidson's account the self-deceptive belief is promoted in a process of practical reasoning, and Gardner distinguishes self-deception from other forms of motivated self-misapprehension by holding self-deception to be an intentional action.

[231] Sigmund Freud, *The Future of an Illusion* (1928) in *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (*SE)*, edited by James Strachey, (London: The Hogarth Press and The Institute of Psycho-Analysis), vol. 21, pp. 30. My italics. „Eine Illusion ist nicht dasselbe wie ein Irrtum, sie ist auch nicht notwendig ein Irrtum. […] Dagegen war es eine Illusion des Kolumbus, dass er einen neuen Seeweg nach Indien entdeckt habe. Der Anteil seines Wunsches an diesem Irrtum ist sehr deutlich. Als Illusion kann man die Behauptung gewisser Nationalisten bezeichnen, die Indogermanen seien die einzige kulturfähige Menschenrasse, oder den Glauben, den erst die Psychoanalyse zerstört hat, das Kind sei ein Wesen ohne Sexualität. Für die Illusion bleibt charakteristisch die Ableitung aus menschlichen Wünschen, sie nähert sich in dieser Hinsicht der psychiatrischen Wahnidee." (Sigmund Freud, *Gesammelte Werke* (*GW*) Ed. Anna Freud (London: Imago Publishing), Band. 14 (1948), s. 353.

We learn from Freud's analysis of this example that an illusion, as opposed to a mere error, is *motivated*; the "strength" of the illusion, the reason why it persists for so long and is so hard to overcome, is that the person *wishes* it to be true, or is in other ways motivated to hold it to be true.[232] Nothing of what Freud says about illusion here suggests that one brings oneself under an illusion in an intentional act of practical reasoning, an act that can be conscious or unconscious. Nothing even suggests that one is aware that the underlying attitude or belief[233] is *a wish* that one harbors. Still, the underlying attitude or belief, as far as it is a wish or other motivation, is intrinsically implicated in the development and maintenance of the illusion, and is equally involved in resistance to give it up when it is criticized, questioned or threatened. According to Jonathan Lear, "[t]he essential problem for an illusion, then, is that we are mistaken about the basis of our commitment to it. We take it to be a belief based on responsiveness to the world; in fact, it is held in place by primordial wishes of which we are unconscious."[234]

Columbus' belief that he had discovered the new sea-route to the Indies, suggests Freud, is an illusion, and not just a mere error, because it is his strong desire to discover such a sea-route that makes him so willing to believe that the land he sees is the Indies and so resistant to the possibility that this is not the case. When under the trawl of an illusion, one tends to see what fits in with what one wants to believe and to ignore or re-interpret anything that can threaten one's prior understanding. The impressions that inform Columbus' understanding are therefore not perceived and considered neutrally, nor are the conclusions that he draws from them unbiased, but his wishes filter and guide what he sees and how he understands it. Columbus' project of searching for a new sea-route to the Indies and his strong desire to find such a route makes him inclined to interpret what he finds as confirmation that the land that he has found is the Indies. Similarly, the Victorian attitude toward sexuality as something shameful or even perverse made Freud's contemporaries unwilling and therefore unable even to entertain the thought of children as sexual beings.[235]

The examples of illusion that Freud provides and discusses in *The Future of an Illusion* seem at first sight to have little to do with oneself (or rather one's

---

[232] The motivation is not always as straightforward as in the cases above. An illusion can also be a preconceived idea or "false knowledge", such as, for example, a preconceived idea of one's body as huge, fat, ugly etc. typical of anorexia.

[233] That the land is the Indies; that the Indo-Germanic race is the only one capable of civilization; that children are creatures without sexuality.

[234] Jonathan Lear, *Freud*, p. 204.

[235] Another important feature of the quote about illusion above that should be kept in mind is that it expresses a close connection between 'illusion' (*Illusion*) and 'psychiatric delusion' (*Wahnideen*). In this case, as in many others, Freud makes no attempt to demarcate the pathological from the normal, but characterizes illusion through comparison with psychiatric delusion. He does, however, distinguish between pathological and normal behavior. We will return to this later.

self). Columbus' illusion is about what part of the world it is that he has reached; the nationalist's illusion is that one race, his own, is the only one capable of civilization; and a common illusion of Freud's time, he says, regards the asexual nature of children. But these examples show that there is no sharp border between what constitutes a delusion about oneself and what constitutes a delusion about something else. This is one of the implications of Freud's claim that illusions are derived from human wishes. Assuming that Columbus' greatest desire is to find a new sea-route to the Indies, it will *matter* to him if he finds it or not. Furthermore, we can easily imagine that there are further personal motivations for this wish such that achieving the desired goal would be important for his reputation, and thus his self-esteem. Similarly, it is hard to imagine that people other than Indo-Germans would proclaim the Indo-Germanic race to be superior; the superiority that the nationalist claims for the race, he claims for himself. One can imagine, of course, that this wish involves a sense of competition, threat, or compensation. The refusal to consider children as sexual beings can reveal a discomfort with the implication that one was not free from such morally condemned fantasies and actions even as a child. Nor are one's own children innocent in this respect. The three examples of illusions that Freud discusses show that illusions are difficult to uncover and surrender because they are so intimately connected with one's understanding of oneself, which, in turn, is motivated by what one wishes oneself to be.

I have suggested that other important aspects of self-deception become salient if we consider self-deception in the light of illusion, delusion, appearance etc. rather than seeing it in terms of simply lying to oneself. I have also tried to show that the connotations of the German *Selbsttäuschung* point in this direction: they show that self-deception need not be intentional and may directly affect apprehension. Freud differentiates illusion from error by virtue of the role wishes play in illusions. In illusion, apprehension and understanding are biased because wishes guide what one apprehends and how one interprets that which one apprehends. Freud writes that the ego is disposed already at the perceptual level to acknowledge that which it finds pleasurable and prevent unpleasant excitations from becoming acknowledged. Freud calls this disposition to see, believe or do whatever is most pleasurable the *Pleasure Principle*". In *Inhibitions, Symptoms and Anxiety*, he says that the perceptual system receives excitations both from without and from within, "and endeavors, by means of the sensations of pleasure and displeasure which reaches it from these quarters, to direct the course of mental events in accordance with the pleasure principle."[236] Thus, in Freud's view, illusions and psychiatric delusions

[236] Freud, *Inhibitions, Symptoms and Anxiety* (1926[1925]), SE, vol. 20, p. 92. Freud describes this as an instinctual reaction. „[E]s empfängt Erregungen nicht nur von aussen, sondern auch von innen her und mittels der Lust-Unlustempfindungen, die es von daher erreichen, versucht es, alle Abläufe des seelischen Geschehens im Sinne des Lustprinzips zu lenken." (GW, Band 14, s. 119.)

are misconceptions that arise because one apprehends what one wishes to be the case, or what one is in some way motivated to apprehend.

If being guided by a wish or other motivation is a central element in self-deception, as I believe it is, Freud's discussion of illusion, delusion, appearance etc. can help us understand self-deception better. We have seen that illusion is often already at work at the level of apprehension, and thus that it is not a result of practical reasoning. A central characteristic of self-deception in analogy with what Freud writes on illusion is that one takes for granted preconceived ideas. I take biased apprehension and understanding, as well as diminished self-reflection and deficient soul-searching, to be central traits of self-deception. In this respect, my view of self-deception differs from Gardner's and Davidson's, where it is assumed that self-deception[237] is an intentional action in which beliefs are manipulated. In these two accounts, little if any attention is paid to the relation between apprehension and the formation of beliefs; beliefs and evidence are already presumed.

Pathology and the Normal

The cases we have discussed so far all belong to what would count as normal behavior according to a common understanding of what is normal and what deviates from it: the crazy, neurotic, etc.[238] A normal person, perhaps in somewhat unusual circumstances, can be deluded by something he sees e.g., in seeing the accumulated fog as a person, he is deluded by the senses. In implicitly trusting that the form of the argument or the practice of logical argumentation will also assure that the conclusion expresses a true fact, he suffers from a delusion of reason. And finally, in allowing taboos or prejudices to limit or guide what he sees or understands, he suffers from what we might term "delusions of the will", or perhaps "delusions of judgment": he believes what he wants to believe and/or what he thinks he ought to believe.

Unlike Gardner, I do not proceed from an assumption that self-deception can and should be sharply distinguished from pathology (or, following Gardner's distinction, deeper forms of irrationality). My approach is rather to turn to texts where Freud discusses psychical illness and pathology with the aim of trying to understand self-deception through these texts. My investigation of self-deception thus requires a discussion of the relation between self-deception and pathology. A central assumption in Gardner's account it will be recalled, is that self-deception, as a form of ordinary irrationality, and neurosis, as a form of deep irrationality, are different in kind and thus require different forms of explanation. I aim to show that Freud's texts provide little support for drawing

[237] In Gardner's case, strong self-deception, and in Davidson's, core cases of self-deception.

[238] With the exception of psychiatric delusion, which Freud mentions in his characterization of illusion.

a rigid distinction between ordinary and deep forms of irrationality (between what is normal and what is pathological). I will argue that self-deception is better understood if we consider its similarities to pathological conditions rather than if we start off from the view that self-deception must be distinguished from such conditions.

The very title of one of Freud's most well-known texts, *Psychopathology of Everyday Life* makes clear that there is no absolute distinction to be made between "the normal" and "the pathological". Here Freud analyzes ordinary "slips"[239], such as slips of the tongue (often his own), or the forgetting of impressions and intentions etc., as pathologies of the everyday. Like obsessional neurotic actions, for instance, they are seen not as chance events but as influenced by a motive, and as such are open to psychoanalytic explanation.[240] In this text, Freud notices that even in healthy, non-neurotic people there is resistance to recalling distressing impressions and thoughts. Freud offers the example of how a patient reacted to his remark about something that she had told him earlier – that her children were bed-wetters: she denied ever having said any such thing. Freud concludes that she must have "forgotten" what she had said. This forgetting, Freud notes, is motivated. In the accompanying footnote, Freud lists a number of authors who, like himself, "appreciate the influence of affective factors on the memory and who – more or less clearly – recognize the contribution towards forgetting made by the endeavour to fend off unpleasure."[241] Affective factors influence what is or is not recalled in healthy, non-neurotic people and in neurotics alike, but it can be seen most clearly in neurotics, says Freud:

> There are thus abundant signs to be found in healthy, non-neurotic people that the recollection of distressing impressions and the occurrence of distressing thoughts are opposed by a resistance. But the full significance of this fact can be estimated only when the psychology of *neurotic* people is investigated. We are forced to regard as one of the main pillars of the mechanism supporting hysterical symptoms an *elementary endeavour* of this kind *to fend off* ideas that can arouse feelings of unpleasure – an endeavour which can only be compared with the flight-reflex in the presence of painful stimuli.[242]

---

[239] *Fehlleistungen*.

[240] Freud, *Psychopathology of Everyday Life* (1901), SE, vol. 6, p. 4.

[241] Ibid. p. 146. „A. Pick hat kürzlich eine Reihe von Autoren zusammengestellt, die den Einfluss affektiver Faktoren auf das Gedächtnis würdigen und – mehr oder minder deutlich – den Beitrag anerkennen, den das Abwehrbestreben gegen Unlust zum Vergessen leistet." (GW, Band 4, s. 162, Anmerkung 2.)

[242] Ibid. pp. 146. „Man findet also auch bei gesunden, nicht neurotischen Menschen reichlich Anzeichen dafür, dass sich der Erinnerung an peinliche Eindrücke, der Vorstellung peinlicher Gedanken, ein Widerstand entgegensetzt. Die Wolle Bedeutung dieser Tatsache last sich aber erst ermessen, wenn man in der Psychologie neurotischer Personen eingeht. Man ist genötigt, ein solches *elementares Abwehrbestreben* gegen Vorstellungen, welche Unlustempfindungen erwecken können, ein Bestreben, das sich nur dem Fluchtreflex bei Schmerzreisen an die Seite stellen last, zu einem der Hauptpfeiler des Mechanismus zu machen, welcher die hysterischen Symptome trägt." (GW, Band 4, ss. 162.)

Freud is not suggesting here a qualitative difference between the healthy and neurotics, only that the reaction of fending off displeasure is strong and basic in neurosis.[243]

Defensive Reactions

I have suggested earlier that a characteristic trait of self-deception is the lack of impetus towards self-knowledge or self-reflection. As in the case of illusion, this passive attitude is also motivated in self-deception. The motivation can be the desire to hold on to an idea that one wants to be true, and by so doing, to avoid facing what evokes anxiety. But I argued further that something more than the absence of an impetus towards self-knowledge and transparency of thought is needed to account for many cases of self-deception. I will suggest that Freud's treatment of different forms of defensive reactions, or defense mechanisms, can capture this other element. In this chapter, I will turn to Freud's account of repression and other forms of defensive reactions in order to look for a framework for understanding self-deception where it is not conceived in terms of rational, goal-directed action. My aim here is chiefly to suggest that self-deception can be seen as a defensive reaction which can include various other kinds of defense mechanisms, that is, as a response to the anxiety provoked by some perceived threat.[244] Looking at self-deception in this light will help us develop a more nuanced account of self-deception than what we have so far found in the accounts offered by Davidson and Gardner. With this purpose in mind, I will discuss the following five defensive reactions, considering their relation to self-deception: inhibition, symptom formation, undoing what has been done, isolation and repression. I believe that these defensive reactions are central to understanding self-deception, but I do not claim that this list is exhaustive.

One of the texts that I will study most in depth in this chapter is Freud's *Inhibitions, Symptoms and Anxiety*, In this text, Freud explores the central role of anxiety in causing symptoms, inhibitions (*Hemmung*) and other forms of

---

[243]Neither is Freud suggesting that this reaction of fending off what is not pleasurable is always possible. "The assumption that a defensive trend of this kind exists cannot be objected to on the ground that often enough one finds it impossible, on the contrary, to get rid of distressing memories that pursue one, and to banish distressing affective impulses like remorse and the pangs of conscience. For we are not asserting that this defensive trend is able to put itself into effect in every case, that in the interplay of psychical forces it may not come up against factors which, for other purposes, aim at the opposite effect and bring it about in spite of the defensive trend." (Freud, *Psychopathology of Everyday Life*, p. 147.)

[244] Hans Sjöbäck's study *Psykoanalysen som livslögnsteori: Läran om försvaret* (*Psycho-Analysis as a Theory of the Life-Lie: A Doctrine of Defense*) relates to my project. Sjöbäck studies literary examples of the life-lie, such as Hjalmar's life-lie in Ibsen's *The Wildduck*, in the light of Freudian psychoanalysis as well as exposes Nietzsche's influence on Freud's theories. (Hans Sjöbäck, *Psykoanalysen som livslögnsteori: Läran om försvaret* (Lund: Bo Cavefors Bokförlag, 1977.)

defense[245], and he discusses the relation between the normal and the pathological with regard to psychological reactions to perceptual influence.[246] Freud describes inhibition and symptom formation as both common reactions to, and protections from, anxiety. They restrict a person's actions and reflection in certain ways, and can be imposed to protect the person from a worry that arises from within or from the effect some outer event has upon him. I consider inhibition and symptom formation for two reasons. First, they give us examples of how a person can react in trying to handle anxiety. Second, this discussion is of interest since it raises the question of a distinction between normal and pathological. In discussing symptoms and inhibitions, Freud says:

> The two concepts are not upon the same plane. Inhibition has a special relation to function. It does not necessarily have a pathological implication. One can quite well call a normal restriction of a function an inhibition of it. A symptom, on the other hand, actually denotes the presence of some pathological process. Thus, an inhibition may be a symptom as well. Linguistic usage, then, employs the word *inhibition* when there is a simple lowering of a function, and *symptom* when a function has undergone some unusual change or when a new phenomenon has arisen out of it.[247]

Here Freud makes a distinction between normal and pathological: inhibition and symptom formation are both reactions to anxiety, but inhibition is portrayed as spanning the gap between what we call normal and what we call pathological, while symptom formation is always a pathological reaction. The difference is that inhibition is a lowering of the function of carrying out one's wish, a difference of degree, while symptom formation is *another*, merely substitutive, way in which to try to realize one's wish. A qualitative change has taken place in symptom formation; it is a sign of, and a substitute for, an instinctual drive that has been repressed and replaced.[248]

A symptom occurs when the repression of an instinct has, to greater or lesser extent, failed, and the repressed instinct seeks satisfaction in a substitutive obsessional action. What exactly is it that makes symptom formation something pathological? A part of the answer is that the neurotic action cannot fulfill the original desire: "when the substitutive impulse is carried out there is no

---

[245] Symptoms are not strictly a form of defense but rather the formation in which a defensive struggle results. It is a substitute for an action denied in repression. We will come back to this.

[246] Perceptual influence here stands for influence both from within and from without. I perceive my pain and sadness as well as what goes on in my surroundings.

[247] Freud, *Inhibitions, Symptoms and Anxiety*, p. 87. „Die beiden sind nicht auf dem nämlichen Boden erwachsen. Hemmung hat eine besondere Beziehung zur Funktion und bedeutet nicht notwendig etwas Pathologisches, man kann auch eine normale Einschränkung einer Funktion eine Hemmung derselben nennen. Symptom hingegen heisst soviel wie Anzeichen eines krankhaften Vorganges. Es kann also auch eine Hemmung ein Symptom sein. Der Sprachgebrauch verfährt dann so, dass er von Hemmung spricht, wo eine einfache Herabsetzung der Funktion vorliegt, von Symptom, wo es sich um eine ungewöhnliche Abänderung derselben oder um eine Leistung handelt." (GW, Band 14, s. 113.)

[248] Compare with *Inhibitions, Symptoms and Anxiety*, p. 91.

sensation of pleasure; its carrying out has, instead, the quality of a *compulsion*."[249] In other passages, Freud describes cases in which carrying out a substitutive impulse does provide satisfaction, but the fact that the impulse is substitutive nevertheless has implications for satisfaction, since a substitutive action cannot fulfill the original, real desire.[250] It can only replace the desire and the satisfaction of having it fulfilled with the satisfaction of performing a substitutive obsessional action. In obsessional neurosis, desires that the person reacts to as bad or forbidden, perhaps as taboos, are prevented from being fulfilled. Described in the terminology of the mental structure of id, ego and super-ego that Freud applies in *Inhibitions, Symptoms and Anxiety*,[251] it is the super-ego – the moral and regulating instance – which prevents these desires stemming from the id from being fulfilled or even acknowledged by the ego.[252]

---

[249] Ibid. p. 95 „Wenn er [der Ersatz] vollzogen wird, kommt keine Lustempfindung zustande, dafür hat dieser Vollzug den Charakter des Zwanges angenommen." (GW, Band 14, s. 122.)

[250] In one of Freud's examples, a neurotic man is obsessed with squeezing blackheads, which, according to Freud, gives him great satisfaction. On the basis of therapy and discussions with this man, Freud concludes that this symptom is a substitutive action for the desire to masturbate. In this case, although this substitutive obsessional action can be successful in preventing him from masturbating *and* it gives him some satisfaction, it cannot fulfill the real desire. (Freud, "The Unconscious" (1915), SE, vol. 14, pp. 199.)

[251] I have tried to leave out discussions of, and references to the psychoanalytic apparatus (for example, the divisions into id, ego and super-ego) of this study as much as possible but there are sections of highly relevant passages in which Freud's technical vocabulary cannot be avoided.

[252] The accounts of the psychical that Freud provides are manifold and under constant revision and development. I will attempt a short summary, based on what Freud writes in one of his later texts, *The Ego and the Id* (1923). The division of the psychical into what is conscious and what is unconscious, Freud says, is "the fundamental premise of psychoanalysis". The unconscious can be of two different kinds, latent unconscious and dynamic unconscious. The latent unconscious is capable of becoming conscious at any time; Freud comes to call it the preconscious. He reserves the term unconscious for the dynamic unconscious; that which is repressed. We thus arrive at the division into conscious, preconscious and unconscious. (pp. 13) According to Freud, these distinctions proved inadequate and insufficient, and thus he introduced the distinctions ego, id and superego. Regarding the ego he writes: "We have formed the idea that in each individual there is a coherent organization of mental processes; and we call this the ego. It is to this ego that consciousness is attached; the ego controls the approaches of motility – that is, to the discharge of excitations into the external world; it is the mental agency which supervises all its own constituent processes, and which goes to sleep at night, though even then it exercises the censorship on dreams. From this ego proceed repression, too, by means of which it is sought to exclude certain trends in the mind not merely from consciousness but also from other means of effectiveness and activity." (p. 17) There is also unconscious mental content in the ego, which is revealed by resistance. Freud says that there can be no question that resistance emanates from the ego and belongs to it. Resistance produces powerful effects without itself being conscious and it requires special work before it can be made conscious, just as the repressed mental content does. The antithesis of conscious and unconscious thus has to be substituted for another antithesis: "the antithesis between the coherent ego and the repressed which is split off from it" (p. 17). That which is split off is the id, which harbors the instincts (sexual and death instincts). (p. 40) The third mental structure is the ego-ideal, or the super-ego. Historically, it is a residue of the id's identification with the mother and the father and, at the same time, a reaction formation against these identifications: "Its relation to the ego is not exhausted by the precept 'You ought to be like this (like your father)', it also comprises the prohibition: 'You may not be like this (like your father) – that is, you may not do all that he does; some things are his prerogative.'" (p. 34)

In neurosis, the ego is under the control of the super-ego; Freud says that it is fully accessible to the influence of the super-ego but has shut out the wishes of the id by means of repression. The consequence is that what is conscious is only a distorted substitute of the id's wish, distorted either by being vague or so travestied as to be unrecognizable.

The vocabulary of id, ego and super-ego, with the ego standing between the other two as mediator as well as on the border between the outer and the inner world, can be read as a way of characterizing the struggle within the person or the self with respect to its own needs and values, but also with respect to the requirements of the world, be that struggle neurotic or normal. In the case of obsessional neurosis, the person's moral demands on himself are unusually strong and demanding and give the spontaneous desires and wishes little room to express themselves. In a summary of the tendency of symptom formation in obsessional neurosis, Freud asserts: "The over-acute conflict between id and super-ego which has dominated the illness from the very beginning may assume such extensive proportions that the ego, unable to carry out its office as mediator, can undertake nothing which is not drawn into the sphere of that conflict."[253] And, "The result of this process, which approximates more and more to a complete failure of the original purpose of defence, is an extremely restricted ego which is reduced to seeking satisfaction in the symptoms."[254] The demanding moral requirements of the neurotic subject allow desire to be fulfilled only in the highly compromised form of a substitutive action. Rather than being defended against the instinctual demands of the id, the ego suffers under the strong moral requirements of the super-ego.

We have seen that, in inhibition, satisfaction of certain demands of the id is reduced. This has the consequence that the ego is restricted in certain ways. Freud gives an example of inhibitions in professional activities where the super-ego acts punitively and forbids the ego (the person) to engage in activities where he can attain success. In symptom formation, the function of the ego in promoting the wish (of the id) is not just reduced but the wish is distorted, and satisfaction can only be obtained in a compromised form in the substitutive obsessional act. In this chapter, we have seen Freud make a distinction between acting out a symptom, which is pathological by its very nature, and inhibition, which can be pathological in extreme cases but need not be. An important difference is that the symptom is a substitutive action that can take the place of

---

Moreover, Freud remarks: "It is easy to show that the ego ideal answers to everything that is expected of the higher nature of man". (p. 37) (Freud, *The Ego and the Id*, SE, vol. 19.)

[253] Freud, *Inhibitions, Symptoms and Anxiety*, p. 118. „Der überscharfe Konflikt zwischen Es und Über-Ich, der die Affektion von Anfang an beherrscht, kann sich so sehr ausbreiten, dass keine der Verrichtungen des zur Vermittlung unfähigen Ichs der Einbeziehung in diesen Konflikt entgehen kann. (GW, Band 14, s. 148.)

[254] Ibid. „Ein äusserst eingeschränktes Ich, das darauf angewiesen ist, seine Befriedigungen in den Symptomen zu suchen, wird das Ergebnis dieses Prozesses, der sich immer mehr dem völligen Fehlschlagens des anfänglichen Abwehrstrebens nähert." (Ibid.)

the desired and repressed action without at all revealing the desired action, which it replaces, to the subject.

Having noted Freud's distinction between inhibition and symptom formation, it is of greater importance in trying to understand self-deception to pay attention to what Freud describes as something that they share: they stand in the way of carrying out one's wishes. They compromise the enjoyment that is to be had from having one's desire fulfilled. These reactions are caused by anxiety, and, as can be seen from Freud's analysis, anxiety arises because there is a conflict between conceptions and preconceptions of what is right, suitable, moral etc. (demands from the super-ego, the moral and regulating instance) and the wish to carry out the desire. It is the pressure caused by these contrary "forces" that causes anxiety and, in the end, also inhibition or symptom formation. Here we can see a close parallel between many cases of self-deception and inhibition/symptom formation. The case of Anna Karenina is an excellent example. The tension between her strong desire for Vronsky, on the one hand, and her self-perception as a married woman and a mother, with all the expectations on herself, sense of responsibility, pride, meaningfulness, security etc. that comes with it, on the other, puts strong pressure on her and causes anxiety. While Gardner sees Anna's self-deception as a means for fulfilling her desire, I want to look at it from this angle: as a defense against the anxiety which arises under the pressure of two contradictory desires, or, in the case of Anna Karenina, two possible ways of life. Her self-deception is an inhibition of her desire to be with Vronsky. Vronsky's reaction shows this too. When Anna finally admits to her love for him, he is relieved, since this acceptance, or understanding on Anna's part, gives him hope that they might have a future together. The analogy with a formation of a substitute can also reveal Anna's reaction. When she thinks of Vronsky she doesn't think of him as her loved one, or of his feeling for her as love. Although that is what she desires, this desire is in conflict with her commitments and desires as a wife and mother. Under the pressure from both, her "compromise" is to think of her feelings for Vronsky as something other than love, and when he expresses his love for her it upsets her. One could say that when she thinks of their relationship, she can only accept it in terms such as friendship, perhaps attraction etc., all these descriptions being "substitutes" for love which "diminishes" the feeling but which also can co-exist with the idea of the person she takes herself to be and wants to be, what we could call moral requirements.

I have discussed symptom formation and inhibition as responses to desires and as reactions to anxiety. Symptom formation is, as we have seen, a sign that repression is occurring. It also shows that repression isn't completely successful; a substitutive action needs to be carried out for the repression of the wish to be maintained and for the subject to remain in control of the anxiety. Thus, repression arises as a reaction to anxiety. Throughout Freud's works, repression

stands in relation to anxiety. In his early works, Freud holds that anxiety is a *consequence* of repression.[255] In *Inhibitions, Symptoms and Anxiety*, Freud reverses the picture and holds that anxiety is one of the chief motivating forces leading to repression. Further, Freud changed his mind many times as to the *meaning* of repression: as including all the defenses, as being a special form of defense, etc.[256] For my purposes, it is not important how or if the different forms of defense can be clearly distinguished from each other. Rather, I consider what Freud says about certain forms of defense because it sheds light on the different ways in which a person can avoid experiencing the anxiety provoked by an event, a thought, a feeling etc.

Repression is perhaps best characterized by looking at how it arose as a term in psychoanalytic theory. The phenomenon of *resistance* in clinical practice gave rise to the concept. Patients are often resistant towards admitting to, or talking about, attitudes or deeds that they find degrading, embarrassing, threatening, etc. Freud discovered that resistance could be used as an "instrument" in clinical practice because it was an indicator that the analysand had found something in analysis troubling during analysis, something that he could not or did not want to recognize, and in some cases, about something that had previously been *repressed*.[257] The resistance that is revealed in analysis is a re-awakening of the resistance that had once caused repression, according to Freud. He writes: "One of the vicissitudes an instinctual impulse may undergo is to meet with resistances which seek to make it inoperative. Under certain conditions […] the impulse then passes into the state of 'repression' [*Verdrängung*]."[258] In accounting for the form of avoidance that repression is, and in contrasting it with other forms, Freud says:

> If what was in question was the operation of an external stimulus, the appropriate method to adopt would obviously be flight; with an instinct, flight is of no avail, for the ego cannot escape from itself. At some later period, rejection based on judgment (*condemnation*) will be found to be a good method

[255] Freud, "Repression" (1915), SE, vol. 14, p. 153 and p. 155. This is also discussed in the Editor's Note on "Repression".

[256] In Freud's early text *Studies in Hysteria* (1893-95), 'defense' is the word used to describe the process of repression. Later (in "Repression"), 'repression' is used generally and replaces 'defense', while in "Inhibitions, Symptoms and Anxiety" Freud explores the different mechanisms and restricts the term 'repression' to one particular mechanism, while reviving 'defense' as "a general designation for all the techniques which the ego makes use of in conflicts which might lead to neurosis" ("Repression", vol. 14, Editor's Note, pp. 144; Freud, *Inhibitions, Symptoms and Anxiety*, vol 20, pp. 163.)

[257] See, Freud, *Introductory Lectures on Psycho-Analysis*, chapter XIX: "Resistance and Repression", SE, vol. 16. (p. 294)

[258] Freud, "Repression", p. 146. „Es kann das Schicksal einer Triebregung werden, dass sie auf Widerstände stösst, welche sie unwirksam machen wollen. Unter Bedingungen […] gelangt sie dann in den Zustand der *Verdrängung*." (GW, Band 10, s. 248.)

> to adopt against an instinctual impulse. *Repression is a preliminary stage of condemnation, something between flight and condemnation*.[259]

Repression is not based on judgment, it is a more primitive reaction: "*the essence of repression lies simply in turning something away, and keeping it at a distance, from the conscious."*[260] As we have seen, judgment plays a central role in Davidson's and Gardner's accounts of self-deception. Carlos judges that, all things considered it is better to avoid anxiety, therefore he deceives himself. Anna Karenina judges what means are required to allow her and Vronsky's love to grow. I want to suggest that self-deception ought not to be conceived of as a rational, goal-directed action (involving judgment) at all.

Before concluding the discussion on repression and anxiety, I will look at two other forms of defense, which Freud calls "surrogates" for repression. He thinks that these are well suited to illustrate the purpose and technique of repression. One of these surrogates is to *undo what has been done*. I bring this form of defense into the discussion because it provides examples of avoidance that is *motivated* but not *intentional*. Further, these examples can help illuminate the tendency to over-rationalize an act in trying to understand or explain it. The pattern of undoing what has been done is common in cases of obsessional neurosis. In the first chapter, we discussed an extract from "Notes upon a Case of Obsessional Neurosis" where Freud tells of a patient who, when he stumbled on a branch that lye on the path in the park, threw it into the hedge. Later, on his way home, where

> he was suddenly seized with uneasiness that the branch in its new position might perhaps be projecting a little from the hedge and might cause an injury to someone passing by the same place after him. He was obliged to jump off his tram, hurry back to the park, find the place again, and put the branch back in its former position – although any one else but the patient would have seen that, on the contrary, it was bound to be more dangerous to passers-by in its original position than where he had put it in the hedge.[261]

---

[259] Ibid. My italics. Upon this citation follows: "it is a concept which could not have been formulated before the time of psycho-analytic studies." „Handelte es sich um die Wirkung eines äusseren Reizes, so wäre offenbar die Flucht das geeignete Mittel. Im Falle des Triebes kann die Flucht nichts nützen, dann das Ich kann sich nicht selbst entfliehen. Später einmal wird in der Urteilsverwerfung (Verurteilung) ein gutes Mittel gegen die Triebregung gefunden werden. Eine Vorstufe der Verurteilung, ein Mittelding zwischen Flucht und Verurteilung ist die Verdrängung […]" (Ibid.)

[260] Ibid. p. 147. „[I]hr *Wesen nur in der Abweisung und Fernhaltung vom Bewussten besteht."* (Ibid. s. 250.)

[261] Freud, "Notes upon a Case of Obsessional Neurosis (The Rat Man)" (1909), SE, vol. 10, p. 192, n. 2. „[Ü]berkam ihm plötzlich die Sorge, in der neuen Lage könnte der jetzt vielleicht etwas vorragende Ast zum Anlasse eines Unfalles für jemand werden, der nach ihm an derselben Stelle vorbeigehe. Er musste von der Trambahn abspringen, in den Park zurückeilen, die Stelle aufsuchen und den Ast in die frühere Lage zurückbringen, obwohl es jedem anderen als dem Kranken einleuchten würde, dass die frühere Lage doch für einen Passanten gefährlicher sein müsste als die neue im Gebüsche." (GW, Band 7, ss. 414, Anmerkung 3.)

Freud continues: "The second and hostile act, which he carried out under compulsion, had clothed itself to his conscious view with the motives that really belonged to the first and philanthropic one."[262] Freud seems to be suggesting that there is no reason or intention in the second act. It is rather a compulsion, but the neurotic man believes that there is an intention: to prevent passers-by from getting hurt. As we have seen, Davidson argues that this irrational behavior can be made sense of by showing that it consists of different elements which, in themselves, have rational structures. But as we saw in the first chapter, this rational reproduction of the man's irrational action has been called into question.[263] I want to show that this critique has support in Freud. The case of obsessional neurosis above is discussed in a footnote in Freud's case study on the Rat Man, which Freud considers to be a parallel case. Let us see how Freud understands these cases of obsessional neuroses.

The Rat Man has conflicting feelings towards his girlfriend. His feelings manifest themselves in obsessional behavior that is acted out when she is on her way to visit her sickly grandmother and he is unhappy about her leaving.

> On the day of her departure he knocked his foot against a stone lying on the road, and was *obliged* to put it out of the way by the side of the road, because the idea struck him that her carriage would be driving along the same road in a few hours' time and might come to grief against the stone. But a few minutes later it occurred to him that it was absurd, and he was *obliged* to go back and replace the stone in its original position in the middle of the road.[264]

Before the day of his girlfriend's departure, during their holidays together, the Rat Man had interpreted an utterance by her to be an expression of a wish not to be with him. When they talked it over, she had been able to convince him that he had misinterpreted her. Since that time, the Rat Man had an obsession with understanding, and a fear of misunderstanding. Freud interprets this obsession with understanding as an expression of the Rat Man's lingering doubt as to whether or not he had understood his girlfriend correctly, and if he had good reason to take her expressions of affection for him as being genuine. He doubted her love. Freud says:

> A battle between love and hate was raging in the lover's breast, and the object of both these feelings was one and the same person. The battle was represented in a

---

[262] Ibid. p. 193. „Die zweite Feindselige Handlung, die sich als Zwang durchsetzte, hatte sich vor dem bewussten Denken mit der Motivierung der ersten, menschenfreundlichen, geschmückt." (Ibid.)

[263] By, for example, Jonathan Lear, *Freud*, p. 26-30, see Chapter One, p. 50.

[264] Freud, "Notes upon a Case of Obsessional Neurosis (The Rat Man), p. 190. „Am Tage, als sie abreiste, stiess er mit dem Fusse gegen einen auf der Strasse liegenden Stein und musste ihn nun auf die Seite räumen, weil ihm die Idee kam, in einigen Stunden werde ihr Wagen auf derselben Strasse fahren und vielleicht an diesem Stein zu Schaden kommen, aber einige Minuten später fiel ihm ein, das sei doch ein Unsinn, und er musste nun zurückgehen und den Stein wieder an seine frühere Stelle mitten auf der Strasse legen." (GW, Band 7, s. 412.)

> plastic form by his compulsive and symbolic act of removing the stone from the road along which she was to drive, and then of undoing this deed of love by replacing the stone where it has lain, so that her carriage might come to grief against it and she herself be hurt.[265]

Freud elucidates: "We shall not be performing a correct judgement of this second part of the compulsive act if we take it at its face value as having merely been a critical repudiation of a pathological action. The fact that it was accompanied by a sense of compulsion betrays it as having itself been a part of the pathological action, though a part which was determined by a motive contrary to that which produced the first part."[266] Of the two cases considered, the man who places and replaces the branch (earlier referred to as Mr. S) and the Rat Man, Freud says:

> Compulsive acts like these, in two successive stages, of which the second neutralizes the first, are a typical occurrence in obsessional neuroses. The patient's consciousness naturally misunderstands them and puts forward a set of secondary motives to account for them – *rationalizes* them, in short. [...] But their true significance lies in their being a representation of a conflict between two opposing impulses of approximately equal strength: and hitherto I have invariably found that this opposition has been one between love and hate.[267]

These acts are not rational acts, but compulsive. When one gives an explanation of them according to which the person has justifications for each step of the action, one imposes a rational structure on a behavior that is not itself rational. The compulsive acts reveal opposing feelings. In the quote above, Freud says that the patient himself tends to rationalize his behavior. He remarks: "naturally an attempt is made to establish some sort of logical connection (often in defiance of all logic) between the antagonists."[268] In Chapter One, I accused

---

[265] Ibid. p. 191. „Es tobt in unserem Verliebten ein Kampf zwischen Liebe und Hass, die der gleichen Person gelten, und dieser Kampf wird plastisch dargestellt in der zwanghaften, auch symbolisch bedeutsamen Handlung, den Stein von dem Wege, den sie befahren soll, wegzuräumen und dann diese Liebestat wieder rückgängig zu machen, den Stein wieder hinzulegen, wo er lag, damit ihr Wagen an ihm scheitere und sie zu Schaden komme." (Ibid. s. 414.)

[266] Ibid. pp. 191. „Wir verstehen diesen zweiten Teil der Zwangshandlung nicht richtig, wenn wir ihn nur als kritische Abwendung vom Krankhaften Tun auffassen, wofür er sich selbst ausgeben möchte. Dass auch er sich unter der Empfindung des Zwanges vollzieht, verrät, dass er selbst ein Stück des krankhaften Tuns ist, welches aber von dem Gegensatz zum Motiv des ersten Stückes bedingt wird." (Ibid.)

[267] Ibid. p. 192. „Solche zweizeitige Zwangshandlungen, deren erstes Tempo vom zweiten aufgehoben wird, sind ein typisches Vorkommnis bei der Zwangsneurose. Sie werden vom bewussten Denken des Kranken natürlich missverstanden und mit einer sekundären Motivierung versehen – *rationalisiert*. Ihre wirkliche Bedeutung liegt aber in der Darstellung des Konfliktes zweier annähernd gleich grosser gegensätzlicher Regungen, soviel ich bisher erfahren konnte, stets des Gegensatzes von Liebe und Hass." (Ibid.)

[268] Ibid. „[N]atürlich nicht ohne dass der Versuch gemacht würde, zwischen den beiden einander feindseligen eine Art von logischer Verknüpfung – oft mit Beugung aller Logik – herzustellen." (Ibid.)

Davidson of "over-rationalization" in his account of irrational action, with regards to both obsessional neurotic behavior and to forms of irrationality such as self-deception. Davidson's explanation of the compulsive act suggested that it can, and should, be understood as consisting of justifiable parts. As Freud points out above, this tendency of "over-rationalization" is typical for someone trying to understand his own behavior as well.

The examples above show the pattern of undoing what has been done, which is a pattern that runs through cases of obsessional neuroses. To take the first case as an example, the neurotic first acts by removing the branch from the path and putting it in the bush, with the aim of preventing passers-by from stumbling over it and hurting themselves. Later he undoes, or neutralizes, this action by putting the branch back in its original position on the path. His intervention in the world is undone by this second action. The neurotic, according to Freud, is typically motivated to act by opposing impulses of love or hate, and indeed it seems impossible to find a sensible reason for moving the branch back to the path as long as his aim is to prevent passers-by from being harmed. The aim in this action is to undo what has been done, as Freud writes: "to 'blow away' not merely the *consequences* of some event but the event itself."[269] Freud says that the "endeavor to undo shades off into normal behavior in the case in which a person decides to regard an event as not having happened."[270] What separates what Freud considers to be normal behavior from what he calls pathological behavior seems to be that while in the normal case one *decides* not to pay further attention to the event or its consequences, the obsessional neurotic will *take action* to deny the event and make it non-existent by undoing it. While the first case is a decision regarding what attitude to take towards something, the second is an attempt to deny the fact of the event.

In "The Loss of Reality in Neurosis and Psychosis", Freud makes a distinction between neurosis and psychosis: "in neurosis a piece of reality is avoided by a sort of flight, whereas in psychosis it is remodeled [...] Neurosis does not disavow the reality, it only ignores it; psychosis disavows it and tries to replace it."[271] It is typical of neurosis that the part of reality that one doesn't want to recognize is avoided by flight, while in psychosis the flight is followed

---

[269] "Undoing what has been done […] is, as it were, negative magic, and endeavors, by means of motor symbolism, to 'blow away' not merely the *consequences* of some event (or experience or impression) but the event itself. I choose the term 'blow away' advisedly, so as to remind the reader of the part played by this technique not only in neuroses but in magical acts, popular customs and religious ceremonies as well." (Freud, *Inhibitions, Symptoms and Anxiety*, p. 119.) „Sie ist sozusagen negative Magie, sie will durch motorische Symbolik nicht die Folgen eines Ereignisses (Eindruckes, Erlebnisses), sondern dieses selbst ‚wegblasen'." (GW, Band. 14, s. 149.)

[270] Ibid. p. 120. „Seine Abschattung zum Normalen findet das Streben zum Ungeschehenmachen in dem Entschluss ein Ereignis als ‚*non arrive*'." (Ibid. s. 150.)

[271] Freud, "The Loss of Reality in Neurosis and Psychosis" (1924), SE, vol. 19, p. 185. "[B]ei der Psychose folgt auf die anfängliche Flucht eine aktive Phase des Umbaues, bei der Neurose auf den anfänglichen Gehorsam ein nachträglicher Fluchtversuch. [...] Die Neurose verleugnet die Realität nicht, sie will nur nichts von ihr wissen; die Psychose verleugnet sie und sucht sie zu ersetzen." (GW, Band 13, s. 365.)

by an active phase of replacing.[272] I will consider this distinction between defenses against taking in reality, but first I wish to note a complication: the undoing that occurs in obsessional neurosis seems to be a case of replacement of the kind that Freud ascribes to psychosis in the text "The Loss of Reality in Neurosis and Psychosis". As we have seen in "Notes upon a Case of Obsessional Neurosis", the neurotic's obsessional behavior consists in replacing an earlier action. Freud makes clear that no sharp distinction between neurosis and psychosis should be made.[273] In any case, the exact relation between neurosis and psychosis is of little importance to this investigation of self-deception. Rather, the discussion is of interest to the extent that it introduces an important contrast between 1) deciding to treat something as not having happened; 2) not wanting to know and avoiding knowledge; 3) distorting that which one doesn't want to know by replacing it with something else.

What is the difference between the normal and the pathological here? There seems to be three steps where neurosis bridges the gap between the normal and the pathological. According to Freud, in the normal case of undoing, one makes a *decision* in full view of the situation to regard an event as not having happened, or rather, to disregard that it has happened, i.e. *to treat it as if* it never happened. If we consider neurosis as described in "The Loss of Reality in Neurosis and Psychosis", there is, I believe, an important shift between the normal case of making a decision and the description of the neurotic as "not wanting to know about reality" or "ignoring it". In the latter case, there is no suggestion of a decision being made to regard something as not having happened; the neurotic does not know that which she doesn't want to know. The way in which the neurotic misleads herself is not by lying to herself about something that she knows, but by avoiding rather than seeking the truth and, perhaps, by shying away from situations, reflections, and comments which would make it difficult for her to avoid recognizing the fact of the matter. Neither does the psychotic have a full view of the situation; the distortion in psychosis (or in obsessional neurosis, which, as we noted earlier, is not entirely distinct) is not a distortion of something the sense of which one knows well.

The other surrogate for repression that Freud discusses, and the last of the defense mechanisms that I will present here, is the technique of *isolating*. Freud says that isolation has an effect similar to that which an amnesiac has in repression, except that in the case of isolation, nothing is forgotten, but elements that belong together are held apart.[274] Here, as in the case of undoing, Freud accounts for different cases along the spectrum of isolation. He points to

---

[272] Ibid.

[273] "A neurosis usually contents itself with avoiding the piece of reality in question and protecting itself against coming into contact with it. The sharp distinction between neurosis and psychosis, however, is weakened by the circumstance that in neurosis, too, there is no lack of attempts to replace a disagreeable reality by one which is more in keeping with the subject's wishes." (Freud, "The Loss of Reality in Neurosis and Psychosis", p. 187.)

[274] Freud, *Inhibitions, Symptoms and Anxiety*, pp. 121.

concentration as a normal phenomenon that has the same form as the neurotic procedure of isolation. When we concentrate, "what seems to us important in the way of an impression or a piece of work must not be interfered with by the simultaneous claims of any other mental process or activities."[275] Freud adds that concentration is not purely a matter of focus but also an effort to neglect the contradictory: "even a normal person uses concentration to keep away not only what is irrelevant or unimportant, but, above all, what is unsuitable because it is contradictory."[276] In obsessive behavior, it can take the form of disrupting a sequence: "when a neurotic isolates an impression or an activity by interpolating an interval, he is letting it be understood symbolically that he will not allow his thoughts about that impression or activity to come into associative contact with other thoughts."[277] We see that isolation – the phenomenon that can have the advantageous aim and consequence of successfully and effectively getting work done – can also have a negative and pathological aim or consequence. It can, for example, prevent one from understanding one's own behavior, as in the neurotic's case above, or, from recalling an event as a traumatic experience, since, when only scattered pieces are recalled, the feeling of danger or fright is not perceived. In this case, the motivation for the defensive reaction of isolation would be to avoid recalling, and so re-experiencing, the traumatic incident.

Isolation, like the other defensive reactions that I have considered, is a way of handling pressure and demands coming from different directions. These demands can run the gamut from the most ordinary circumstances, such as intentionally shutting something out which calls for one's attention in order to focus on something else which one considers more important or urgent, to pathological cases, such as when the Rat Man compulsively interjects a "not" in his prayer and thus disrupts, or "attacks", his own activity of praying.[278] In the latter case, isolation refers to something which is not an intentional action but compulsive behavior. I think that isolation can capture well what goes on in a number of cases of self-deception, such as the case of Anna Karenina. She is aware of the happiness she feels when she meets Vronsky, of having changed social circles although she really preferred the first, of Vronsky's interest in her,

---

[275] Ibid. p. 121. „Was uns bedeutsam als Eindruck, als Aufgabe erscheint, soll nicht durch die gleichzeitigen Ansprüche anderer Denkverrichtungen oder Tätigkeiten gestört werden." (GW, Band 14, s. 151.)

[276] Ibid. „Aber schon im Normalen wird die Konzentration dazu verwendet, nicht nur das Gleichgültige, nicht Dazugehörige, sondern vor allem das unpassende Gegensätzliche fernzuhalten." (Ibid.)

[277] Ibid. p.122. „[W]enn der Neurotiker auch einen Eindruck oder eine Tätigkeit durch eine Pause isoliert, gibt er uns symbolisch zu verstehen, dass er die Gedanken an sie nicht in assoziative Berührung mit anderen kommen lassen will." (Ibid. s. 152.) Freud says that isolation involves the "taboo on touching" and describes isolation as "removing the possibility of contact" and as a "method of withdrawing a thing from being touched in any way".

[278] Lear discusses this case in *Freud*, pp. 42.

etc., but she prevents herself from reflecting on what her emotions and reactions might mean. This may be understood in terms of isolation.

I have introduced five kinds of defensive reactions that Freud discusses in the essay *Inhibitions, Symptoms and Anxiety* in order to consider what light Freud's characterization of the different defenses, and what their relation to each other can shed on our discussion on self-deception. I believe that we can either see them as present in, or as analogous to, self-deception, and that they therefore can help bring certain elements into sharper relief. In the following section, I will discuss these defensive reactions in the context of illusion and misapprehension.

Flights from Anxiety

In our study of Freud's examples of illusion at the beginning of the present chapter, we noted that to be subject to an illusion is different from being in error regarding something in that the illusion is guided by what the person wishes to see and believe. This suggests that someone who is subject to an illusion is biased in his apprehension, understanding and reasoning: he acknowledges only that which fits in with his preferred ideas, or self-understanding, and he "ignores" that which falls outside of how he wants things to be. Or perhaps better put, he does not want to know reality when it does not conform to what he wants to be the case. "Not wanting to know" appears to be a rather passive attitude. But it might prove difficult to guard oneself against one's own critical suspicions and nearly impossible to evade questions and critical remarks made by others. What remains in one's power under circumstances where questions, doubts, or facts incoherent with one's beliefs are forced upon one is the capacity to reform or re-figure these facts, questions or doubts so as to defuse their critical potential. Consider the example of Carlos. It is possible that Carlos does not even grasp the implicit criticism in the instructor's comments, although it is obvious. But assuming that Carlos grasps the critical implication, he can nonetheless reject it as meaning what it does, i.e., as expressing doubt that he will pass the driver's test. Instead, he might find alternative interpretations: the teacher is cautious because he is afraid of giving guarantees; the teacher is a pessimistic man who refrains from encouraging students to register for taking the driver's test; the teacher dislikes Carlos, etc. Or Carlos can take the teacher's doubts seriously, but convince himself that he can make up for his weaknesses in two days and manage to pass the test anyway. The veil of ignorance is pierced by the instructor's comments, and what shines through must be covered or muted in Carlos' understanding of it. The self-deceiver can protect himself from questions and criticism by rationalizations.

Compare how the self-deceiver reacts with rationalizations when something that he does not want to recognize confronts him, and how symptoms arise when repression has not quite succeeded. We could say that symptoms arise

when there is a tear in repression: the substitutive actions that make up the symptoms arise because the action that one originally wished to perform, but which was deemed "unacceptable", and therefore repressed, hasn't quite been silenced by the repression. The wish requires some form of satisfaction, but it is only allowed satisfaction in a masked form, i.e., through the substitute that does not cause anxiety. In the case of symptom formation, the arousal of the repressed wish makes the person anxious in the same way it did when it was first repressed; he becomes anxious because this is a wish he doesn't want to acknowledge, perhaps because it doesn't fit with the moral requirements he places on himself. There is an analogy here with the case of Carlos; the teacher's comments express something that threatens to rupture Carlos's self-understanding and, since this makes him anxious, he reacts by ignoring the critique or defusing it. Moreover, Davidson writes that "the thought of failing at anything is particularly galling to Carlos"[279], which suggests that the threat that Carlos experiences at this point is a re-awakening of a well-known and dreaded feeling.

This can be expressed in the psychoanalytic vocabulary of id, ego and superego. According to Freud, acting out symptoms is a reaction to the conflict between the instinctual requirements of the id and the requirements of the super-ego (the ego-ideal) that one should be moral, strong, successful, etc. The ego, under the influence of the super-ego, represses wishes which are deemed unacceptable and, when the wishes are not held in check, the conflict gives rise to symptom formation. Carlos's self-deception and the rationalizations that keep it going can be understood as resulting from the pressure of the ego under influence of the super-ego, which does not want to acknowledge any aberration from the idealized self-image. The instructor's comments conflict with Carlos's self-conception and are therefore re-interpreted or re-formed in self-deception so as to make them appear not to be. Self-deception and symptom formation can both be seen as ways of undoing what has been done. In the case of the neurotic, by performing the substitutive action, the wish to perform the original action can be covered up. In the case of Carlos' self-deception, the threatening character of the instructor's comments is concealed or "undone" by Carlos through rationalizations.

In Freud's discussion of the "surrogates of repression" – isolation and undoing – he brings in "normal" cases of isolation and undoing as well as pathological cases. According to Freud, as we have seen, a normal case of undoing is to decide to treat something as not having happened. In Freud's characterization of neurosis and psychosis, a distinction is drawn between not wanting to know (neurosis) and distorting that which one doesn't want to know by replacing it (psychosis). I find Freud's characterization of the normal

[279] Davidson, "Deception and Division" in *Problems of Rationality*, p. 209.

case problematic.[280] This tripartite distinction is nevertheless of interest for the discussion of self-deception, since it helps to clarify a difference between Davidson's account, on the one hand, and my suggestion as to how we should understand the problem, on the other. In Davidson's picture, self-deception is a case of deciding to treat something that one knows is true as if it weren't true. Self-deception starts with something that the self-deceiver takes to be true, i.e. with a belief that is the target of self-deception: that which the self-deceptive intention aims to bury. In contrast, in Freud's work I have found the suggestion that self-deception is characterized not by a conscious belief, but rather by an avoidance of becoming conscious, that is, of forming a belief. In line with Freud's work, I argue that self-deception ought rather be characterized by the attitude of not wanting to know, together with the struggle to maintain this condition of not knowing, or, alternatively, of "reforming" a potentially threatening reality (including thoughts, impressions, perceptions) to one that conforms with one's desires and wishes.

While the example of Carlos shows the perseverance of his wish that he will pass the test and prevents potential criticism from shattering that hope, Anna Karenina's self-deception can be seen as a prophylactic device to prevent her desire to be with Vronsky from becoming conscious. In the case of Carlos, the wish to pass the test is part of the greater wish to see himself as someone who does not fail. In the case of Anna Karenina, the motivation to keep the desire to be with Vronsky unnoticed, we can imagine, is to maintain her self-image as Karenin's wife and a good mother to their child, as well as to avoid choosing between Vronsky and her son. The desire for Vronsky, which she has not yet made explicit to herself, contains a potential threat to Anna Karenina's ego – to her ideas about herself – which can be kept intact as long as the desire does not take the form of a belief, that is, is not made conscious. Keeping a forbidden wish unconscious is a central element in obsessional neurosis. In the symptom formation that is typical of obsessional neurosis, the action that would satisfy one's wish is replaced by a distortion of it, a substitute. One can understand Anna Karenina's self-deception in an analogous way; as resistance to becoming aware of her wish to see Vronsky *as* a wish to see the man she is about to fall in love with. This resistance involves her finding other ways of understanding her feelings of delight when meeting him. She might think of him as a very good friend, or she might admit to being attracted to him but not to the strength and

[280] Freud says that when undoing shades off into normal behavior, it is a *decision* to treat something as not having happened. This seems to be too rational a rendering even of a "normal" case of ignorance, or undoing. Even in the case where we are conscious of what has happened or what it is that we have done and do not face up to it or take responsibility for it, it seems doubtful that this "ignorance" is best described as the result of a decision. There are cases of self-deception that include ignoring something which one is somehow aware of having done, but I believe, and will later argue, that in these cases ignorance is rather a defensive response and not a decision. I will return to Freud's inclination to make this type of rationalization later in this chapter. (Niklas Forsberg has made me aware of this problem in Freud's characterization of the normal case of undoing.)

quality of her feelings. In these cases, she finds substitute descriptions (understandings) that do not capture what he really means to her, but which offer escape from the anxiety that the belief that she is in love with Vronsky would provoke. We have seen that Anna Karenina does not become conscious of her love for Vronsky because she does not acknowledge what she feels. Isolation can partly account for how repression is achieved and upheld, since isolation prevents her from reflecting upon what her emotions and reactions might mean.

A central characteristic of self-deception is thus that it is a flight from anxiety, and different defense mechanisms, such as the ones discussed above, play important roles in this escape. Gardner distinguished self-deception from deep forms of irrationality, it will be recalled, in part on the grounds that neurosis etc. includes repression but self-deception does not. I object to this distinction; repression is common in self-deception insofar as the self-deceiver typically avoids awareness of that about which she deceives herself.[281] This "neurotic" element is present in many cases of self-deception. Anna Karenina, for example, is deceived in not understanding her feelings for Vronsky as love; the acknowledgment of her desire is repressed. The different defenses serve the role of protecting the ego, for example, by suppressing memories of a traumatic experience. But this protection also includes the maintenance of one's self-image through a variety of defense mechanisms: the self-deceiver neglects to notice self-contradictions and implicit or explicit criticisms, for instance. Or she "reforms" the facts of the matter to buttress the self-image she has.

The discussion of misapprehensions associated with illusion shows that one is prone to apprehend that which one wishes to be the case. In these cases, the person is directed toward and motivated by that which she wants to believe. Repression and other forms of defense are motivated by the same thing, but with emphasis on avoiding remembering or understanding something, or avoiding understanding that causes anxiety. I have tried to show that defensive reactions to anxiety and threats play a decisive role in what we call self-deception. The tendency to hold on to what one wants to be the case and to avoid that which provokes anxiety both find expression in what Freud calls the *Pleasure Principle*: we tend to believe or do what gives most pleasure and we steer clear of whatever signals displeasure.

These tendencies can overlap. Freud's analysis of the Rat Man, for instance, shows that he knows that the reason why he was removing the stone from the road was that he wanted to protect the lady who would soon be passing in her carriage. But the Rat Man has no idea why he "had to" replace it. His love for the lady in part allows him to repress his conflicting feelings of anger and hate. Freud says of the Rat Man's compulsive acts: "The love had not succeeded in extinguishing the hatred but only in driving it down into the unconscious; and in the unconscious the hatred, safe from the danger of being destroyed by the

[281] I will discuss Freud's exploration of repression in more detail later in this chapter.

operations of consciousness, is able to persist and even to grow."[282] We can consider the cases of Carlos and Anna Karenina, respectively, in analogous ways. By holding on to the belief that he will pass the test, Carlos succeeds in repressing the anxiety that is evoked by sensing that he is not the ambitious and talented person whom he wants and expects himself to be. Similarly, Anna Karenina seriously believes that she is displeased with Vronsky for pursuing her, a belief which is appropriate considering that she is married. Later she is forced to realize that she desires to be with him. Thus, self-deception can be understood as resulting from an unwillingness to allow one's preconceived ideas to be disturbed combined with an active attempt to repulse any idea, thought, memory, feeling, or perception that may threaten the security which they provide. Freud's writings on pathology provide vast material for understanding avoidance, such as the avoidance of anxiety-provoking situations, the avoidance of coming to awareness (of forming a belief), etc.

Let me now summarize my argument up to this point. I have discussed three points that ought to be considered with regard to the phenomenon of self-deception. First, self-deception ought not be sharply distinguished from the phenomena captured by descriptions such as 'delusion' (*Täuschung*, *Wahnidéen*) and 'illusion'. Second, the difference between self-deception and what Gardner calls "deeper forms of irrationality" is largely one of degree rather than of kind, thus a comparison of self-deception and neurotic symptom formation can be instructive. Third, self-deception is a defensive reaction and is to be seen as a flight from anxiety.[283] I will now turn to Freud's use of the term 'unconscious intention'. In previous chapters, I have criticized Davidson's and Gardner's rendering of self-deception as intentional. In my view, nonetheless, there is something to be gained from a consideration of how it might be meaningful to speak of self-deception as an intentional action.

What is an Unconscious Intention?

We have seen how Freud characterizes the relation between the normal and the pathological as mainly one of degree rather than kind.[284] In his discussion of isolation and undoing, Freud also attends to how these can function advantageously in normal cases. In the case of deciding to treat an event as if it has not happened, the undoing is voluntary, as is the isolation in concentration.

---

[282] Freud, "Notes upon a Case of Obsessional Neurosis ("The Rat Man")", p. 239. „Die Liebe hat den Hass nicht auslöschen, sondern nur ins Unbewusste drängen können, und in Unbewussten kann er, gegen die Aufhebung durch die Bewusstseinswirkung geschützt, sich erhalten und selbst wachsen." (GW, Band 7, s. 455.)

[283] My ambition here has been to understand self-deception through texts where Freud, in great parts, writes on pathology. Clearly it is possible to distinguish more carefully than I do here ways in which self-deception *differ* from neurosis, psychosis, and other forms of pathology.

[284] There are some exceptions, such as taking symptom formation to always be pathological.

If we consider a case of undoing such as the obsessional neurotic Mr. S's removing and replacing of the branch, it is clear that he is repeating a pattern of action over which he has no control. A mechanism sets in when he is anxious that results in this compulsive behavior. Despite this difference between normal cases and pathological ones, such as obsessional actions, Freud holds that obsessional action also involves intention. What does he mean by that, and how is his view related to Gardner's contention that what distinguishes "deeper forms of irrationality" from self-deception is that the latter is intentional?

In the section on undoing what has been done, Freud speaks of an *aim* to undo. In speaking of an aim to undo and not just of a pattern of undoing, Freud is suggesting that there is an intention at work here; that this is an action rather than a mere behavior. One goal of Freud's therapy is to make his patient aware of the intentions that show themselves in patterns of behavior but of which the patient is not aware. In discussing the meaning of symptoms in the *Introductory Lectures*, Freud mentions the case of a woman who suffers from obsessional neurosis. Freud describes her obsessional behavior thus: "She ran from her room into another neighbouring one, took up a particular position there beside a table that stood in the middle, rang the bell for her housemaid, sent her on some different errand or let her go without one, and then ran back into her own room."[285] Freud asks the woman why she does this. At first, she can't answer him. One day, however, she suddenly recalls an experience, namely, the memory of her wedding night. Her husband had come running to her room many times during night, but every time without being able to consummate the sexual act. In the morning, he expressed concern that he would be humiliated in the eyes of the housemaid when she came to make the bed, so he took up a bottle of red ink and poured its contents over the sheet. But he missed the spot where a bloodstain would have been likely. The parallel between the obsessional action and the memory of the wedding night is the stain, since there is also a stain on the tablecloth on the table by which the lady stands when she rings for the housemaid. Freud sees an intimate connection between her obsessional action and the experience from the wedding night. He concludes: "It already seems proved that *the obsessional action had a sense*; it appears to have been a representation, a repetition, of the significant scene."[286] But Freud hopes to know more. He continues:

> But we are not obliged to come to a halt here. If we examine the relation between the two more closely, we shall probably obtain information about something that goes further – about the intention of the obsessional action. Its

[285] Freud, *Introductory Lectures* (1916-17), SE, vol. 16, p. 261. „Sie lief aus ihrem Zimmer in ein anderes nebenan, stellte sich dort an eine bestimmte Stelle bei dem in der Mitte stehenden Tisch hin, schellte ihrem Stubenmädchen, gab ihr einen gleichgültigen Auftrag oder entliess sie auch ohne solchen und lief dann wieder zurück." (GW, Band 11, s. 269.)

[286] Ibid. p. 262. My italics. „Der Beweis, dass die Zwangshandlung sinnreich ist, wäre bereits erbracht; sie scheint eine Darstellung, Wiederholung jener bedeutungsvollen Szene zu sein." (Ibid. s. 270.)

> kernel was obviously the summoning of the housemaid, before whose eyes the patient displayed the stain, in contrast to her husband's remark that he would feel ashamed in front of the maid. Thus he, whose part she was playing, did not feel ashamed in front of the maid; accordingly the stain was in the right place. We see, therefore, that she was not simply repeating the scene, she was continuing and at the same time correcting it; she was putting it right. But by this she was also correcting the other thing, which had been so distressing that night and had made the expedient with the red ink necessary – his impotence. So the obsessional action was saying: No, it's not true. He had no need to feel ashamed in front of the housemaid; he was not impotent." It represented this wish, in the manner of a dream, as fulfilled in a present-day action; it served the purpose of making her husband superior to his past mishap. [287]

According to Freud's interpretation, what the woman could contribute by telling him about the wedding night and making associations between that experience and her obsessional action was enough to show that the action had a *sense*: the repetition of the scene from the wedding night. In his interpretation of *why* the woman performed the obsessional action, Freud thought that he had found the *unconscious intention* of the action: the intention was to correct the embarrassing incident by showing the maid that the stain was in the right place and thus deny that her husband was impotent and redeem his reputation.[288]

On the face of it, Freud's interpretation would seem to spell out the sense of the action, rather than provide intention about the action, but Freud's theory assumes intention from the outset. He says: "My interpretation carries with it the hypothesis that intentions can find expression in a speaker of which he himself knows nothing but which I am able to infer from circumstantial evidence."[289] In Freud's usage of 'intention', the person who performs an intentional action need not be conscious of the intention; an observer can apprehend an intention in something someone else says or does, of which that person is himself not aware. We should note that neither the case discussed

---

[287] Ibid. pp. 262. „Aber wir sind nicht genötigt, bei diesem Schein Halt zu machen; wenn wir die Beziehung zwischen den beiden eingehender untersuchen, werden wir wahrscheinlich Aufschluss über etwas Weitergehendes, über die Absicht der Zwangshandlung erhalten. Der Kern derselben ist offenbar das Herbeirufen des Stubenmädchens, dem sie den Fleck von Augen führt, im Gegensatz zur Bemerkung ihres Mannes: Da müsste man sich vor dem Mädchen schämen. Er – dessen Rolle sie agiert – schämt sich also nicht vor dem Mädchen, der Fleck ist demnach an der richtigen Stelle. Wir sehen also, sie hat die Szene nicht einfach wiederholt, sondern sie fortgesetzt und dabei korrigiert, zum Richtigen gewendet. Damit korrigiert sie aber auch das andere, was in jener Nacht so peinlich war und jene Auskunft mit der roten Tinte notwendig machte, die Impotenz. Die Zwangshandlung sagt also: nein, es ist nicht war, er hatte sich nicht vor dem Stubenmädchen zu schämen, er war nicht impotent; sie stellt diesen Wunsch nach Art eines Traumes in einer gegenwärtigen Handlung als erfüllt dar, sie dient der Tendenz, den Mann über sein damaliges Missgeschick zu erheben." (Ibid.)

[288] This case of obsessional neurosis shares the trait that Freud ascribed to psychosis in "The Loss of Reality in Neurosis and Psychosis": the attempt to replace or revise reality.

[289] Freud, *Introductory Lectures* (1915-16), SE, vol. 15, pp. 64 „Meine Deutung schliesst die Annahme ein, dass sich bei dem Sprecher Intentionen Äussern können, von denen er selbst nichts weiss, die ich aber Indizien erschliessen kann." (GW, Band 11, s. 59.)

above nor the hypothesis suggest that the intention is something that the person has first been aware of and then suppressed; rather the intention shows itself without the person being or having been aware of having it. (This is an important difference between Freud's use of intention and Davidson's and Gardner's, where the self-deceiver "knows what she is up to".) There is an intention in the behavior that has *always been* unconscious. Thus one could say that such actions are characterized by not being recognized as actions (at least not as the actions that they are) by their agents. They do not see the intention in, or the full meaning of, their behavior.

I want to recall the discussion of intention in the first chapter, and more particularly Anscombe's remarks. According to her, something can be an intentional action even when the answer to the question "Why did you do it?" is "For no particular reason". My interpretation of Anscombe's view was that although there need not be a reason for an intentional action, it is typical of an intentional action that it makes sense to ask for a reason for it. But, as we saw, Anscombe distinguishes between cases in which one does something for no particular reason, and actions for which a reason seems to be required but where the actor seems surprised by his own action and, in seeking a reason for it, assuming that there must be one, cannot find one. Anscombe holds that in these latter cases there simply is no reason, "even if psychoanalysis persuades him to accept something as his reason, or he finds a reason in a divine or diabolic plan or inspiration, or a causal explanation in his having been previously hypnotized. [ ] I myself have never wished to use these words in this way, but that does not make me suppose them to be senseless. They are curious intermediary cases."[290] In short, of course, one *can* call the explanations that the subject later comes to adopt as explanations for his behavior reasons, but we can also choose not to. While Freud claims that intentions of which a speaker is unaware can be revealed to someone else by what he says, Anscombe rejects this use of 'intention' or 'reason'. Rather, as Anscombe says, "an action of this sort is voluntary, rather than intentional."[291] What does she mean by that?

Anscombe refers to an example of a man who is pumping the water supply for a house. The water is poisoned and, in the version of the example that is relevant for our discussion, the man who is pumping knows that the water is poisoned. The question is: how are we to understand his action if he claims that his intention is not to poison the people in the house but only to make a living? "In that case", Anscombe says, "although he knows concerning an intentional act of his – for it, namely replenishing the house water-supply, is intentional by our criteria – that is *also* an act of replenishing the house water-supply with *poisoned* water, it would be incorrect, by our criteria, to say that this act of replenishing the house supply with poisoned water was intentional."[292] I take it

[290] This quote is discussed in Chapter One, p. 46. Anscombe, *Intention*, § 17, p. 26.
[291] Anscombe, § 17, p. 26.
[292] Ibid. §25, p. 42.

that the distinction between voluntary and intentional comes in here: replenishing the house with poisoned water is in this case a voluntary act but not an intentional one. Still, she notices, this view has its problems since it can lead one to think that if knowing of the side effects of one's intentional action is not enough for calling the causing of those side effects an intentional action, intention seems to be some interior motive. (As if one could point inwards and say: "The intention in my head was only to supply the house with water.") Anscombe says that there is some truth in the statement "Only you can know if you had such-and-such an intention or not",[293] but it should not be understood as knowing what goes on in one's head. Anscombe goes on to specify the circumstances in which one can say that someone did something intentionally in such a way as to emphasize the person's actions and choices rather than interior motives. According to Anscombe, if the man was hired to poison the water, his action is intentional, even if the man says that he only did it for the money. This case is different from the former in that the man didn't simply go on doing what he had done for many years, knowing that this time the water he would bring up to the house would be poisoned. In the latter case, what he agreed to do, the "job description" so to say, was to carry poisonous water to the house: "So that while we can find cases where only the man himself can say whether he had a certain intention or not; they are further limited by this: he cannot profess not to have had the intention of doing the thing that was a means to an end of his."[294] Thus, although all that he was interested in was the money, he agreed to earn his money by poisoning people, which must be seen as part and parcel of his intention.

What Anscombe discusses above is different from the case of unconscious intentions discussed by Freud, since, in cases of unconscious intention, the actor cannot say anything about why he acted as he did; he is not aware of his action either as a desired goal or as a means towards a desired goal. Anscombe remarks: "'Roughly speaking, a man intends to do what he does.' But of course this is *very* roughly speaking."[295] In cases where someone is unable to provide a reason for his actions, he might simply not have had an intention with his action. This seems to be Anscombe's take on such cases, contrary to Freud's view that "intentions can find expression in a speaker of which he himself knows nothing".[296] Of course, there is no point from which to judge objectively

---

[293] Ibid. p. 44.
[294] Ibid. p. 44.
[295] Ibid. p. 45.
[296] Richard Moran writes: "with respect to knowledge of one's own intentions, philosophers sometimes invoke a distinction between certainty that is based on evidence or discovery, and certainty that is based on a decision made by the person. According to this distinction, uncertainty about what one intends to do is normally a matter of one's having not yet fully *formed* an intention, and this uncertainty is ended by a decision about what to do rather than by a discovery of an antecedently formed intention. The question expressing this uncertainty will *not* indicate a situation in which there is something I intend to do but I do not yet know what it is. Rather, the question expresses the fact that my intention itself is uncertain." (Richard Moran,

which use of 'intention' is the "right" one. Language use is malleable and conventions change. In this respect, Freud's use of "unconscious intention" is legitimate. Problems arise when, in describing this "new concept", 'unconscious intention', one understands it as an intention for which one can provide reasons. "Intention" in the first use seems to cover a different aspect or element in behavior. In my critique of Freud, my aim is not to establish the "correct" use of 'intention', but to point to where I think that Freud conflates or vacillates between unconscious intention and conscious intention and how this causes confusion.

I argue that Freud's account of unconscious intention pays too little attention to an important difference between being able to say why one did something (to articulate one's reason or intention) and not being able to. In his analysis of the woman who repeats what she experienced on her wedding night, Freud seems to think that there is a full-fledged intention, albeit unconscious, to be "discovered" even in the case when the person cannot give an answer to the question "Why?" Freud takes his interpretation to have revealed not the sense of her behavior (to repeat, an interpretation which she can provide herself), nor the deeper meaning, or fuller sense, of her repetition of the scene (to "correct" the event), but also a "plan" to show the stain for the maid as proof that her husband wasn't impotent and, by so doing, to rehabilitate his reputation.[297] Anscombe, on the other hand, seems to hold that not being able to provide an answer reveals a *significant* difference regarding the (normal) case

---

Authority and Estrangement: An Essay on Self-Knowledge, (Princeton: Princeton University Press: 2001), pp. 55.) Moran refers to Anscombe's *Intention* here. As we see, this picture of intention is different than Freud's, in that in Freud an intention is something that can be discovered in retrospect.

[297] There are parallels between my argument in this chapter and Marcia Cavell's argument in *The Psychoanalytic Mind: From Freud to Philosophy* (Marcia Cavell, *The Psychoanalytic Mind: From Freud to Philosophy*, Harvard: Harvard University Press, 1996) In most respects, we seem to agree in our understanding of unconscious mental states and also on which parts of Freud's writings offer problems in this area. Cavell says, for example, that Freud is of two minds in the matter of how to articulate the unconscious in "viewing primary processes as chaotic, and repressed mental content not only as perfectly rational but also as containing just those thoughts that will emerge after repression has been overcome." (p. 173) She continues: "It seems more likely that, as Freud acknowledges (1937) ["Constructions in Analysis", SE, vol. 23], what analyst and patient together construct by way of 'memory' or an earlier understanding was in many cases not thought of even unconsciously in just that way at the time. […] to say that someone is not oriented to the truth, even that she is in flight from it, does not necessarily imply that she already has those beliefs she might otherwise have come to hold." (p. 174) There are, however, also strong differences between Cavell's view and my own, especially as regards our views on Donald Davidson's account. In defending Davidson's account from the critique Mark Johnston presents in "Self-deception and the Nature of Mind" (see Chapter One where I affirmed Johnston's critique), she says: "I conclude […] that rationality is constitutive of the mental; and that we are forced to posit partitioning whenever holism is violated to a significant degree, as it surely is in some cases of self-deception and even in repression." (p. 201.) A main point of disagreement is that while Davidson and Cavell claims that rationality is constitutive of the mental, Johnston's tropist account is an argument against this view. (Cavell, p. 200; Johnston, p. 67, in *Perspectives on Self-Deception*).

of intentional action, which, in her view, makes 'intentional' an unsuitable description for the case that she discusses.

What makes one want to call the meaning of an act that the subject grasps in retrospect her intention? In a paragraph in Wittgenstein's *Philosophical Investigations* that Anscombe discusses, Wittgenstein asks:

> Why do I want to tell him about an intention too, as well as telling him what I did? – Not because the intention was also something which was going on at that time. But because I want to tell him something about *myself*, which goes beyond what happened at that time. I reveal to him something about myself when I tell him what I was going to do. – Not, however, on grounds of self-observation, but by way of response (it might also be called an intuition).[298]

In expressing one's intention in an action, one expresses something about oneself. One says something more than what is displayed in the action itself. Telling someone about one's intention is like giving a response to an implicit question: one wishes to say something about why one acted as one did, what one was thinking, what one was aiming at, etc. What one wants to reveal about oneself in telling of one's intention is not a "fact", i.e. not something one has observed. Freud, on the other hand, speaks of unconscious intentions as intentions of which the subject is not aware. One cannot tell of one's unconscious intentions; at the moment one can do so, they have become conscious.[299] Why would one want to use the word 'intention' about something of which the subject is unaware? Perhaps because an action such as the neurotic woman's strictly regulated action "says more" than what is strictly going on. It quite obviously shares elements with what she tells Freud about her experience on her wedding night. It also appears obvious that she is not at ease with those memories, which seems to trigger the obsessive behavior. Freud holds that her obsessive action reveals the intention of correcting the experience of her wedding night. Though she does not tell him about her intention in acting (she is unable to, since she is not aware of it herself), *her action* tells him a great deal about *her*: that she finds her husband's impotence embarrassing. But should we ascribe an unconscious intention to someone on the grounds that her action reveals something to us, and perhaps later to herself, of which she is not herself aware? I will now turn to Jonathan Lear's discussion of how the unconscious content of the mind differs from conscious intention, which raises important issues involved in Freud's account of unconscious intentions and unconscious beliefs, fears, etc.

Lear analyzes a moment in Freud's therapy of the Rat Man in which the Rat Man starts to heap gross and filthy abuse on Freud and his family only to

---

[298] Ludwig Wittgenstein, *Philosophical Investigations*, (Cornwall: Blackwell Publishing, 2001), § 659.

[299] As I will argue in Chapter Four, when one finally *can* tell someone about the intentions one takes oneself to have had earlier, though they were unconscious at the time, one has reached this knowledge on grounds of self-observation.

excuse himself immediately ever so much for uttering these insults. Freud writes:

> While he talked like this, he would get up from the sofa and roam about in the room, – a habit which he explained at first as being due to delicacy of feeling: he could not bring himself, he said, to utter such horrible things while he was lying there so comfortably. But soon he himself found a more cogent explanation, namely, that he was avoiding my proximity for fear of my giving him a beating. If he stayed on the sofa he behaved like someone in desperate terror trying to save himself from castigations of terrific violence […] He recalled that his father had had a passionate temper, and sometimes in his violence had not known where to stop.[300]

Lear notes that the Rat Man has a difficult time accounting for *why* he is cringing before Freud; he comes up with an explanation only to abandon it for, as Freud describes it, "a more cogent explanation". Lear takes this to imply that the Rat Man is suffering from a "*reflexive breakdown*": from the inability to give a full or coherent account of what he is doing.[301] The Rat Man glosses over this breakdown with various self-interpretations. He stops at the interpretation that he fears that Freud is going to give him a beating. Lear says:

> The Rat Man himself implicitly understands that if he is to interpret himself as afraid of Freud, he must at the same time come up with a reason for his fear. Thus he suggests that he is afraid that Freud is going to give him a beating. Now the Rat Man is interpreting himself as having not only an unconscious fear of Freud, but an unconscious belief about him. And when the Rat Man asks himself why he should believe that, he himself comes up with the thought that Freud reminds him of his violent father. The Rat Man is well on his way to interpreting himself as having an Unconscious Mind with its own beliefs and intentions.[302]

The Rat Man's way of rationalizing his fear in terms of unconscious emotions and unconscious beliefs gives the impression that his cringing was an action motivated by reasons. Lear continues: "Cringing has burst forth, and the Rat Man wants to make it intelligible to himself by giving it a reason. But consciously the Rat Man knows that he doesn't really have anything to fear from Freud. […] So the reasons for the cringe must be Somewhere Else. And

---

[300] Freud, "Notes upon a Case of Obsessional Neurosis", p. 209. „Bei diesen Reden pflegte er vom Diwan aufzustehen und im Zimmer herumzulaufen, was er zuerst mit Feinfühligkeit motivierte; er bringe es nicht über sich, so grässliche Dinge zu sagen, während er behaglich daliege. Er fand aber bald selbst die trieftigere Erklärung, dass er sich meiner Nähe entziehe, aus Angst von mir geprügelt zu werden. Wenn er sitzen belieb, so benahm er sich wie einer, der sich in verzweifelter Angst vor masslosen Züchtigungen schützen will [...] Er erinnerte, dass der Vater jähzornig gewesen war und in seiner Heftigkeit manchmal nicht mehr wusste, wieweit er gehen durfte." (GW, Band 7, s. 429.)

[301] Jonathan Lear, *Open Minded. Working out the Logic of the Soul.* (Harvard University Press, 1999), p. 81.

[302] Ibid. p. 91.

the reason for the Somewhere Else is the felt need to give reasons for the cringe."[303] Thus, the reaction of providing a reason for everything we do in trying to understand the sense of our actions and behaviors leads us to posit different rational structures of the mind. This leads to a conception of the mind similar to Davidson's, i.e. as consisting of rational wholes standing in conflict with each other. Lear is strongly critical of accounts that posit a Second Mind. What must be resisted, Lear holds, is the assumption that there is a reason-explanation to be given for each and every one of our acts and behaviors.[304]

Lear says, that "[i]n cringing, the Rat Man *acts out* fear", and that acting out is "an activity which isn't an action". [305] Lear describes the Rat Man's cringe as "fearful" rather than as an expression of fear. The point is to separate this reaction from a manifestation of fear as a propositional attitude for which there is a reason-explanation: "The very use of language to describe these mental states and activities tends to make them look more rational than they are. Just by giving these mental states a name, we make them seem already to be *in the domain of logos*, while what we are in fact trying to capture is their not (yet) being there."[306] Lear argues that when Freud accepts the Rat Man's rationalization that he is *unconsciously afraid* that Freud is going to beat him, he is led off in the wrong direction. In holding the cringe to be an expression of fear, we look for a reason for this fear. As Lear says:

> If we fear something we *believe* it to be a danger – and we take that danger to be a legitimate cause of our fear. And thus if one assumes Mr. R [the Rat Man] is afraid, and there is no conscious belief that Freud is a threat, then there is conceptual pressure to conclude there must be an unconscious belief. […] They [the Rat Man and Freud] are already en route to conceptualizing the

---

[303] Ibid.

[304] In discussing the Rat Man case in *Open Minded*, Lear writes about obsessionals: "As anyone who has worked with them will know, obsessionals tend to interpret themselves as being more rational than they are. *Rationalization is among the favorite forms of obsessional defense*." (Lear, *Open Minded*, p. 92. My italics.) As the Rat Man's case shows, rationalization, seemingly the result of a genuine quest for knowledge, can actually be a hindrance in the attempt to identify the real cause since in accepting the rationalization as an explanation for his behavior, the Rat Man stops his self-interrogation. I recommend Joshua Landy's book *Philosophy as Fiction: Self, Deception, and Knowledge in Proust* for a careful analysis of a complex instance of evasion. Landy is analyzing the narrator's obsession with obtaining evidence for, and "knowledge" of, his lover Albertine's fidelity or lack of thereof, and how this search for knowledge actually serves to avoid confrontation with facts which, the narrator himself suspects, could be evidence for or against her fidelity. (Joshua Landy, *Philosophy as Fiction: Self, Deception, and Knowledge in Proust*, Oxford: Oxford University Press, 2004, Chapter 2: *Self-deception (Alberine's Kimono)*).

[305] Lear, *Open Minded*. p. 92. My italics.

[306] Ibid. p. 93. My italics. There are similarities between the view Lear gives expression to here and Lacan's/Bruce Fink's view expressed in *Lacan to the Letter*. Fink discusses Lacan's critique of Freud's ascription of an intention to his patient to deceive him about dreams that she reports having had. Fink reproduces Lacan's interpretation (in Seminar IV, 108): "the woman had not yet formed an *intention* to trick him; 'it was a desire.' Freud makes it into something more than a desire, something real, by naming it, by symbolizing it, by 'interpreting it too soon'." (Bruce Fink, *Lacan to the Letter*, Minneapolis: University of Minnesota Press, 2004.)

> unconscious as a rational structure with a "mind of its own". For, as we have seen, it makes no sense that there should be only one unconscious belief on its own.[307]

If we think of fear as a *propositional attitude*, for which there is a rationalizing structure of supporting beliefs and other propositional attitudes, assuming that someone is unconsciously afraid and that this fear cannot be rationalized by *conscious* beliefs etc. which the person holds, we implicitly presuppose an *unconscious mind* in which this fear makes sense. Lear's point is that unconscious hopes, fears, wishes etc. need to be understood differently than conscious beliefs, desires etc., as other than *propositional attitudes*: "What the cringe lacks is, in the literal sense of the term, *information*. It has not yet been fully formed, because it has not been taken up into logos and embedded in the web of beliefs, expectations, and desires which would help to constitute it as fear."[308] The cringe is not a response to a situation that is based on rational judgment in the light of other beliefs, desires and expectations. Instead of calling the cringe an expression of fear, Lear calls the cringe an expression of *phantasy* which operates "in relation to, but relatively free of, the rationalizing constraints of logos – the holistic system of an agent's beliefs and desires, fears, angers, and other propositional attitudes."[309] Lear says of phantasy that it "will typically 'show' a meaning where it does not say'"[310] and that this is a way in which it remains relatively cut off from conscious understanding. Indeed, Lear says, it is *because* phantasy is relatively free from the constraints of logos, because it doesn't have to interact with other beliefs that the person holds, that an expression of phantasy survives for long and is so powerful.[311]

Instead of picturing irrationality as a conflict between different elements of the mind that are in themselves rational wholes, Lear argues that irrationality should be seen as *immanent* in mind. While what Lear calls the "Two-Minds Schema" assumes that "rationality is built into the very ideas of agency, action and mind",[312] he wants to show "that it is intrinsic to the very idea of mind that mind must be sometimes irrational. Rather than see irrationality as *coming from the outside* as from an Unconscious Mind which disrupts Conscious mind, one should see irrational disruptions as themselves an inherent expression of mind."[313] Although human beings are capable of acting for reasons, and

---

[307] Lear, *Freud*, pp. 31.
[308] Lear, *Open Minded*, p. 94.
[309] Ibid. pp. 92.
[310] Ibid. p. 92.
[311] Ibid. p. 93.
[312] Ibid. p. 81.
[313] Ibid. p. 84. Herbert Fingarette writes in his paper "Self-Deception Needs no Explaining" that "[e]xplaining correctly how the mind works reveals self-deception as non-puzzling and in no need of any special explanation." ("Self-Deception Needs no Explaining" in *Self-Deception*, Berkeley: University of California Press, Ltd., 2000, p. 176) Though, in Fingarette's account this does not mean to explain that the mind is sometimes irrational. Rather, he holds that self-deception can be explained fully in terms of attention; as avoidance of directing one's attention towards something.

although we distinguish ourselves from the rest of nature by being self-interpreting animals, we are also sometimes unable to give a full or coherent account of what we are doing. In Lear's terms, we sometimes suffer *reflexive breakdowns*, "disruptions in our capacity to be self-interpreting animals".[314]

Lear argues that the Rat Man and Freud are mistaken in thinking that there must be a reason for the cringe, and in treating the Rat Man's fearful response as if it were a rational response for which the agent has reasons. The untangling of the problems involved in not making a distinction between fear (of something in particular, conscious fear) and a fearful reaction, I argue, reveals problems inherent in Freud's use of 'unconscious intention' in the analysis of the neurotic woman. Recall Anscombe's definition of an intentional action as an action for which one can give reasons. When the obsessional woman's behavior is seen as an action, Freud is led to think that there must be an intention in her action, a further reason for why she acts as she does. But when she is asked for a reason, she cannot provide one. He then thinks that there must be an unconscious intention, namely, to correct what happened on her wedding night. The unconscious intention that he ascribes to her is a plan for how to rehabilitate her husband's reputation. Lear's distinction between fearing something and acting out anxiety (or fearfulness) shows that the Rat Man's cringe is not an action for which there are reasons – it is not "in the domain of logos", "embedded in the web of beliefs, expectations, and desires which would help constitute it as fear." Analogously, I argue, the neurotic woman's obsessive behavior is not an intentional action with the further aim of correcting an earlier incident. What happens is rather that she acts out the discomfort she feels when the memory of the incident is recalled. Freud ascribes too much reason to the woman's behavior when he assumes an unconscious intention of correcting the incident on the night of the wedding. In Freud's analysis of the case, this includes a quite intricate plan similar to the rationalizing explanation that the Rat Man gave for his cringe. Analogous to Lear's discussion of the Rat Man, I would argue that her behavior is not a rational response to a situation with interrelated intentions, means and goals, but rather a case of acting out of her discomfort.

We have seen that Lear's critique of the Two-Minds Accounts also applies to Freud in the cases where rationalizations of premature reactions (e.g. fearfulness) as mature feelings (fear) lead him to accept or assume a context for which this feeling makes sense. I agree with Lear in his analysis of the Rat

---

For example: "the way our mind goes about producing self-deceptive states of mind is the same as for the non-deceptive ones. [ ] Suppose, for example, that I have done something shameful. I take account of my conduct and its significance for me. However, just because this particular shame is deeply wounding to me, given my sense of self, I avoid focusing my attention on the event, or at least on its shameful features." (p. 169) I object to Fingarette's portrayal the self-deceiver as having knowledge of "his conduct and its significance". In Fingarette's portrayal self-deception takes place on a conscious/preconscious level, I argue that it is better understood if we allow for an unconscious level as well.

[314] Lear, *Freud*, p. 81.

Man's cringe, and I have argued that the problems carry over to Freud's analysis of the neurotic woman's obsessive behavior.

In *Wittgenstein Reads Freud: The Myth of the Unconscious*, Jacques Bouveresse points out that a major problem with Freud's theory of the unconscious is that Freud portrays unconscious mental processes as if they were *just like* conscious one's, only not perceived.[315] This critique hits the mark in a number of Freud's many attempts to account for the difference between the conscious and the unconscious. In a characterization of the spatial idea of the unconscious and the conscious/preconscious in his *Introductory Lectures*, Freud describes the unconscious and the preconscious as different rooms with a watchman on the threshold between the two rooms, and consciousness as an eye placed in the room of the preconscious: "But even the impulses which the watchman has allowed to cross the threshold [to the preconscious] are not […] necessarily conscious as well; they can only become so if they succeed in catching the eye of consciousness."[316] We see that what distinguishes unconscious mental content is that it belongs to the room of the unconscious where it is not accessible to consciousness, since it is spatially cut-off from its field of vision. Bouverresse's objection, inspired by his reading of Wittgenstein and Kurt Koffka, among others, is that Freud does not pay attention to the "grammatical" difference between conscious and unconscious states or processes. Bouveresse quotes Koffka: "When one found it necessary to go beyond consciousness in the description and exploration of mind, one imagined the non-conscious parts of mind to be fundamentally alike to the conscious one, fundamentally alike, that is, in its aspects or properties with the exception of being conscious."[317] He also cites Wittgenstein's remark on the unconscious in *The Blue Book*:

> The idea of there being unconscious thoughts has revolted many people. Others again have said that these were wrong in supposing that there could only be conscious thoughts, and that psychoanalysis had discovered unconscious ones. The objectors to unconscious thought did not see that they were not objecting to the newly discovered psychological reactions, but to the ways in which they were described. The psychoanalysts on the other hand were misled by their own way of expression into thinking that they had done more than discover new

---

[315] Jacques Bouveresse, *Wittgenstein Reads Freud: The Myth of the Unconscious*, (Princeton: Princeton University Press, 1995).

[316] Freud, *Introductory Lectures*, SE, vol. 16, p. 296. „Aber auch die Regungen, welche der Wächter über die Schwelle gelassen, sind darum nicht notwendig auch bewusst geworden; sie können es bloss werden, wenn es ihnen gelingt, die Blicke des Bewusstseins auf sich zu ziehen." (GW, Band 11, s. 306.) It should be noted that Freud calls these ideas "crude" and says, "more than that, I know that they are incorrect". I read Freud here as saying that this is merely a "picture", which he means can throw *some* light on conscious, preconscious and unconscious content of the mind.
(In her novel *The Sacred and Profane Love Machine*, Iris Murdoch plays with insight and witt with the Freudian mental apparatus in her story of the therapist Blaise's double-life. Iris Murdoch, *The Sacred and Profane Love Machine* , Penguin Books Ltd, 1976.)

[317] Kurt Koffka, "On the Structure of the Unconscious" in *The Unconscious. A Symposium* (New York: Alfred A. Knopf, 1928), 43-68, pp.43. Cited in Bouverresse, p. 25.

> psychological reactions; that they had, in a sense, discovered conscious thoughts which were unconscious.[318]

It is true that many of Freud's attempts to describe unconscious thought, feelings, wishes etc. fail to characterize these as qualitatively, or grammatically, different from conscious beliefs, feelings, desires, etc. But I want to acknowledge that there are texts in which Freud *does* account for a qualitative or grammatical difference. Now I wish to turn to Freud's paper "The Unconscious" in which he analyses repression in a way that is not susceptible to the problems discussed by Lear and Bouveresse above.

Idea versus Belief

What is the status or quality of that which is being repressed?[319] Freud writes that the ego "is able by means of repression to keep the idea which is the vehicle of the reprehensive impulse from becoming conscious. Analysis shows that the idea often persists as an unconscious formation."[320] The idea is "kept from becoming conscious", which suggests that it was not conscious before the repression. But how are we to understand this idea, which is an unconscious formation? Is it propositionally formulated? Is it a propositional attitude? Is it a belief? We first notice that the repressed idea, like the unconscious intention, can be inferred by the analyst (or, by someone else) from someone's action or speech, without this person himself being aware of harboring this idea. The repressed idea can be inferred from symptoms, for example, since symptoms are at one and the same time prohibitions against ideas being made conscious and distorted expressions of ideas. But the fact that the interpreter can express what he apprehends in the patient's behavior in a proposition does not imply that the repressed idea has been propositionally formulated as a belief by the patient. Even in the case where the patient agrees to what the analyst claims that he unconsciously feels, e.g. that he is angry with his father, this does not mean that the patient has formed a belief that he is angry with his father, at least not if holding a belief implies being able to give a reasoned explanation as to why one holds that belief, as in Davidson's and Gardner's accounts of beliefs as propositional attitudes. The patient's acceptance of the wording of the analyst's interpretation doesn't necessarily imply insight. It can be mere repetition,

---

[318] Ludwig Wittgenstein, *The Blue and Brown Books*, (Oxford: Blackwell, 1958), p. 57.

[319] In the German original, Freud most often calls that which is repressed *'Vorstellung'* (See, for example, Editor's note in "The Unconscious", p. 174, n. 1 and, "Repression", vol. 14, p. 152, n.1); in SE it is translated as 'idea' or 'ideational representative' (See for example Editor's note in "The Unconscious", p. 201, n. 1). Other possible translations would be, for example, 'image' or 'presentation' (See Editor's comment, "The Unconscious", p. 174, n.1).

[320] Freud, *Inhibitions, Symptoms and Anxiety*, p. 91. „Das Ich erreicht durch die Verdrängung, dass die Vorstellung, welche der Träger der unliebsamen Regung war, vom Bewusstwerden abgehalten wird." (GW, Band 14, s. 118.)

perhaps combined with a belief that whatever the authority figure (the analyst) says is true. As we will see later, Freud stresses the importance of conceptual formulation in becoming aware of something; for this mere repetition of someone's interpretation is not enough. There are stages of awareness, or knowing:

> Knowledge is not always the same as knowledge: there are different sorts of knowledge, which are far from equivalent psychologically. […] The doctor's knowledge is not the same as the patient's and cannot produce the same effects. If the doctor transfers his knowledge to his patient as a piece of information, it has no result. No, it would be wrong to say that. It does not have the result of removing the symptoms, but it has another one – of setting the analysis in motion, of which the first signs are often expressions of denial. The patient knows after this what he did not know before – the sense of his symptom, yet he knows it just as little as he did. Thus we learn that there is more than one kind of ignorance. […] But our thesis that the symptoms vanish when their sense is known remains true in spite of this. All we have to add is that the knowledge must rest on an internal change in the patient such as can only be brought about by a piece of psychological work with a particular aim.[321]

This quote is rich in content. It says that there are different kinds of knowledge, and of ignorance. It also provides support for an important distinction between 1) something expressing itself (in behavior or words) in a speaker of which he himself is unaware; 2) the analysand's repetition of the analyst's interpretation of his behavior; 3) a person speaking his mind (his beliefs, intentions etc.) directly. If we assume that the patient repeats the interpretation that the analyst has transferred to him as a piece of information, he is, of course, uttering a sentence, a propositional formation, but not as an *expression* or *avowal* of *his* intention, wish, fear etc. We could perhaps say that the difference between the patient who merely repeats the psychoanalyst's interpretation and the patient who has become aware that a motivation or idea expresses itself in his behavior is that the latter has overcome his inability to apprehend his own motivations, wishes, desires, etc. This motivation can now be promoted in an intentional action, and what was earlier an unconscious wish, desire or feeling can now be expressed as a belief; e.g. someone's belief that he fears his father. The patient has re-acquired the natural, first-person ability to apprehend and avow his own

---

[321] Freud, *Introductory Lectures*, vol. 16, p. 281. „Wissenund Wissen ist nicht dasselbe; es gibt verschiedene Arten von Wissen, die psychologisch gar nicht gleichwertig sind. [...] Das Wissen des Arztes ist nicht dasselbe wie das des Kranken und kann nicht dieselben Wirkung äussern. Wenn der Arzt sein Wissen durch Mitteilung auf den Kranken überträgt, so hat dies keinen Erfolg. Nein, es wäre unrichtig, es so sagen. Es hat nicht den Erfolg, die Symptome aufzuheben, sondern den anderen, die Analyse in Gang zu bringen, wovon Äusserungen des Widerspruches häufig die ersten Anzeichen sind. Der Kranke weiss dann etwas, was er bisher nicht gewusst hat, den Sinn seines Symptoms, und er weiss ihn doch ebensowenig wie vorhin. Wir erfahren so, es gibt mehr als eine Art von Unwissenheit. [...] Aber unser Satz, dass die Symptome mit dem Wissen auf einer inneren Veränderung im Kranken beruhen muss, wie sie nur durch eine psychische Arbeit mit bestimmtem Ziel hervorgerufen werden kann. (GW, Band 11, s. 291.)

intentions, wishes, desires, etc., an ability we have, for example, when we are not restricted by anxiety and the reactions it can provoke.[322]

I have discussed 'idea' (*Vorstellung*) in relation to the previous discussion of 'unconscious intention' and suggested that the repressed idea lacks propositional formulation, in contrast to a belief. I find support for this interpretation in Freud's "The Unconscious". Here Freud describes the distinction between that which has assumed a propositional form and that which has not as his discovery of what distinguishes repressed material from conscious material, i.e. the distinction or division between the conscious and the unconscious. Earlier, Freud thought of the division between conscious and unconscious as physical, i.e. as spatial division. Here he presents his "discovery" that the difference is between something that is propositionally formulated and something that is not. He writes:

> What we have permissibly called the conscious presentation of the object can now be split up into the presentation of the *word* and the presentation of the *thing*; the latter consists in the cathexis[323], if not the direct memory-images of the thing, at least of remoter memory-traces derived from these. We now seem to know, all at once, what the difference is between a conscious and an unconscious presentation. The two are not, as we supposed, different registrations of the same content in different psychical localities, nor yet different functional states of cathexis in the same locality; but *the conscious presentation comprises the presentation of the thing plus the presentation of the word belonging to it, while the unconscious presentation is the presentation of the thing alone.*[324]
>
> Now, too, we are in a position to state precisely what it is that repression denies to the rejected presentation in the transference neurosis: what it denies to the presentation is translation into words which shall remain attached to the object.

---

[322] In Chapter Four, I continue the discussion of privileged access (first-person knowledge) and how it might fail.

[323] German: *Besetzung*. The term 'cathexis' (*Besetzung*) stems from Freud's hypothesis of psychic energy. Cathexis refers to the process that attaches psychical energy, essentially libido, to an object, whether this is the representation of a person, body part, or psychic element. The term also denotes the binding of psychic energy to interconnected representations in the progressive organization of the psyche." (Alain de Mijolla (ed.), *International Dictionary of Psychoanalysis*, Macmillan Library Reference, 2004, "Cathexis", p. 259.)

[324] Freud, "The Unconscious", p. 201. My italics. „Was wir die bewusste Objektvorstellung heissen durften, zerlegt sich uns jetzt in die *Wortvorstellung* und in die *Sachvorstellung*, die in der Besetzung, wenn nicht der direkten Sacherinnerungsbilder, doch entfernterer und von ihnen abgeleiteter Erinnerungsspuren besteht. Mit einen Male glauben wir nun zu wissen, wodurch sich eine bewusste Vorstellung von einer unbewussten unterscheidet. Die beiden sind nicht, wie wir gemeint haben, verschiedene Niederschriften desselben Inhaltes an verschiedenen psychischen Orten, auch nicht verschiedene funktionelle Besetzungszustände an demselben Orte, sondern die bewusste Vorstellung umfasst die Sachvorstellung plus der zugehörigen Wortvorstellung, die unbewusste ist die Sachvorstellung allein." (GW, Band 10, s. 300.)

> A presentation which is not put into words, or a psychical act which is not hypercathected, remains thereafter in the Ucs. in a state of repression.[325]

We learn that it is *presentation put into words* that is denied in repression. The instinct itself is not thereby hindered, but the *acknowledgment* of it is, it remains unconscious. In his treatment of the repression of instincts in "The Unconscious", Freud points out that it cannot be the impulse itself but rather it must be the 'idea' (*Vorstellung*) of the impulse that is repressed:

> We have said that there are conscious and unconscious ideas; but are there also unconscious instinctual impulses, emotions and feelings or is it in this instance meaningless to form combinations of the kind? [ ] I am in fact of the opinion that the antithesis of conscious and unconscious is not applicable to instincts. An instinct can never become an object of consciousness – only the idea that represents the instinct can. Even in the unconscious, moreover, an instinct cannot be represented otherwise than by an idea. If the instinct did not attach itself to an idea or manifest itself as an affective state, we could know nothing about it. When we nevertheless speak of an unconscious instinctual impulse or of a repressed instinctual impulse […] [w]e can only mean an instinctual impulse the ideational representative of which is unconscious.[326]

The idea is a mental representation of the instinct that can be conscious or unconscious. How are we to understand this? What is the presentation that gets rejected if it is not put into words? When Freud writes that the unconscious presentation is the presentation of the thing alone without the presentation of the word belonging to it, does that mean that the unconscious presentation – the idea – is a non-linguistic entity? This doesn't seem right. How could something non-linguistic represent? Further, we know that in repression it is the unwanted, anxiety-provoking idea *together with other ideas* (unconsciously) *associated with it*, that gets repressed; how can *association* be accounted for if the unconscious presentation is not linguistic? As Immanuel Kant famously says in *The First Critique*: "Thoughts without content are empty, intuitions without

---

[325] Ibid. p. 202. „Wir können jetzt auch präzise ausdrücken, was die Verdrängung bei den Übertragungsneurosen der zurückwiesenen Vorstellung verweigert: Die Übersetzung in Worte, welche mit dem Objekt verknüpft bleiben sollen. Die nicht in Worte gefasste Vorstellung oder der nicht überbesetzte psychische Akt bleibt dann in *Ubw* als verdrängt zurück." (Ibid.)

[326] Ibid. p. 177. „Wir sagten, es gäbe bewusste und unbewusste Vorstellungen; gibt es aber auch unbewusste Triebregungen, Gefühle, Empfindungen, oder ist es diesmal sinnlos, solche Zusammensetzungen zu bilden? [ ] Ich meine wirklich, der Gegensatz von Bewusst und Unbewusst hat auf den Trieb keine Anwendung. Ein Trieb kann nie Objekt des Bewusstseins werden, nur die Vorstellung, die ihn repräsentiert. Er kann aber auch in Unbewussten nicht anders als durch die Vorstellung repräsentiert sein. Würde der Trieb sich nicht an eine Vorstellung heften oder nicht als ein Affektzustand zum Vorschein kommen, so könnten wir nichts von ihm wissen. Wenn wir aber doch von einer unbewussten Triebregung oder einer verdrängten Triebregung reden, so ist dies eine harmlose Nachlässigkeit des Ausdrucks. Wir können nichts anderes meinen als eine Triebregung, deren Vorstellungsrepräsentanz unbewusst ist, denn etwas anderes kommt nicht in Betracht." (GW, Band 10, ss. 275.)

concepts are blind."[327] How can intuitions (or "things") alone provoke anxiety and give rise to repression and other defense mechanisms? It would seem that the idea must belong to the realm of language; that it must be a sign.

What does Freud mean then when he says that the unconscious presentation is "not put into words", that it is a "presentation of the thing alone", and that *this* is the difference between the conscious and the unconscious? Freud uses the psychoanalytical term 'hypercathexis' to describe wherein the difference between the unconscious and the conscious presentation lies. Concerning the distinction between the unconscious and conscious in terms of thing-presentation and thing- and word-presentations, Freud says:

> The system *Ucs.* contains the thing-presentation of the objects, the first and true object-cathexes; the system *Pcs.* comes about by this thing-presentation being hypercathected through being linked with the word-presentations corresponding to it. It is these hypercathexes, we may suppose, that bring about a higher psychical organization […].[328]

Earlier, in his topological account[329], Freud described the difference between the system of the preconscious/conscious and the system of the unconscious in terms of potential: that what belongs to the preconscious/conscious is not necessarily conscious, but it is "*capable of becoming conscious"*.[330] "Capable of becoming conscious" here was understood as meaning that the mental content had passed the threshold between the Unconscious and the Preconscious, and was thus accessible to the "eye of attention".[331] Now, "capable of becoming conscious" means that the thing- and word-presentations (*Sach- und Wortvorstellung*) are linked.[332] What is it that is lacking in that which is unconscious? It is not merely that attention hasn't been directed at it. Neither is it that what is unconscious is walled-off from what is preconscious/conscious, since Freud has now given up the spatial division of the topological account as

---

[327] Immanuel Kant, *Critique of Pure Reason*, transl. Norman Kemp Smith, (London: Macmillian Press Ltd.), A51/B 75, p. 93.

[328] Freud, "The Unconscious", pp. 201. „Das System *Ubw* enthält die Sachbesetzungen der Objekte, die ersten und eigentlichen Objektbesetzungen; das System *Vbw* entsteht, indem diese Sachvorstellung durch die Verknüpfung mit den ihr entsprechenden Wortvorstellungen überbesetzt wird. Solche Überbesetzungen, können wir vermuten, sind es, welche eine höhere psychische Organisation herbeiführen [...]" (Ibid, s. 300.) Freud refers to primary and secondary processes in accounting for the difference between the unconscious and the preconscious. Upon what I have cited follows: "and make it possible for the primary processes to be succeeded by the secondary processes which is dominant in the *Pcs*."

[329] "The first topology" refers to Freud's account of the psyche as developed in *The Interpretation of Dreams*: as divided into *conscious*, *unconscious* and *preconscious*. "The second topology", which was developed in *The Ego and the Id*, refers to the division of the psyche into *id*, *ego* and *super-ego*.

[330] Freud, "The Unconscious", p. 173. "Es ist noch nicht bewusst, wohl aber *bewusstseinfähig*." (GW, Band 10, s. 272.)

[331] I discuss this on p. 139 as well.

[332] What makes something preconscious rather than conscious is, still, that it is not presently at the center of one's attention.

explanation. Rather, *that which is unconscious lacks a certain psychical organization that comes with "being put to words",* that is, *articulated.*[333]

We have seen that when the sense or intention of the act is unconscious, it is typical that the person is unable (on a conscious level) to make certain associations, and is limited in her ability to reflect upon her behavior. The obsessive woman, for example, can describe her obsessional act by telling Freud about her experience during her wedding night, but she is unable to understand what her obsessional repetition of this experience *means*. In other cases, the person finds a cogent explanation for his behavior, but it is a rationalization and not an expression of insight into his motivations. The rationalization can be an unconscious effort to *avoid* understanding his motivations. Anna Karenina, for example, thinks that she is angry with Vronsky because he shows his feelings for her so openly, when it is really her own feelings and actions that bother her the most. We could say that Anna cannot articulate her discomfort even to herself, but this does *not* mean that she is aware of feelings that she cannot communicate. Rather, in becoming able to articulate her feelings, she becomes aware of what feelings it is that she harbors. She was aware of her feelings for Vronsky as having a certain quality before, that being with him made her happy, that he was attracted to her, and perhaps also that she was attracted to him. Anna was also *sensitive* to the articulation of her feelings before she had articulated it herself, as the refutation of Vronsky's expression of love shows. Understanding what it is that one feel, thus, doesn't mean to move from a non-linguistic understanding of one's feelings to a linguistic one (what ever that would mean); rather, it means finding the right articulation for one's feelings. (And this can be frightening, something which one avoids.) Thus, "to understand" is often a matter of understanding *better* what one somehow already knew. In Anna Karenina's case, it means understanding and accepting her feelings of happiness when in Vronsky's company, her more frequent engagement with the social circle where she might meet him, her disappointment when he did not turn up as she expected, etc. *as* manifestations of loving Vronsky. Anna's recognition in words, her thought "I am in love with Vronsky", is not something separate from her feelings, something "added on", but "the thing" (love) itself, brought to awareness. When she articulates her love for Vronsky, her feelings can be reflected upon, they can be evaluated in the context of other feelings, values, duties, etc.

If we understand the unconscious presentation, the "idea" (*Vorstellung*), as part of the linguistic realm, but different from the "word-presentation" (*Wortvorstellung*), i.e. in lacking in psychical organization, one way of characterizing this difference is that the unconscious presentation is not a propositional attitude. As we have seen, holding a propositional attitude means

[333] At this point in the text, the translators of SE alter the translation of *Vorstellung*. I judge that this is not of direct relevance for this discussion. See note 1, p. 201, vol. 14 in SE for an explanation.

that one can give a reason explanation, or a *justification for why* one believes so-and-so. Propositional attitudes thus imply a high level of psychical organization, since they are connected to other propositional attitudes together with which they make up a rational whole. The idea seems to differ from the belief, or any propositional attitude, by not being subject to attempts at justification. Recall Lear's distinction between fearing something and fearful behavior, where he described the latter as not yet being in the domain of logos. Here we find another, perhaps better, characterization of the idea: it is *not yet* in the domain of logos. It does not yet belong to a structure of other beliefs, desires, feelings etc. that makes it reasonable. The person who is unconscious of something has trouble making *sense* of her behavior, feelings, reactions, etc. She might *avoid* attempts at understanding her behavior, feelings and reactions, or she might struggle to understand, but the best she can do is to come up with possible interpretations. This shows that the unconscious content of the mind is not integrated with the rest. Although unconscious content is also linguistic, it lacks the articulation needed to connect it conceptually and rationally to a person's conscious knowledge and attitudes.

How are we to understand this failure in integration? In a passage which I cited a few pages back, Freud spoke of different kinds of knowledge which are not psychologically equivalent: "The doctor's knowledge is not the same as the patient's and cannot produce the same effects. If the doctor transfers his knowledge to his patient as a piece of information […] [i]t does not have the result of removing the symptoms […] Thus we learn that there is more than one kind of ignorance." The patient learns something when the doctor shares his interpretation with him, but this is not the same thing as understanding the sense of his symptoms, and it does not suffice to make him able to stop acting them out. Being informed and understanding something is not the same thing. Thus, although repeating the analyst's interpretation is "to put his problems into words", it is not enough for becoming conscious of the sense of the symptoms. To "put something to words" in a sense that is relevant for making it conscious must mean more than merely repeating someone else's words. The articulation is in this case just "borrowed", rather than a result of an achievement of understanding on the part of the patient, such as coming to hold a belief.

Let the Rat Man serve as an example. In analyzing the Rat Man, Freud discovers that he seems to have an unconscious feeling of anger and hatred towards his father, while at the same time maintaining that his father is the person he loves most of all. But it is not enough for the Rat Man simply to cease protesting against Freud's interpretation, or that he verbally acquiesces to it, if he is to gain genuine insight into his attractions (of love and hate) and what motivates them. He has to do the work of articulation himself. One could say that, at the stage of mere acquiescence, the patient does not yet know – or hold the belief – that he holds competing feelings of strong anger and hatred

towards his father, on the one hand, and great love, on the other. He has, at best, just begun considering Freud's proposed interpretation.

We might understand the difference between the unconscious presentation as a presentation of the thing alone, and the conscious presentation as a presentation of a word and thing, as the difference, on the one hand, between a largely unstructured pattern of associations between impulses, instincts, feelings and pictures and, the understanding of these *as* i.e., a feeling of fear, hate or love on the other. In perceiving these attitudes as parts of a whole instead of experiencing them as isolated, the person can confront his feeling of, e.g. hatred towards his father. This makes it possible for him to see how he might try to change his behavior and/or attitudes, which is really the same thing as coming to grips with what frightens him. Perceiving the situation as a whole makes rational, intentional thought and action possible. But it also means that one cannot escape from apprehending that which has been concealed when the apprehension and understanding was distorted. If it is the recollection of a traumatic experience, one is no longer prevented or protected from facing the anxiety. We recall that isolation is one of the defenses against, or flights from, that which threatens or evokes anxiety. To stay at the level of an non-verbalized and thus indeterminate, apprehension of something can be seen as a case of isolation: something is kept unconscious by not being included in the realm of articulated attitudes, knowledge, etc. It is prevented from becoming a determinate problem, and thus the patient is freed from the possibility, and therewith responsibility, of acknowledging it as something demanding his attention and perhaps even action. Although this reaction can reduce the anxiety, or even conceal it completely, no means have been taken to confront the threat and quiet the anxiety for good. That requires breaking the isolation and thus acknowledging that which has been isolated from one's conscious thought. What is required, in short, is that one *faces* one's anxiety.

Gardner neatly summarizes the function of ideas in his discussion of *Inhibitions, Symptoms and Anxiety*. He says that symptoms display "ideas, whose relations to one another are not such as to form propositions, but rather […] *relations of association*."[334] While the content of consciousness has a propositional character, the repressed unconscious is characterized by being prevented from assuming a propositional form.[335] Therefore, according to Gardner: "*Identifying the material of repression as ideas makes it possible to understand how repression can occur without the instrument of conscious thought*: in Freud's later theory [*Inhibitions, Symptoms and Anxiety*], anxiety is the signal which causes repression, and it is triggered by ideas, not judgments."[336] Gardner

[334] Gardner, p. 104. My italics.
[335] Gardner: "The basic philosophical intuition behind Freud's account is not hard to discern. Consciousness is bound up with the predominantly *propositional* character of its contents, so a good way of accounting for the impossibility of something's becoming conscious is to suppose that it is prevented from assuming a propositional form." (Ibid.)
[336] Ibid. My italics.

agrees with Freud's description of repression as "something between flight and condemnation", and the motive for repression as an impulse's "incompatibility" with the ego, which is clearly concerned with something below the level of propositional inconsistency.[337] Here Gardner argues that judgment doesn't belong to repression, and that the incompatibility is not to be understood as incompatibility between propositions. Thus, to say that the unconscious presentation has not taken propositional form seems to be a way of saying that it is not something for which the subject has justification, or for which he can give a reason-explanation. Further, to say that the impulse, or the idea which represents the impulse, is "incompatible with the ego" is not to say that the idea is propositionally articulated and that what it expresses in incoherent with other propositional attitudes of the subject. It is rather to say that the subject is yet not able to recognize this impulse as *his* in the first place.

In Gardner's rendering of Freud, "anxiety is the signal which causes repression, and it is triggered by ideas, not judgments."[338] Judgment is commonly understood as involving decision based on an evaluation of evidence. This is not the case with ideas. That which triggers repression, for Freud, is thus not a judgment based on beliefs, but something *prior to* judgment or perhaps even prior to articulation. The discussion of 'idea' above is intended to highlight something of uttermost importance for understanding self-deception, namely, that what one deceives oneself about is not an articulated belief or even a belief that one can articulate, nor any other propositional attitude which one has judged to be the case on the basis of the available evidence, but rather something which one senses but avoids acknowledging (for example, by forming a judgment or articulating for oneself that one senses). One could even say that it is characteristic of self-deception that one instinctively refuses to allow oneself any opportunity to become cognizant of that about which one has an inkling (by associations) but avoids confronting. The avoidance is not a *decision to avoid a belief which one holds*, as Davidson's account proposes. Self-deception is not a *decision* to avoid or deceive at all. Self-deception can thus be characterized as a reaction to a perceived threat, where one is unable and/or unwilling to acknowledge even that one feels threatened, much less why or wherefore. Self-deception, I suggest, is a form of flight from anxiety[339].

---

[337] Ibid.

[338] It will be recalled that Gardner holds that the neurotic should not be seen as exercising a preference. He means that this distinguishes it from self-deception.

[339] Freud's description of 'anxiety', as distinguished from 'fear' and 'fright', also lends support to the view that what one is anxious of is typically not something which one can express in a proposition, i.e. discursively: "'Fright', 'fear' and 'anxiety' are improperly used as synonymous expressions; they are in fact capable of clear distinction in their relation to danger. 'Anxiety' describes a particular state of expecting the danger or preparing for it, even though it may be an unknown one. 'Fear' requires a definite object of which to be afraid. 'Fright', however, is the name we give to the state a person gets into when he has run into a danger without being prepared for it […]" (Freud, *Beyond the Pleasure Principle* (1920), SE, vol. 18, p. 12.)

Before we move on, I will summarize the gist of my argument in this section. One of the most debated topics stemming from Freud's theory is how to understand unconscious mental content as distinct from conscious thought. I have criticized Davidson's and Gardner's accounts of self-deception as reasonable and intentional action, and have found that Freud's descriptions of motivated misapprehensions and defenses, such as repression, provide a more adequate basis for understanding the phenomenon. This entails that we also understand Freud's conception of the unconscious. Of the different ways in which Freud tries to work out the difference between conscious and unconscious mental content, I have focused on the difference between what is articulated or what one is able to articulate, and what is not. Freud's conclusion in the paper "The Unconscious" is that the difference between a conscious and an unconscious presentation is the organization of thought which articulation, or "the word", brings to the thought when it becomes conscious. The understanding of unconscious mental content as lacking in articulation and organization has great merits as opposed to, for example, the topographical account, since the critique which Lear directs against Two-Minds accounts also apply to Freud's understanding of the distinction between conscious and unconscious mental content as a spatial division. Unconscious mental content is, in Freud's later account, rather *qualitatively* different from conscious content in not being articulated, that is, in being relatively cut-off from propositional attitudes. In "showing" rather than "saying", the suggested spatial division proposed in the topographical account becomes unnecessary and irrelevant since what is unconscious doesn't come with a rationalizing structure, i.e. "an unconscious mind". As unconscious mental content is qualitatively different from conscious mental content, unconscious mental processes such as defensive reactions are qualitatively different from conscious one's, such as intentional actions. The thrust of the foregoing has been to show that this qualitative difference has far-reaching consequences for how we understand self-deception.

Does Self-Deception Involve Intention?

Are one's own intentions something that one can discover in retrospect when reflecting on one's own behavior? As we have seen, this is what Freud argues. In discussing Freud's interpretation of the woman's obsessional behavior as expressing an unconscious intention, I asked if therapy, when successful, should not rather be seen as helping the patient to further spell out the sense in her behavior rather than revealing the intention. I took issue with Freud's use of 'intention' because it seemed wrong to ascribe an intention *post facto* to someone's action of which she herself was unaware of having had. As we have seen, Anscombe does not want to use the term intentional for cases in which the subject cannot explain why he acted as he did, although he senses that he *should* be able to. I am inclined to agree with Anscombe on this point. It seems

more accurate to say that when a person comes to understand why she acted as she did, it is because she comes to understand better the *sense*, or meaning, of the action rather than because she has discovered an intention that was there all along.

Nonetheless, one might think that there is something about obsessive behavior and self-deception, for instance, which cannot be captured by 'sense', where Freud's notion of 'unconscious intention' may be significant. Sense is something that an observer can find in someone's behavior or the person herself can read into her own behavior in retrospect. If it is correct to say that analysis helps the patient make sense of an action, how does it do this? Should we think of "making sense of" the patient's actions as something which is similar to how a reader "makes sense" of a text, independently of the author's intentions? To do so would seem to suggest that there is no intrinsic sense to the act (or text), but that the sense arises in the act of analysis (or interpretation)[340]. If we think of making sense of the neurotic's behavior in this way, we do not regard the behavior as *someone's* behavior or action, as an expression of her feelings, motivations or intentions, which we can understand or fail to understand. This would be a problem; as a rule, we respect first-person authority with regard to one's own feelings, intentions etc. *even if* one can sometimes fail to recognize what one feels, is afraid of, etc. In ascribing an unconscious intention to the woman, however, Freud claims to have identified the intention with which the woman was acting all along, although she was at no point aware of herself as having that intention. Freud's writings are full of examples of how people understand their own behavior in retrospect and require help from an analyst to understand their behavior and themselves. In other words, they need help in understanding the very thing to which they are supposed to have unique privileged access, e.g., their intentions.

Freud is not just claiming something about the sense of the patient's behavior but also something about the patient herself: that she acted with an intention.[341] Freud's practice, accounted for in his case studies, shows that people do fail to understand what they do and why they do it, and that they can be helped to understand this in therapy. Therapy aims at laying bare this intention in making it conscious.[342] But why does he assume that the

---

[340] This is largely Lacan's view. See, for example, *Ecrites*.

[341] Repression and resistance is to be understood in this context as well. It is common that the patient strongly resists the intention that the analyst ascribes to her. Freud means that this resistance proves that a process of repression of the intention has taken place and that resistance is a re-awakening of the repression. (See, for example, Freud, *Introductory Lectures*, vol. 16, pp. 436.) We have seen that the assumption that there is a repressed intention (belief etc.) does not carry with it the assumption of an initial awareness that then gets rejected; repression refers to the cases in which one has never been aware of the intention to begin with. Nevertheless, repression involves unpleasant and/or anxiety-provoking thoughts or ideas.

[342] Freud often compares his therapy to archeology and its discoveries, for example, in "Studies on Hysteria", SE, vol. 2, p. 139. "[I]n this, the first full-length analysis of a hysteria undertaken by me, I arrived at a procedure which I later developed into a regular method and employed

identification of an unconscious intention in the patient's behavior is what is required for therapy to be successful? A patient could come to understand the motivations for her behavior, one might think, without recognizing that there was an intention behind it (or in it) all along, of which she was unaware. The therapist can help the patient to better understand his situation without identifying something like an intention – portrayed by Freud as a plan with a further aim – underlying the action. When Freud emphasizes his discovery of the hidden intention behind the action as the essential moment in therapy, it suggests that what is repressed is an identifiable element, such as a belief, an intention etc. But we do not need to see repression as a bracketing of something specific as much as a disruption of recollection and reflection that sets in as a reaction to something discomfiting. In his talk of the identification of the intention behind an action as the critical moment in therapy Freud falls back into thinking of unconscious motivations as parallel cases to conscious ones, i.e. to intentions.

Let us consider the case of the obsessive woman once more. In Freud's interpretation, the woman carries out the obsessive behavior with the aim of correcting the actual experience; this is the unconscious intention of the action. Although she can provide what Freud calls the sense of the action in observing that her action resembles what took place on her wedding night, she is prevented from understanding, and therefore from expressing, that she is repeating the scene in order to put it right. The repetition is carried out as obsessive behavior, and the recognition of it as an intentional action is rejected. Freud assumes that the fact that certain things are left out or altered in the woman's obsessive behavior in comparison with the real event show that the woman has an *unconscious intention to correct reality*. We saw earlier that Freud begins with the hypothesis that a speaker's intentions, of which he himself is unaware, can be inferred from what he says by the analyst in light of circumstantial evidence. What this means, in essence, is that the analyst is thought to be able to find a *purpose* in the action, or rather, a further purpose behind the agent's behavior.

It seems to me that this assumption about an intention is superfluous. The woman can simply be described as *reacting* to the anxiety and embarrassment that the recollection of the shameful moment of her wedding night provokes – without having a purpose or intention in doing so – by blocking out any discursive recollection of the real event by physically changing that reality. I suggest that this case does not need be understood as involving an unconscious intention to correct. Rather, the obsessive behavior shows the woman's preoccupation with this past event at the same time as it prevents her form recalling what was so distressing about it. The obsessive behavior can be

---

deliberately. This procedure was one of clearing away the pathogenic psychical material layer by layer, and we liked to compare it with the technique of excavating a buried city."

understood as instinctual rather than as an intentional action.[343] What I find problematic is not so much the description of the obsessive behavior, as including unconscious intention as how Freud accounts for this unconscious intention: as directed at correcting. I have criticized Gardner for holding that Anna Karenina's self-deception is directed at creating time for her and Vronsky to be together – at obtaining an advantage. But, one could object, is it not the case that the means of avoidance which Freud describes as defensive reactions are also advantageous, insofar as they allow the person to escape from anxiety? Perhaps, but the difference lies in that avoidance of anxiety is not for the sake of obtaining something *further*. In my interpretation, Anna does not deceive herself *in order to* create time for her and Vronsky's love to grow but *in* deceiving herself she avoids anxiety. Similarly, the obsessive woman is not carrying out her act *in order to correct* and redeem her husband's reputation, but *in* acting (or rather, acting out) she is, in some way, dealing with the painful memory. We can imagine that the memory of the wedding night will not be erased but will continue to haunt her.[344] However, she "cannot" accept what actually happened; her pride[345] doesn't want to know the truth. Caught between these "pressures", she reacts by acting out the event, altering the elements that are painful to recall. The problem, I hold, is that Freud accounts for this unconscious intention as if it were a conscious intention directed at a further goal.[346]

In his analysis of the obsessive woman, Freud holds that what she already knows is the *sense* of her action, but what therapy aims at revealing is the *intention* behind her action. Yet sometimes Freud expresses himself more restrictively, in line with my suggestion above. In his analysis of the case of the Rat Man, who Freud described as acting out conflicting feelings of love and hate, with no specific intentions in his behavior, Freud steered away from assuming an intention. Regarding different kinds of knowledge, he writes that understanding the *sense* makes the symptoms vanish, and that psychological work aims at bringing the patient to this insight. Freud seems to waver between two notions of the purpose of therapy. In the one, the goal is to uncover "unconscious intentions", i.e. real and determinate *causes* of behavior. In the other, the aim is to help the patient make sense of himself, his actions and his behavior. This wavering could, perhaps, be explained by Freud's ambition for psychoanalysis to attain the status of a *science* (the first interpretation) while he is at the same time quite aware of the difficulties that such an ambition

[343] As a defensive reaction (mechanism).

[344] This can be compared with the pressure of a forbidden desire in the case of obsessional neurosis that I discussed in footnote 250, p. 114. It cannot completely be silenced by repression but it expresses itself in the person's behavior, although in masked form.

[345] Or, ego under the pressure of the superego.

[346] In the appendix at the end of the book I consider Freud's analysis of fetishism as an illuminating analogy. I find it helpful in reflecting upon the evasive behavior of the self-deceiver as well as upon to what extent the self-deceiver is aware of that of which she deceives herself.

presents.[347] Freud's practice, accounted for in his case studies, shows that people do fail to understand what they do and their motivations for doing so, and that therapy can help them. But a problem for Freud is to explain *how* this is possible; how can he show that his practice actually results in producing the *right* explanation for his patient's behavior, and not merely in getting them to accept the explanation he offers? A task that Freud takes to be a central element in therapy is to identify the unconscious intention behind the patient's behavior and not to be satisfied with simply disclosing the sense. Maybe this is so important to Freud since the former would be a discovery, i.e., scientific to the extent that it yields new facts, while the latter seems simply to be an interpretation among various possible interpretations.

Self-Deception in the Light of Unconscious Intention

The difference between that which is conscious and that which is unconscious is defined in "The Unconscious", as we have seen, as being a difference between that which is conceptually formulated and that which lacks conceptual formulation. What repression denies to the rejected presentation is "translation into words which shall remain attached to the object": a repressed presentation lacks propositional form and conceptual connections. In the discussion of "The Unconscious", we saw that coming to awareness and conceptualization or discursive formulation, are two ways of describing the same thing, and I spoke above of the avoidance of the formation of a determinate belief or thought as a reaction to anxiety. According to Davidson and Gardner, self-deception begins with holding a belief '*p*' which self-deception aims at deceiving oneself about. According to Freud's suggestion that the difference between the unconscious and conscious is understood as the difference between that which does not have propositional form and that which has propositional form, this implies that the person is conscious at the outset of that about which he deceives himself. Self-deception is then a case of *deciding* to regard something that one has (conscious) knowledge of as not having happened.[348]

In this view, the self-deceiver knows what he fears and lies to himself in a process that is, to a high degree, transparent to him, which Gardner's claim that "the self-deceiver knows what he is up to" shows. While Davidson and Gardner depict self-deception as an intellectual act of manipulating beliefs that one already holds, Freud's discussion of illusion and delusion, on the other hand, suggests that the self-deceiver is strongly influenced already in his initial

[347] Richard Wollheim expands on the topic that Freud's accounts on mental abnormality must be understood in the light of his ambition to provide a general theory of mind. (Richard Wollheim, *Sigmund Freud,* Cambridge: Cambridge University Press, 1995.)

[348] Recall the tripartite division that I made in discussing the case of undoing where the normal case, according to Freud, is to *decide* to treat something as not having happened**.** (p. 121 and p. 125.)

apprehension and understanding by what he wants to be true, and, equally, by what he doesn't want to be true. The relation between illusion, delusion, psychiatric delusions and self-deception, suggested by the concept *Selbsttäuschung*, is ignored in Davidson's case, and denied significance in Gardner's. In Freud's discussion, the various defenses provoked by anxiety are reactions to something more basic than beliefs, namely, vague ideas, impressions and associations that threaten one's self-understanding. In this sense, the misapprehension involved in self-deception is best understood as the avoidance of a danger that has yet to be discovered. Even a case such as Carlos's, when he is confronted with the instructor's remarks, can be understood in this way; if the defense sets in already at the stage of apprehension, the instructor's judgment can mean almost anything, or nothing at all. Not wanting to know something about oneself or the world is not a judgment, but a disposition or attitude in the everyday sense. The psychoanalytical notion of "rationalization", one could argue, has become a part of ordinary parlance precisely because it is so familiar to everyone. We all recognize, both in ourselves and in others, a human tendency to want to avoid unpleasant truths.

Earlier, I quoted Gardner's summarized description of how Freud spells out the role of ideas in the characterization of repression in *Inhibitions, Symptoms and Anxiety*. Gardner says that the relation between ideas does not take the form of propositions but are relations of association.[349] If Gardner would allow self-deception to be seen as involving ideas and not beliefs, self-deception could be seen as not simply one's attitude towards something one knows, but already influencing the formation of knowledge. In Gardner's view, however, self-deception is a form of ordinary irrationality, and ordinary irrationality is propositionally transparent, that is, it is constituted and defined by a particular structure of propositional attitudes.[350] By defining his object thus at the outset, Gardner excludes the very possibility of seeing self-deception in terms other than that of beliefs.

Davidson and Gardner construe intention as a means in a rational process. Gardner says that the self-deceiver, but not the neurotic, knows what she is up to,[351] indicating that the intention in self-deception is conscious and that the process is transparent. Freud's use of 'unconscious intention' differs from

---

[349] Gardner says: "Identifying the material of repression as ideas makes it possible to understand how repression can occur without the instrument of conscious thought: in Freud's later theory [*Inhibitions, Symptoms and Anxiety*], anxiety is the signal which causes repression, and it is triggered by ideas, not judgments." See discussion in Chapter Three, p. 147.

[350] Chapter Two, p. 67.

[351] We recall: "We have seen that in order to explain irrational phenomena of a grade transcending ordinary psychology, psychoanalytic explanation attributes motives, such as hatred and its conflict with love, and assigns a function of self-misinterpretation. Although this gives neurotic symptoms the appearance of self-deception, the crucial ingredients of intention and preference are missing: *in so far as a person is self-deceived, she knows what she is up to, but in so far as she is neurotic, she does not*." (Gardner, p. 109.) See discussion Chapter Two, p. 89.

Gardner's and Davidson's in a number of ways, here I will discuss three that I take to be especially relevant to the present discussion.

1) Intentions can be unconscious. Freud says: "My interpretation carries with it the hypothesis that intentions can find expression in a speaker of which he himself knows nothing but which I am able to infer from circumstantial evidence"[352] A person's behavior can display an intention of which the person is unaware. Freud's analysis of the case of the woman who obsessively repeats the scene from her wedding night is an example of Freud's use of 'unconscious intention'.

2) Unconscious intentions occur in processes that are far from rational, such as in neurotic behavior. (I have, however, argued that Freud, in his ascription of unconscious intentions, tends to rationalize these acts and that this is a problem.)

3) Freud use of 'unconscious intention', I have suggested, can be replaced by 'unconscious motivation' or 'wish'. There is a motivation or wish that guides the person's thoughts and actions, but she is not herself aware of it.[353]

We have seen that one of the two important characteristics that distinguishes self-deception from motivated self-misrepresentation – or weak self-deception – in Gardner's account, is that self-deception is not simply motivated but intentional. Further, the self-deceiver is aware that she is manipulating her beliefs (she knows what she is up to), which seems to suggest that the self-deceiver is aware that she is deceiving herself. In contrast, Freud's use of unconscious intention does not involve manipulation of beliefs, but rather the inhibition of the process of belief-formation. When Freud claims that there are unconscious intentions in the neurotic action, it doesn't mean that the neurotic knows what she is up to (e.g. the woman does not know that she repeats what was happening on the wedding night *in order to correct it*).

In my view, it is better to refer to neurotic actions and self-deception as *motivated* rather than as guided by an unconscious intention, the most important reason being that the vocabulary of unconscious intention so easily leads one to read more control, transparency and "staging" into neurotic and self-deceptive behavior than we have reason to assume. Freud makes precisely this mistake in his interpretation of the unconscious intention, which he ascribes to the neurotic woman who repeats the event of the wedding night. I

[352] I discuss this quote on p. 130.

[353] Anna Karenina's sincere belief that she is displeased with Vronsky's pursuit of her can serve as an example: in order to avoid acknowledging for herself what she actually feels for him, she instinctively reacts by rejecting his advances. Holding the belief that she is displeased with his pursuit of her is motivated by the anxiety that her own feelings for him provoke; an anxiety which arises because she senses a threat to her entire world: her values, her self-image and her way of life.

take this move to be a consequence of Freud's vacillating between two conceptions of the point and purpose of psychoanalysis: as an interpretative craft primarily concerned with meaning or as a scientific discipline concerned with causal explanation. The notion of "unconscious intentions" is a product of the latter.

The Ego Organization and Exclusion

We will now consider another picture of the difference between the conscious and the unconscious than in terms of propositional formulation, that of the ego as an organization and of that what belongs to the id as outlaws. While in the former picture, the focus was on that which is unconscious, unformulated and excluded, it is now on the ego as an organization that *excludes*. It will be seen that the idea of the ego as an organization is analogous to Freud's characterization of the conscious/preconscious as having reached a higher level of psychical organization than the unconscious, as given in "The Unconscious". In the passages that I consider here, Freud accounts for the mental in the terms of id, ego and superego (the so called "second topology"). Like the division into conscious and unconscious, this tripartite division is posited by Freud as a fundamental assumption of psychoanalytic theory.[354] I am interested in this division as a way of representing competing characteristics (drives, values etc.) within a person and in his or her relation to the environment as far as it can be perspicuous. (I am equally aware that it *can* lead one astray.)

Freud discusses the strength and weakness of the ego in *Inhibitions, Symptoms and Anxiety*. He claims that the controversy over the strength or weakness of the ego arises because we take the division between the ego and the id too literally, when it is actually an abstraction. A consideration that, Freud says, makes it plausible to divide the ego from the id is that the ego is an organization while the id is not.

> We were justified, I think, in dividing the ego from the id, for there are certain considerations which necessitate that step. On the other hand the ego is identical with the id, and is merely a specially differentiated part of it. […] if a real split has occurred between the two, the weakness of the ego becomes apparent. But if the ego remains bound up with the id and indistinguishable from it, then it displays its strength. The same is true of the relation between the ego and the super-ego. In many situations the two are merged; and as a rule we can only distinguish one from the other when there is a tension or conflict between them. *In repression the decisive fact is that the ego is an organization and the id is not.* The

[354] The division of the psyche into what is conscious and what is unconscious is, Freud says, "the fundamental premise of psychoanalysis". According to Freud, these distinctions proved inadequate and insufficient, which is what motivated the introduction of the new distinctions *ego, id* and *super-ego*. See footnote 252, p. 114 above for a short account of Freud's metapsychological terms.

> ego is, indeed, the organized portion of the id […] As a rule the instinctual part which is to be repressed remains isolated. [355]

One of the main ambitions of Freud's psychoanalysis is to make that which used to belong to the id a part of the ego, that is, to make it fit in with the ego organization. This means to make the ego accept the wishes, desires, drives etc. of the id, to accept them as part of itself, as belonging to one's self; or differently put, to bring repressed thoughts and feelings into consciousness. But when the impulses of the id, described in the vocabulary of the second topology, lead the way and leave the ego behind to rationalize the morally problematic behavior, or when the super-ego is dominant and the ego follows its demands without acknowledging the desires and wishes of the id, the ego-organization, whose function it is to unify and communicate between the id and the super-ego, is weak and malfunctioning. Freud's vocabulary of id, ego and super-ego can be seen in the light of acknowledging the struggle within a subject as a struggle between fundamental needs of having one's desires and wants fulfilled, and creating and preserving oneself as a person with a direction in life and with certain values and responsibilities. In short, Freud's vocabulary is a way of talking about moral and existential struggles without reliance on traditional moral or metaphysical vocabulary, concepts and doctrines.

It is characteristic of Freud's style of writing that he uses many different ways of exploring and expressing an idea and that he often uses pictures and analogies for this exploration. Previously, we saw Freud spelling out his insight that what separates the conscious from the unconscious is that when something is conscious, words are connected to the object. I have described this by saying that that which is conscious is propositionally formulated, or discursive, i.e. possible to identify and express. In the picture outlined here, the ego is an organization that shuts out all elements that disrupts the system. Whatever cannot be integrated into its organization is left outside of it, which is to say, left unconscious. These are two different ways of representing the difference between what is conscious and unconscious, but there is an important similarity between them, namely, that what is conscious or belongs to the ego belongs to a *structure*. In the first picture, the conscious is seen as belonging to a discursive structure, which makes it open to reasoning and reflection. The second picture

---

[355] Freud, *Inhibitions, Symptoms and Anxiety*, p. 97. My italics. „Die Scheidung des ichs vom Es scheint gerechtfertigt, sie wird uns durch bestimmte Verhältnisse aufgedrängt. Aber anderseits ist das Ich mit dem Es identisch, nur ein besonders differenzierter Anteil desselben. [...] hat sich ein wirklicher Zweispalt zwishen den beiden ergeben, so wird uns die Schwäche dieses Ichs offenber. Bleibt das Ich aber mit dem Es verbunden, vom ihm nicht unterscheidbar, so zeigt sich seine Stärke. Ähnlich ist das verhältniss des Ichs zum Über-Ich; für viele Situationen fliessen uns die beiden zusammen, meistens können wir sie nur unterscheiden, wenn sich eine Spannung, ein Konflikt zwischen ihnen hergestellt hat. Für den Fall der Verdrängung wird die Tatsache entscheidend, dass das Ich eine Organization ist, das Es aber keine; das Ich ist eben der organisierte Anteil des Es [...] in der Regel bleibt die zu verdrängende Triebregung isoliert." (GW, Band 14, s. 124-125.)

contrasts the ego as an organization with that which is not.[356] Why would something be excluded from the organization of the ego? There can be a number of reasons, but in general one can say that the alien element conflicts with the person's values, self-image or world-view so that it can at best be approximated to the "organization" or structure, but not expressed.

These images are closely related. As a general description of the motive for repression, Gardner suggests "an impulse's 'incompatibility' with the ego, which is clearly concerned with something below the level of propositional inconsistency."[357] In other words, what is excluded from the ego is not a propositional attitude that is denied because it is rationally inconsistent with the ego, but something that is not yet formulated in a proposition. It thus seems that the denial cannot be on intellectual grounds. The idea, although it is not recognized or conceptualized, can evoke unpleasant associations, which the ego seeks to repress.

In the following sections of *Inhibitions, Symptoms and Anxiety*, where Freud discusses the weakness of the ego and the status of repressed material, he says that the weakness of the ego in relation to the other parts should not be taken as a cornerstone of psychoanalysis. To begin with, it is unclear what we mean by referring to the ego as strong or as weak:

> Although the act of repression demonstrates the strength of the ego, in one particular it reveals the ego's powerlessness. […] For the mental processes which had been turned into a symptom owing to repression now maintains its existence outside the organization of the ego and independently of it.[358]

> If the ego succeeds in protecting itself from a dangerous instinctual impulse, through, for instance, the process of repression, it has certainly inhibited and damaged the particular part of the id concerned; but it has at the same time given it some independence and has renounced some of its own sovereignty. This is inevitable *from the nature of repression, which is, fundamentally, an attempt at flight. The repressed is now, as it were, an outlaw; it is excluded from the great*

[356] The description of the ego as an organization, a structure, reminds one of the description Freud uses in "The Unconscious" where he distinguishes the conscious/preconscious from the unconscious by saying that the former has a "higher psychical organization". He further suggests that this characteristic is internally related to the other: that in the conscious/preconscious thing-presentations and word-presentations are linked. "The system *Ucs.* contains the thing-presentation of the objects, the first and true object-cathexes; the system *Pcs.* comes about by this thing-presentation being hypercathected through being linked with the word-presentations corresponding to it. It is these hypercathexes, we may suppose, that bring about a higher psychical organization […]" (Freud, "The Unconscious", p. 202) See my discussion on p. 144.)

[357] Gardner, p. 104.

[358] Freud, *Inhibitions, Symptoms and Anxiety*, p. 97. „Hat der Akt der Verdrängung uns die Stärke des Ichs gezeigt, so legt er doch in einem auch Zeugnis ab füt dessen Ohnmacht [...] Denn der Vorgang, der durch die Verdrängung zum Symptom geworden ist, behaubtet nun seine Existenz ausserhalb der Ichorganization und unabhängig von ihr." (GW, Band 14, s. 125.)

> *organization of the ego* and is subject only to the laws which govern the realm of the unconscious.[359]

Repression demonstrates the power of the ego, but it also shows the ego to be weak with respect to its function of mediating between and unifying the powers of a person. In repression, the ego fails in its role as mediator between the requirements of the super-ego and the wishes and drives of the id by ignoring the latter. It lets the demands of the super-ego dominate it.

In calling the repressed "an outlaw", Freud suggests, I believe, that the exclusion is not on intellectual grounds, but on moral ones. That which does not belong to the values and morality built into the organization of the ego – which is determined by the ego ideal, or super-ego – can evoke ideas and associations that make the ego anxious, and the ego can react by excluding it before it has tried to confront it on intellectual grounds (where it could judge and condemn it). This exclusion, I suggest, should not primarily be seen as the exclusion of something that is in and of itself incoherent, but rather as the exclusion of something that threatens the coherence or organization of the ego. In short, what is excluded is anything that threatens one's self-understanding. To return to one of our examples, when Anna Karenina deceives herself of her feelings for Vronsky, it is not simply because it is *incoherent* with the fact that she is married and a mother; it is because her feelings for him threaten her self-conception – they threaten everything that she takes herself to be. In that respect, they threaten the very existence of "Anna Karenina" as such. Naturally, such a threat evokes anxiety. Even if Anna Karenina cannot identify and acknowledge her true feelings, and is thus incapable of intellectually grasping that they pose a threat, she can still react instinctively to them. This is the respect in which self-deception should be seen as a kind of flight. Anna Karenina's feelings of happiness in the company of Vronsky, Vronsky's exclamations of love etc. hint at something forbidden and frightening. Anna Karenina has an inkling of something uncanny[360] and, rather than facing it and

---

[359] Ibid. p. 153. „Wenn es dem Ich gelungen ist, sich einer gefährlichen Triebregung zu erwehren, z. B. durch den Vorgang der Verdrängung, so hat es diesen Teil des Es zwar gehemmt und geschädigt, aber ihm gleichzeitig auch ein Stück Unabhängigkeit verzichtet. Das folgt aus der Natur der Verdrängung, die im Grunde ein Fluchtversuch ist. Das Verdrängte ist nun ‚vogelfrei', ausgeschlossen aus der grossen Organisation des Ichs, nur den Gesetzen unterworfen, die im Bereich des Unbewussten herrschen." (GW, Band 14, p. 185) The word „vogelfrei" is used to refer to someone who is being excluded from some priviledges of society because his behavior is deemed despicable.

[360] Freud's paper "The Uncanny" (*Das Unheimliche*) is of interest in reflecting upon self-deception. Freud describes the paper as an investigation of aesthetics, where this is understood to mean "the theory of the qualities of feeling", and he explores the feeling of *heimlich* (meaning: "belonging to the house", "not strange", "familiar", "tame", "intimate", "friendly", etc.) and its intimate relation to *unheimlich* (meaning: "secret", "untrustworthy", "hidden", "withdrawn from knowledge", etc.). A thought that runs throught the paper is that the "uncanny proceeds from something familiar which has been repressed" (p. 247) and that which is familiar yet uncanny can be an affect which is "in reality nothing new or alien" but "something which ought to have

trying to understand it better, she reacts by fleeing from the danger of psychological annihilation. In fleeing from her own feelings, she is, in effect, fleeing from herself. She can do this in different ways. She can ignore her feelings for Vronsky completely, she can rationalize them by seeing Vronsky as a good friend etc.

Davidson and Gardner understand self-deception to be an intentional action. Davidson holds that the action is directed at holding a belief which one wants to hold (although one knows that the total evidence speaks against it). In Gardner's theory, the action aims at fulfilling a morally problematic desire. As we have seen, Freud also sometimes accounts for behavior, the motivation or sense of which is not known by the subject as involving an intention of sorts, an unconscious intention, for example in his analysis of the woman who repeated the scene from her wedding night as an intentional action of putting it right. I have questioned Freud's use of intention in such cases and suggested that such behavior should be seen as a reaction or response to something painful, or to a recollection of something painful, rather than as an intentional action with a further aim, such as correcting the event so as to exculpate her husband. I will now turn to a passage in which Freud discusses pathological cases of symptom formation, since it displays the urge to think that a person's actions are always directed at obtaining something further (i.e., that it has a purpose that reaches beyond protecting oneself and preserving one's view of oneself) as well as how this conception is problematic. This passage says something important about the character of the ego – the person – in general.

In *Inhibitions, Symptoms and Anxiety*, Freud describes the ego as based on reciprocal influence between all its parts and, therefore, its very nature obliges it to synthesize, to bind together and unify. Thus, when repression fails in stifling disquieting instinctual impulses, the struggle is brought to an end by the formation of a symptom. At the stage of symptom formation, according to Freud, the ego presents two faces with contradictory expressions. One continues on the path of repression, but the other is determined by the fact that the ego is an organization. Faced with a symptom, the ego therefore attempts to restore or remake and incorporate these symptoms into its organization.[361] The agoraphobic can serve as an example: when her illness is severe, she adjusts her life and projects to the phobia; she might, for instance, organize her life so that she doesn't have to leave the flat. The symptom gradually becomes the core of the ego's organization. Regarding the ego of the obsessional neurotic, Freud writes: "It makes an adaptation to the symptom – to this piece of the internal world which is alien to it – just as it normally does to the real external world."[362]

---

remained hidden but has come to light" (p. 241). (Sigmund Freud, "The 'Uncanny'" (1919), SE, vol. 17).

[361] Freud, *Inhibitions, Symptoms and Anxiety*. p. 98-100.

[362] Ibid. p. 99. „Es findet eine Anpassung and das ichfremde Stück der Innenwelt statt, das durch das Symptom repräsentiert wird, wie sie das Ich sonst normalerweise gegen die reale Aussenwelt zustande bringt." (GW, Band 14, s. 126.)

Freud emphasizes that this should be seen as an *adaptation to the symptom* and not as a *creation of the symptom* (for some other interest).

> There is a danger, too, of exaggerating the importance of a secondary adaptation of this kind to a symptom, and of saying that the ego has created this symptom merely in order to enjoy its advantages. It would be equally true to say that a man who had lost his leg in the war had got it shot away so that he might thenceforward live on his pension without having to do any more work.[363]

In the eyes of an observer, and with some distance to the accident, it might appear as if the man lives a good and easy life because of his lost leg. The observer might notice that the man has received several advantages from this accident and think that, perhaps, it wasn't such a bad thing that he lost his leg. But should one really think that the accident was something the man had planned, that it was an intentional action with the aim of living on a pension for the rest of his life? For Freud, this would be to exaggerate the importance of adaptation to the circumstances. The same is true of symptom formation, in his view. The neurotic *can* enjoy some advantages from the symptoms or the pleasures that they can bring, such as being cared for by others or enjoying the satisfaction of performing the actions, but that doesn't mean that he *made* the formation of the symptom happen *in order to* be cared for, or because he preferred the pleasure of performing the substitutive action to the pleasure of performing the desired action. Symptom formation should rather be seen as an irrational and pathological instinctual reaction to strong instinctual drives that stand in opposition to strong demands of the super-ego. They are result of a conflict, and play a part in the defense of the ego, but they were not developed as means in an intentional strategy to obtain something further, something advantageous.

Just as Freud writes that the symptom is not created with the aim of obtaining something advantageous, I argue that self-deception is not an intentional action directed at coming to hold a belief or fulfilling a desire. Davidson and Gardner both hold that it is. In Davidson's analysis of the example of Carlos, self-deception is an intentional process directed at coming to hold a belief that Carlos wishes to hold, that he will pass the test, which will bring him relief from the painful thought that he will most probably fail. In Gardner's analysis of Anna Karenina, the intention, which is the motor of her self-deception, is to be with Vronsky and allow their love to grow. No doubt, it can *seem* that self-deception is directed at fulfilling a wish or obtaining something further. The result of self-deception can be that the desire that one

[363] Ibid. „Man kann die Bedeutung dieser sekundären Anpassung an das Symptom auch übertreiben, indem man aussagt, das Ich habe sich das Symptom überhaubt nur angeschafft, um dessen Vorteile zu geniessen. Das ist dann so richtig oder falsch, wie wenn man die Ansicht vertritt, der Kriegsverletzte habe sich das Bein nur abschiessen lassen, um dann abreitsfrei von seiner Invalidenrente zu leben." (Ibid.)

condemns morally – which one represses or in other ways prevents oneself from acknowledging – is fulfilled for, when these desires are repressed and, in Freud's expression, made "outlaws", they are not acknowledged and are no longer under the control of the ego. The person can therefore go on to act in a way that allows the desires be fulfilled without the desires or the fulfillment being recognized or acknowledged. But the fact that this can be the consequence of self-deception doesn't mean that there was an intention present in the self-deceiver to fulfill the morally condemned desire. Freud says that just as the man didn't have his leg shot off in order to be eligible for benefits, the obsessional neurotic doesn't develop symptoms in order to enjoy its advantages. Likewise, I argue, the self-deceiver does not intentionally mislead herself in order to have her desires fulfilled.

Discussion of Occurrences of 'Self-Deception' in *SE*

At the beginning of this chapter, I introduced the two German concepts for self-deception: *Selbstbetrug* and *Selbsttäuschung*. These concepts cast light on different aspects of self-deception. I turned our attention to the meaning of the latter, which I take to reveal aspects of self-deception which have been inadequately addressed or left unacknowledged in Davidson's and Gardner's accounts; moreover, I see Freud's works as dealing mainly with those aspects of self-deception captured by the concept of *Selbsttäuschung* rather than *Selbstbetrug*. Up to this point, I have looked at Freud's use and discussion of concepts related to self-deception, without focusing on the passages in which he discusses self-deception specifically. My aim has been to reproduce and further develop a broader and deeper context in which to better understand essential characteristics of self-deception. In the concordance to the English translation of Freud collected works, the word 'self-deception' occurs only three times. I will now look closely at two of these passages; the third I will mention briefly.[364] In the early text "The Psychotherapy of Hysteria", Freud uses the word 'self-deception' (*Selbsttäucshung*) in his discussion of the response the hysterical patient may have to a suggestion made by the therapist.

> If a pathogenic memory or a pathogenic connection which had formerly been withdrawn from the ego-consciousness is uncovered by the work of the analysis and introduced into the ego, we find that the psychical personality which is thus enriched has various ways of expressing itself with regard to what it has acquired. It happens particularly often that, after we have laboriously forced some piece of knowledge on a patient he will declare: "I've always known that, I could have

[364] I have chosen to look closely at these passages rather than the passages in the German texts where *sich betrügen*, *sich täuschen*, *Selbstbetrug*, *Selbsttäuschung* etc. occurs, on several grounds. No doubt, this investigation would have been more thorough if I had analyzed the occurrences of the German concepts as well. I choose to look at the few occurrences of 'self-deception' in order to give each instance adequate consideration.

> told you that before." Those with some degree of insight recognize afterwards that this is a piece of self-deception and blame themselves for being ungrateful.[365]

Freud says that it is a piece of self-deception (*Selbsttäuschung*) for the patient to believe that he has always known and could have told the therapist at any moment this "piece of knowledge". The patient fails to understand, or to admit, that therapy has revealed to him something that he didn't know before, or knowledge that has been repressed for some time. Freud says that the "psychical personality has various ways of expressing itself with regard to what it has acquired". The patient's claim to having always known is familiar as an example of what Freud terms *resistance*: instead of claiming that "he has always known", the patient can also refuse to accept what analysis confronts him with. Both are cases of refusal to admit that analysis has confronted him with something in himself, something of which he wasn't previously aware. One might wonder what the motivation is for the patient to react by claiming that he has always known. The plausible explanation is that it is shocking and unnerving for him to realize that he doesn't know his own motivations, and thus he reacts by denying this to be the case.

At the beginning of this paragraph, Freud writes about "a pathogenic memory", or, "a pathogenic connection", which has been withdrawn from consciousness. In discussing the status of pathogenic memory traces, he states:

> Not at all infrequently the patient begins by saying: "It's possible that I thought this, but I can't remember having done so." And it is not until he has been familiar with the hypothesis for some time that he comes to recognize it as well; he remembers – and confirms the fact, too, by subsidiary links – that he did really once have the thought.[366]

Freud acknowledges that there are cases in which the patient accepts the analyst's interpretation of an underlying thought but is never himself able to remember this thought. He asks what we should think of the status of that which was withdrawn from consciousness and revealed through analysis in these cases.

---

[365] Freud, "The Psychotherapy of Hysteria" (1895), SE, vol. 2, p. 299. "Ist eine pathogene Erinnerung oder ein pathogener Zusammenhang, der dem Ich-Bewusstsein früher entzogen war, durch die Arbeit der Analyse aufgedeckt und in das Ich einfügt, so beobachtet man an der so bereichterten psychischen Persönlichkeit verschiedende Arten sich über ihren Gewinn zu äussern. Ganz besonders häufig kommt es vor, dass die Kranken, nachdem man sie mühsam zu einer gewissen Kenntnis genötigt hat, dann erklären: Das habe ich ja immer gewusst, das hätte ich Ihnen vorher sagen können. Die Einsichtsvolleren erkennen dies dann als eine Selbsttäushung und klagen sich des Undenkens an." (GW, Band 1, s. 305.)

[366] Freud, "The Psychotherapy of Hysteria", p. 299. „Gar nicht selten sagt der Kranke zuerts: Es ist möglich, dass ich dies gedacht habe, aber ich kann mich nicht erinnern, und erst nach längerer Vertrautheit mit dieser Annahme tritt auch durch Nebenverknüpfungen, dass er diesen Gedanken wirklich einmal gehabt hat." (Ibid.)

> The ideas which are derived from the greatest depth and which form the nucleus of the pathogenic organization are also those which are acknowledged as memories by the patient with greatest difficulty. Even when everything is finished and the patients have been overborne by the force of logic and have been convinced by the therapeutic effect accompanying the emergence of precisely these ideas – when, I say, the patients themselves accept the fact that they thought this or that, they often add: "But I can't *remember* having thought it." It is easy to come to terms with them by telling them that the thoughts were *unconscious*. But how is this state of affairs to be fitted into our own psychological views? Are we to disregard this withholding of recognition on the part of patients, when, now that the work is finished, there is no longer any motive for their doing so? *Or are we to suppose that we are really dealing with thoughts which never came about, which merely had the possibility of existing, so that the treatment would lie in the accomplishment of a psychical act which did not take place at the time*?[367]

Let us tie this discussion to a case with which we are now familiar. The Rat Man, Freud claims, holds contradictory feelings towards his father: consciously, he loves him dearly, but unconsciously he harbors great anger towards him. The thought that Freud wants him to accept is the thought that he is angry with his father. In many passages, one can see that Freud, like Davidson and Gardner, is tempted to think that a thought, belief or intention is there before it is recognized by the person and that the thought is *like* the conscious thought. He then sees his own work in analysis to be to bring his patient to a point at which he can himself recall and recognize that underlying thought, belief or intention. In the quotation above, however, Freud can no longer see a *motive* for the patient to withhold recognition of the thought and wonders if we really are to assume that there *is* a thought there to withhold. If this is the case, then, at least sometimes, the work of analysis is not to help the patient remember and admit to having had thought, but rather to make the patient *come to have* a thought, or *come to hold* a belief by acknowledging the motivation for his behavior. Seen in this way, instead of uncovering what was already there, analysis helps to *form* the patient's understanding of his feelings, behavior etc. Freud's various approaches to this topic seem to depend on his ambivalence regarding how much one thought can differ from another (an unconscious thought from a

[367] Ibid. p. 300. My italics. The word '*possibility*' is highlighted by Freud. „Die aus der grössten Tiefe stammenden Vorstellungen, die den Kern der pathogenen Organisation bilden, werden von den Kranken auch am schwierigsten als Erinnerung anerkannt. Selbst wenn alles vorüber ist, wenn die Kranken, durch den logischen Zwang überwältigt und von der Heilwirkung überzeugt, die das Auftauchen gerade dieser Vorstellungen begleitet – wenn die Kranken, sage ich, selbst angenommen haben, sie hätten so und so gedacht, fügen sie oft hinzu: Aber *erinnern*, dass ich es gedacht habe, kann ich mich nicht. Man verständigt sich dann leicht mit ihnen: Es waren *unbewusste* Gedanken. Wie soll man aber selbst diesen Sachverhalt in seine psychologischen Anschauungen eintragen? Soll man sich über dies verweigerte Erkennen von seiten der Kranken, das nach getaner Arbeit motivlos ist, hinwegsetzen; soll man annehmen, dass es sich wirklich um Gedanken handelt, die nicht zustande gekommen sind, für welche bloss die Existenzmöglichkeit vorlag, so dass die Therapie in der Vollziehung eines damals unterbliebenen psychischen Aktes bestünde?" (Ibid.)

conscious one) and still be a thought. In discussing Davidson's and Gardner's accounts of self-deception, I have asked whether self-deception involves manipulation of *beliefs*, where beliefs are understood as propositional attitudes. Now, 'thought' is a term with a much wider application than belief, especially belief in the sense of a propositional attitude; this makes talk of 'unconscious thoughts' far less problematic than talk of 'unconscious beliefs', depending on what one ascribes to a thought. The assumed similarity between an unconscious fear, thought or wish and conscious beliefs as all being part of a network of ideas, beliefs, fears etc., which lend it rationality, leads to perplexities, as Lear points out.

The status of a repressed thought is a topic that Freud discusses in many contexts. I have already discussed the "solution" that he gives to this question in "The Unconscious", where becoming conscious of something is characterized as connecting "the thing" with "a word", i.e. its insertion into the realm of conceptualization. This suggestion fits well with Freud's suggestion above, that to make something which is unconscious conscious is to accomplish the formation of a thought. According to this suggestion, repression sets in *before* a thought is formed, as a rupture in the normal course of events. Psychoanalytic treatment would then not simply uncover the thought that was there but had been hidden from consciousness, but it would help the patient understand something that he had never really understood before; in short, it would help the patient think the hitherto not thought. Differently expressed, the thought was not just *pre*conscious before it was "revealed" but *un*conscious; lacking the psychical organization which comes with conceptualization and thus not yet accessible to consciousness.

In this early text, Freud is groping for an understanding of the status of something before it becomes conscious. It is not clear that he has settled on understanding it as a psychical act that has not been accomplished as a thought in "The Psychotherapy of Hysteria". Freud continues by saying that it is impossible to say anything about the state that the pathogenic material was in prior to the analysis until he has arrived at a clarification of his basic psychological views, especially on the nature of consciousness. At one stage, he points to the constancy of "thought" across different levels of awareness, seemingly suggesting that the psychical act (or thought) is the same prior to and after becoming conscious:

> It remains, I think, a fact deserving serious consideration that in our analyses we can follow a train of thought from the conscious into the unconscious (i.e. into something that is absolutely not recognized as a memory), that we can trace it from there for some distance through consciousness once more and that we can see it terminate in the unconscious again, without this alteration of "psychical illumination" making any change in the train of thought itself, in its logical consistency and in the interconnection between its various parts. Once this train

> of thought was before me as a whole I should not be able to guess which part of it was recognized by the patient as a memory and which was not.[368]

Clearly this can make one prone to the view that the thought is there all along, and at some point becomes conscious. While that idea itself might not cause any great difficulty, given that the sense of 'thought' can vary greatly, what is a problem is that becoming conscious, according to Freud above, "makes no change in the train of thought itself, in its logical consistency and in the interconnection between its various parts". Freud's expression "psychical illumination" suggests that what has become conscious is "illuminated". Freud seems to say that unconscious thought is just like conscious thought. In fact, I think that this is what he is proposing here in his attempt at understanding what unconscious mental content is like. But if we consider Freud's earlier suggestion that the thought was never actually there before therapy, except as a possibility which can be realized in and through it, shouldn't that mean that essential characteristics are lacking in the yet-to-be thought as compared to the conscious thought? Therapeutic success shows that when something becomes conscious and can be expressed, this alters the patient's perception and comprehension. This indicates that the conscious thought is fundamentally different from the unconscious one. My interpretation of the passage quoted above is that Freud is caught up in his perspective as an observer when he describes the constancy of thought across conscious and unconscious stages. It makes little difference to the *observer* in *his* apprehension of a patient's behavior if that which motivates the patient is conscious or not. Further, the observer can see consistency and connections in the behavior when the motivation is unconscious as well as when it is made conscious. The bringing into awareness of a motivation doesn't make much difference to the spectator. Therapeutic success, on the other hand, depends on what difference it makes to the *patient himself* that what was unconscious is brought out in conscious thought. That difference, I argue, only comes about because the conscious thought is fundamentally different from the unconscious one.

What does it mean for someone to become aware of something about himself, something of which he was previously unaware? When I become conscious of something, it can help me to see a whole range of things about myself differently, such as my behavior, reactions, and thoughts, but also decisions and choices that I have made. To the extent that I am aware of what I do and why I do it, I can take responsibility for my actions. From an observer's

---

[368] Freud, "The Psychotherapy of Hysteria", pp. 300. „Es bleibt wohl eine des Nachdenkens würdige Tatsache, dass man bei solchen Analaysen einen Gedankengang aus dem Bewusste ziehen und wieder im Unbewussten enden sehen kann, ohne dass dieser Wechsel der ‚psychishen Beleuchtung' an ihm selbst, an seiner Folgerichtigkeit, dem Zusammenhang seiner seiner einzelnen Teile, etwas ändern würde. Habe ich dann einmal diesen Gedankengang ganz vor mir, so könnte ich nicht erraten, welches Stück vom Kranken als Erinnerung erkannt wurde, welches nicht." (GW, Band 1, s. 306-307.)

perspective, it can seem as if that which I realize when I become conscious was there all along but unconscious, giving guidance and consistency to my actions. There are several difficulties connected with this idea, however. Most importantly, how can we account for the radical difference it makes for someone when he becomes aware of something if we assume that the thought has been there all along? We recall Anna Karenina's strong reaction when, deeply disappointed that Vronsky is not there as she had expected, she realizes that she is in love with him. Is the difference between unconscious and conscious just that when one is unconscious one has not realized a thought that was there all along? It seems rather more accurate to say that Anna Karenina has never had the thought that she loves Vronsky; she has never thought of her feelings for Vronsky as love. Her great disappointment reveals that meaning to her; it makes her see things differently. Of course, her feelings have been there for some time, and others might have thought of Anna Karenina as in love. But this doesn't imply that Anna has had the *thought* that she is in love with Vronsky all along. I find it helpful to recall a passage from Wittgenstein's *Lectures and Conversations* here, which I find fitting as a description of what the psychoanalyst does to a person's reflections and understanding. Wittgenstein says to his students:

> I very often draw your attention to certain differences, e.g. in these classes I tried to show you that Infinity is not so mysterious as it looks. What I am doing is also persuasion. If someone says: "There is not a difference", and I say: "There is a difference" I am persuading. I am saying *"I don't want you to look at it like that." (I am saying I want you to look at the thing in a different way.* –T)"[369]

I will not focus on the element of persuasion here since it would require some discussion to see in what way persuasion can be a part of psychoanalysis without running the risk of forcing views and interpretations onto the patient. I want instead to call attention to the remark "I don't want you to look at it like that", and the point of looking at things "in a different way". When Vronsky tells Anna that he loves her, he articulates what her self-deception has prevented her from articulating. Taken together with the disappointment that overwhelms her when Vronsky does not show up, this makes Anna, "persuades" her, to see things differently, i.e. to see that she is in love with Vronsky, with all that this insight entails. When she recognizes her feelings *as love* it makes a decisive difference for her *life*. The feelings may well have been there for some time, but they were just inarticulate, unidentified impulses. When Anna arrives at the thought, they receive a determinate *meaning*. In realizing that she is in love with Vronsky, Anna is exposed to the anxiety, existential qualms and bad conscience from which self-deception has relieved her, since choosing to be with him excludes the possibility of being with the other person that she loves, her son.

---

[369] Ludwig Wittgenstein, *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief* (Oxford: Basil Blackwell, 2007)§ 35, p. 27. Taylor's remark within brackets. My italics.

The insight implies that she must choose between her loved ones. Articulating her feelings for Vronsky is a turning point in the novel. After this evening, she will turn cold on Karenin and later leave him for Vronsky. She changes her life. Although being self-deceived about what she feels for Vronsky stood in the way of their having a life together, it let her enjoy her life with Karenin and her son as well as her cherished moments with Vronsky.

The extract from Freud's text "The Psychotherapy of Hysteria", which I introduced above first describes a case in which a patient comes to recall a thought which he once really did have and which analysis re-awakens, and, second, a case in which another patient accepts the interpretation, and where this, as in the former case, has a therapeutic effect, but the patient can't remember having thought what analysis suggest that he has though. Since there is no motivation for the patient to resist recalling the thought anymore, it seems that these cases are different. It suggests that, in the first case, the thought had been formed before it was suppressed, while in the latter, the repression took place before the thought was formed, preventing it from being articulated.

The difference which Freud points to in the extract – between being able to recollect a thought which one has had and suppressed, and not being able to recollect any thought but still being convinced of something as the right understanding of one's behavior and thoughts – are elements in self-deception as well. The former case is closer to Davidson's account of self-deception, although it is not clear that Freud's *thought* is best represented as a *belief*, i.e. that it should be understood as having reached the level of consciousness of a propositional attitude – nor is it evident that this thought persists throughout the repression, as Davidson claims of belief in self-deception. In Freud's example, it has not persisted on a conscious level. On the contrary, it would be a piece of self-deception for the patient to claim, "I have always know that and could have told you that before." Davidson says that the self-deceiver must know the belief very well in order to succeed in suppressing it, suggesting that self-deception is an ingenious, rational intentional action. The second of Freud's cases suggests that a thought had never been formed prior to the repression. Freud says that the patient himself can't remember ever having had the thought, but that he is nevertheless "convinced by the therapeutic effect accompanying the emergence of precisely these ideas."[370] The suggestion is accepted as expressing the sense of the patient's hysteria: it makes his hysteria understandable to himself. More importantly, in understanding what his hysteria is about – what it is an expression of – he is better equipped to take control over his life. In knowing his own mind, he achieves genuine agency. The therapeutic effect of grasping the sense of one's behavior consists in achieving this agency, whether or not what is grasped has been thought prior to the therapy.

[370] See p. 164.

Anna Karenina's case is analogous. Earlier, she had found other ways of understanding her feelings and actions, or simply did not reflect on them; struck by her own disappointment at Vronsky's not showing up, however, she is made to wonder at her own reaction. One might say that she becomes a question to herself. She has not *chosen* to arrive at the insight; it is thrust upon her the minute she allows the question, "why am I so disappointed?" to take hold of her. The moment that she realizes that she is in love and that she has been deceiving herself about her feelings, she can also make sense of why she so disliked Vronsky's proclamations of love despite her joy at seeing him; these conflicting attitudes *are* the self-deception, or part of it. As Gardner's words suggest, when she can think the thought "So that's why I told myself that I was displeased with Vronsky's exclamations of love", Anna Karenina comes to recognize that Vronsky's proclamations of love fulfilled her deepest desires, but in so doing simultaneously threatened to rock the very foundations of her world: her reputation, her role as wife and mother, her most cherished values, etc.

Before moving on to the other passage in SE where the word self-deception occurs, I want to round off the discussion of this first passage. Here self-deception consist in the patient thinking that he already knew what the analysis brought out and could have avowed it at any time. Self-deception, in the sense of ascribing knowledge to oneself that one doesn't possess, is thematized in the history of philosophy already by Plato. Socrates is a remarkable character because he does *not* deceive himself: he does not claim to know what he does not know. At the same time, the dialogues make clear that this admirable attitude is no easy thing. Nonetheless, from a Socratic point of view, one might question theoretical constructions that ascribe to the self-deceiver knowledge of that about which he deceives himself. Of the writers whom we have discussed, Davidson goes furthest in this direction when he holds that self-deception starts with a belief about which one wants to deceive oneself. Freud is critical of the patient's reaction of "I already knew", and calls it a piece of self-deception. On the other hand, in his own account of repression, Freud talks as if the patient, on some level, was in control of the situation. He uses expressions such as 'unconscious intention', and he sometimes spells out what is unconscious as if it had the same form or articulation as conscious knowledge. I have argued that his mistake is to ascribe more structure to the patient's behavior than what is there, and more control to the patient than she has.

Let us move on to the other explicit thematization of self-deception in SE, the text "The Psychogenesis of a Case of Homosexuality in a Woman". Here Freud characterizes ways in which a person can fail to know fundamental things about her sexuality.

> I cannot neglect the opportunity of expressing for once my astonishment that human beings can go through such great and important moments of their erotic life without noticing them much, sometimes even, indeed, without having the

> faintest suspicion of their existence, or else, having become aware of those moments, deceive themselves so thoroughly in their judgment of them. This happens not only under neurotic conditions, where we are familiar with the phenomenon, but seems also to be common enough in ordinary life. In the present case, for example, a girl develops a sentimental adoration for women, which her parents at first find merely vexatious and hardly take seriously; she herself knows quite well that she is very much occupied with these relationships, but still experiences few of the sensations of intense love until a particular frustration is followed by a quite excessive reaction, which shows everyone around that they have to do with a consuming passion of elemental strength. Nor had the girl ever perceived anything of the state of affairs which was a necessary preliminary to the outbreak of this mental storm.[371]
>
> It must be admitted that poets are right in liking to portray people who are in love without knowing it, or uncertain whether they do love, or who think that they hate when in reality they love. It would seem that the information received by our consciousness about our erotic life is especially liable to be incomplete, full of gaps, or falsified.[372]

Let me start with the second quote and then return to the first. At first glance, Freud's suggestion here that our erotic life is especially prone to falsification and a lack of clarity appears dubious. But Freud was of course describing 19th-century European bourgeois culture and society, where restrictions and taboos related to sexuality were very much a part of the individual's social and cultural identity and values. And as apprehension of one's desires (sexual or otherwise) is guided by what one can tolerate admitting about oneself, an unconscious self-censure naturally leaves its mark on one's self-understanding. Women's sexuality as such was taboo, and homosexuality all the more so. In such circumstances, it is not difficult to see that issues involving sexuality would be especially susceptible to psychological distortion. Although our attitudes

---

[371] Freud, "The Psychogenesis of a Case of Homosexuality in a Woman" (1924), SE, vol 18, p. 166. „Ich will die Gelegenheit nicht versäumen, auch einmal das Erstaunen darüber zu Worte kommen zu lassen, dass die Menschen so grosse und bedeutungsvolle Stücke ihres Liebeslebens durchmachen können, ohne viel davon zu bemerken, ja mitunter, ohne das mindeste davon zu ahnen, oder dass sie, wenn es zu ihrem Bewusstsein kommt, sich mit dem Urteil so gründlich darüber täuschen. Das geschiet nicht nur unter den Bedingungen der Neurose, wo wir mit dem Phänomen vertraut sind, sondern scheint auch sonst recht gewöhnlich zu sein. In unserem Falle entwickelt ein Mädchen eine Schwärmerei für Frauen, die von den Eltern zuerst nur als ärgerlich empfunden, aber kaum ernst genommen wird; sie selbst weiss wohl, wie sehr sie davon in Anspruch genommen wird, fühlt aber doch nur wenig von den Sensationen einer intensiven Verliebtheit, bis sich bei einer bestimmten Versagung eine ganz exzessive Reaktion ergibt, die allen Teilen zeigt, dass man es mit einer verzehrenden Leidenschaft von elementarer Stärke zu tun hat. Von den Voraussetzungen, die für das Hervorbrechen eines solchen seelischen Sturmes erforderlich sind, hat auch das Mädchen niemals etwas bemerkt." (GW, Band 12, ss. 294.)

[372] Ibid. p. 167. „Man sieht sich so genötigt, den Dichtern recht zu geben, die uns mit Vorleibe Personen schildern, welche lieben ohne es zu wissen, oder die es nicht wissen, ob sie lieben, oder die zu hasse glauben, während sie lieben. Es scheint, dass gerade die Kunde, die unser Bewusstsein von unserem Liebesleben erhält, besonders leicht unvollständig, lückenhaft oder gefälscht sein kann." (Ibid, ss. 295.)

toward, and ideas about, sexuality in general, and homosexuality in particular, have been liberated from many taboos, a situation such as Anna Karenina's can easily provoke self-deception even in our own time and culture. The situation of falling in love with someone while being in a relationship with someone else, especially if there is a child involved, is experienced as morally and emotionally difficult by most people, and can give rise to denial of one's feelings, to interpreting one's feelings in a way that avoids conflict with one's situation as married and with one's self-image as someone's partner and parent.

In the first of the quotes above, Freud says that sometimes we do not have the faintest suspicion of important moments in our erotic life, while at times we have become aware of these moments, but deceive ourselves thoroughly in our judgment of them. He gives as an example a girl who "develops a sentimental adoration for women" and knows that she is greatly preoccupied with these relationships but, as Freud says, "experiences few of the sensations of intense love", until an extreme frustration is followed by a reaction which makes her passion clear to those around her and to herself. Up until this breakthrough in her understanding, many states of affairs and their significance go unnoticed or are in essential ways concealed from her.

Here again, we can see an analogy in the case of Anna Karenina. Anna is aware of many of the moments that she has felt something intensely for Vronsky, and her self-deception consists in what she makes of them. In the case that Freud discusses, the girl experiences feelings of intense love first after she has suffered intense frustration. So too does Anna Karenina realize that she is in love with Vronsky only after her disappointment at not meeting him. In both cases, it would be inaccurate to say that Anna or the lesbian girl had learned something about what they already *knew*, but have chosen to suppress. Rather, in overcoming self-deception, they could see the meaning of their reactions and emotions for the first time. They were earlier aware of something, but now they know what it is. Only now can Anna say: "I believe I am in love with Vronsky!"

Think of how we come to see a rough sketch as a figure. After seeing only lines and curves on a piece of paper, these lines and curves come together in one's apprehension to constitute a figure. This simple picture can, perhaps, illustrate what happens when Anna Karenina comes to understand her condition as that of being in love with Vronsky. Everything was there before, but she didn't see the shape or pattern; all the elements were there in full view, as it were, but she didn't grasp what was there to be grasped. To be able to "put the pieces together", something had to make her look at the situation in a new way. In this case, it was the shock of her heated reaction of disappointment at not meeting Vronsky. The difference between Anna's self-deception and mere failure to apprehend something – anything – in one's surroundings that lies open to view is that self-deception involves a failure to see something which we would expect to fall under the heading of privileged access, or first-person knowledge, that is something which we know immediately, without requiring

recourse to observation and evidence.[373] Second, there is an element of *motivation* involved in self-deception. The self-deceiver, in some respect, does not want to know. As I have argued in opposition to Gardner, however, the motivation in self-deception is not typically to fulfill a wish – in Anna Karenina's case, Gardner proposes, the wish to be with Vronsky – but rather to avoid something which provokes anxiety without (yet) being known. Anna Karenina's reaction of believing that she is displeased with Vronsky when he proclaims his love for her can be seen as a reaction to something which, in combination with her real desires, her awareness of having changed social circle etc. makes her (unconsciously) anxious. The unconscious motivation for self-deception is then rather to avoid dealing with the anxiety and moral confusion of facing the consequences of the clash between her wants and needs, on the one hand, and her highest values, on the other. Rather than being motivated by fulfilling the wish to be with Vronsky, as Gardner claims, I argue – in line with Freud's characterization of the defensive reactions – that Anna Karenina's self-deception can be understood as a reaction to the whole situation she finds herself in with Vronsky, a situation which makes her anxious without her yet knowing why. Her self-deception, I suggest, is motivated rather by the need to feel in control, to feel whole, and to see herself as a good, responsible human being. As we have seen in Freud's discussion of the ego as an organization, a person can react to a perceived threat by repressing, i.e. by excluding from its organization that which it doesn't want to recognize as belonging to itself.

I have discussed two of the three occurrences of the word 'self-deception' in SE. The third occurrence is in the text "From the History of an Infantile Neurosis", in a section where Freud discusses and justifies the psychoanalytic method (SE, vol. 17, p. 53). I will not examine this section here since, in my judgment, it doesn't contribute much new to the discussion of self-deception. Instead, I will turn to *Civilization and its Discontents* to look at a remark that Freud makes on delusion (*Täuschung*). Here Freud's target is illusory ways of understanding the world, through which what is painful or unwanted can be avoided. He says that while "[t]he hermit turns his back on the world and will have no truck with it," one can "build up in its stead another world in which its most unbearable features are eliminated and replaced by others that are in conformity with one's wishes."[374] This is a delusion, but, says Freud, "each one of us behaves in some one respect like a paranoic, corrects some aspect of the world which is unbearable to him by the construction of a wish and introduces

[373] It is true of many, but not all, cases of self-deception that it involves that to which one has privileged access. A more general description is the second; that self-deception involves something that *matters* to oneself, psychologically and morally.

[374] Freud, *Civilization and its Discontents* (1930[1929]), SE, vol. 21, p. 81. „Der Eremit kehrt dieser Welt den Rücken, er will nichts mit ihr zu schaffen haben. Aber man kann mehr tun, man kann sie umschaffen wollen, anstatt ihrer eine andere aufbauen, in der die unerträglichsten Züge ausgetilgt und durch andere im Sinne der eigenen Wünche ersetzt sind." (GW, Band 14, s. 439.)

this delusion into reality."[375] He continues a little later: "No one, needless to say, who shares a delusion ever recognizes it as such."[376]

I mean that it is essential for understanding self-deception, as well as other illusory ways of seeing the world, that it is often an avoidance of unbearable facts about oneself. This can involve "correcting" criticism, the motivation behind one's action, etc. in rationalizations such that one's self-image remains intact. I objected above to Gardner's interpretation of Anna Karenina's self-deception as serving to fulfill the wish to be with Vronsky. I share Freud's view that delusion and self-deception involve correcting or avoiding some aspect of the world or oneself by taking things to be as one wishes them to be. Thus while I want to say that wishes play an important role in self-deception, I reject the claim that self-deception consists in intentional action directed at fulfilling a wish.

Is One Always Unconscious of that about which One Deceives Oneself?

Although I have limited myself to a small representative selection of examples of self-deception, they should suffice to show that the term 'self-deception' can mean rather different things. Carlos' case, for instance, is of another order than Anna Karenina's. At the same time, for the term 'self-deception' to have any sense, there must be some evident similarity between instances. One question that needs to be addressed is to what extent and in what way the self-deceiver can be said to be conscious of whatever it is that he is deceiving himself about. Is the degree of awareness relevant for whether or not we are dealing with a case of self-deception?

One similarity that all our examples share is this: something that stands in conflict with the person's self-understanding or ideals is forced upon him. In the cases of Carlos and Freud's ungrateful patient (who claimed that he had always known), the foreign element is a comment made by someone else. Moreover, this comment is a judgment or an interpretation, which that other person makes about Carlos/the patient. The comment cannot simply be ignored – or rather, if it is ignored, this is also a defensive reaction. The instructor's discouraging comment implies criticism to which Carlos reacts by dismissing or disregarding its importance. Freud's interpretation of the sense of the hysterical patient's behavior likewise provokes a defensive response; the patient's reaction is to deny that this sense was unknown to him and claim that he could have told Freud this himself without the aid of therapy. Freud calls

[375] Ibid. „Es wird aber behaubtet, dass jeder von uns sich in irgendeinem Punkte ähnlich wie der Paranoiker benimmt, eine ihm unleidliche Seite der Welt durch eine Wunchbildung korrigiert und diesen Wahn in die realität einträgt." (Ibid. s. 440.)
[376] Ibid. „Den Wahn erkennt natürlich niemals, wer ihn selbst noch teilt." (Ibid.)

this a piece of self-deception. Even in the case of Anna Karenina's self-deception, the words expressed by someone else are significant; Anna reacts to Vronsky's exclamations of love with displeasure. I have suggested that Vronsky, by telling Anna that he loves her, elicits an instinctive reaction by way of which Anna protects herself from the insight that she is in love with him, an insight which threatens to call into question fundamental beliefs and what she values greatly. The defensive response is a central element in her self-deception. The psychological concept of resistance is applicable here: Anna Karenina's reaction can be compared with how the patient reacts with resistance when something that he has repressed is about to be revealed in analysis. In the case of the girl's understanding of her adoration for women, what is forced upon the girl is not a comment made by someone else. Pent-up desire makes her frustrated, and an untamed passionate reaction comes over her and exposes what previously expressed itself in a premature, or sublimated, form in her sentimental adoration of women.[377]

In my interpretation of these cases, the self-deceiver is neither aware of deceiving himself (herself) nor of what it is that he deceives himself about. Anna Karenina is not aware of being in love with Vronsky; the girl who adores women is not aware of being a lesbian; and the patient is not aware that his claim to have known all along and that he could have told Freud at any moment is a piece of self-deception (if he were, it would be a straight out lie). What about Carlos? The fact that he has failed twice and that he has heard the instructor say discouraging things doesn't add up to Carlos' being aware that all things point to failure; to the extent that he *is* aware, he is immersed in the illusion that he will, nevertheless, surprise the instructor and pass. In my interpretation of these cases of self-deception, when the person recognizes what it is that he has deceived himself about, the deception dissipates. Self-deception ends where insight begins. Insight is the end of self-deception.

Must self-deception look like this? Is it always avoidance of coming to know something that threatens one's self or one's self-understanding? If this is so, self-deception is always manipulation of something that is unconscious or preconscious but not acknowledged, and it ends when it becomes conscious. I want to consider a couple of examples where someone deceives himself about something of which he is aware, and even, to an extent, knows to be problematic. Imagine a case in which a close friend has told me a secret and told me to tell no one. I promise, but in spite of the promise I share her secret with an acquaintance of mine. I am aware of having broken my promise, and I do think that breaking promises is a bad, immoral thing to do. But I tell myself that this case is different. As a matter of fact, it served a good cause: by sharing my friend's story with my acquaintance, I was better able to give her good advice. I tell myself this to ease my conscience, and stop myself from

[377] The moment of insight in the case of Anna Karenina is similar to this.

remembering that I didn't actually break my promise with a good cause in mind; I just couldn't resist sharing this sensational story.

In this case, I know that I have broken a promise and I do initially feel that my action is morally dubious, but I justify it to myself by finding rationalizations in the light of which my telling appears morally justified, or at least justifiable. I am initially aware of that about which I deceive myself, and I also feel the sting of shame. I ease the shame by telling myself that I have nothing to be ashamed of, that I haven't acted wrongly.[378] I manage to do this by building up a context of rationalizations in which my action appears morally acceptable to me. This case is different from the examples above. It is rather close to Davidson's analysis of Carlos, in the sense that insight does not mean the end of self-deception. Even if I at some point see clearly that I did not actually break my promise with the good intention of helping my friend, nothing precludes the rationalizations from setting in again whenever pangs of guilt become too distressing. Rather, this case starts with the insight that what I did was probably wrong, by my own lights, which is why I ever so quickly push away the thought by rationalizations and illusory good causes. We should here recall Freud's discussion of delusion in *Civilization and its Discontents*, in which a person builds up a world in which its most unbearable features are eliminated and replaced by others that are in conformity with one's wishes. This bears similarity to a passage in *Psychopathology of Everyday Life*, which I discussed at the beginning of this chapter. One of Freud's patients had told Freud that her children were bed-wetters, she denies both the fact and ever having mentioned it. Freud holds that it is because this is a shameful fact for her to recall that she forgets. He refers to Nietzsche's famous quote from *Beyond Good and Evil* in discussing the latter example, but it is equally fitting for the case of the broken promise of secrecy. "'I have done that', says my memory. 'I cannot have done that' – says my pride, and remains adamant."[379] There is a memory; I am aware of having done or said something. It is obviously something shameful, since my pride cannot tolerate acknowledging that I have done it. The refusal to face up to one's action – pride's victory – can take different forms: suppression of the memory in which what one has done is made unconscious, rationalization of the action, isolation, in which the memory of the action is freed from the evil intention, etc.

I have added this last example, of someone rationalizing the fact that she committed an action that she would typically see as morally wrong, to show

---

[378] In the paper "Deceiving Oneself Or Self-Deceived? On the Formation of Beliefs "Under the Influence", Ariela Lazar discusses the influence of emotions on formations of irrational beliefs. She argues that self-deceptive beliefs are direct expressions of the subject's wishes, hopes, fears etc. and that self-deceptive states, therefore, are a kind of fantasy. (Ariela Lazar, "Deceiving Oneself Or Self-Deceived? On the Formation of Beliefs "Under the Influence", in *Mind*, Vol. 108. April 1999, Oxford University Press, 1999.)

[379] Friedrich Nietzsche, *Beyond Good and Evil. Prelude to a Philosophy of the Future*, (London: Penguin Classics, 1990). IV, § 68.

more of the spectrum of kinds of self-deception. I want to include cases where one is aware of that about which one deceives oneself and recognizes it as morally problematic. I further want to note that the example of Carlos could be seen in this light as well. It is a plausible interpretation that Carlos has at some point worried that he will not pass the test. But even so, I hold, the "denial" of this thought is not by means of rational reasoning directed at misleading oneself about what one takes as evidence. The denial rather has the form of suppression; for a moment Carlos is aware of the danger that he might not pass the test as a real threat (this is different from knowing that evidence point to failure) and "hides" it under quieting rationalizations which lets him view his situation in a more positive light. Further, the examples of self-deception that I have discussed previously, where the self-deceiver is not aware of that of which he deceives himself nor of deceiving himself, can turn into self-deception where the self-deceiver is conscious of that about which she deceives herself, and even that it involves a moral issue. Let me take Anna Karenina as an example. At the end of the quote that I have been discussing at length, Anna Karenina realizes that "his pursuit was not only distasteful to her, but was the whole interest of her life." I have said that her self-deception ends with this insight that ceases when she becomes aware of this fact. But her self-deception could change form here. After realizing that Vronsky is all that she wants, she can deny the truth of the insight. She can, for example, rationalize her strong emotional response to her disappointment – perhaps under the heavy pressure of social norms and self-expectations – as an over-reaction provoked by her feelings of loneliness and despair at the time. Of course, she is not in love with Vronsky! What really matter to her are her dear little son and loving husband. With this scenario in mind, it is difficult to judge wherein the self-deception lies. Was she deceiving herself in believing that she loved Vronsky, or is she deceiving herself in denying it? Or is it an illusion to believe that there is an either/or answer to be given here? Gardner's discussion of the passage cannot do justice to the complexity of Anna's situation without recapitulating the novel. As the drama of the novel unfolds, her love for Vronsky and her desire for a life with him, and her love for her son and her need to be with him, create an untenable and unbearable situation for Anna Karenina. Here *ambivalence* would be an appropriate term; Anna is torn between two competing desires that cannot both be fulfilled. She must choose one. Clearly, there are situations in which one doesn't know what one wants or feels, or in which one wants and feels contradictory things. These might not be cases of self-deception, strictly speaking; they are perhaps better described as being "at sea". In a desperate attempt to understand, one might drift from one grasp of the situation to the other. If the concept self-deception does apply here, it would seem to refer to the conviction that there is, and must be, *one* "true" answer to be found, one and only one way in which it is right to act.

Self-Deception as a Moral Concept

Self-deception can take many different forms, and different cases of self-deception, or different forms of self-deception, will be placed very differently on the scale unconscious – conscious. I have focused on self-deception as *keeping* oneself unaware. I focus on these cases largely to shed light on an aspect of self-deception which Davidson's and Gardner's accounts fail gravely to describe, cases in which one is driven by unconscious motivations and where self-deception is a reaction to anxiety – where, in contrast to cases involving fear, the object or reason is unknown or unconscious. I have tried to show that self-deception begins already in keeping oneself from becoming aware of something, rather than being an escape from something that one knows. Freud's texts are particularly well-suited to cast light on self-deception as a flight from something which is not yet conscious, but which causes anxiety thanks to his numerous accounts of the unconscious, anxiety, defensive reactions, etc. I have argued that self-deception is a flight from anxiety. I take this to be one of the most fundamental and general descriptions that can be given. A great many cases of self-deception can be seen as a flight from something that (one knows) could be in one's control and for which one should take responsibility, but which one is either unwilling or unable to confront and/or conquer.

What someone deceives himself about is often something that he does not want to believe about himself, or a complexity in his feelings or beliefs which he avoids facing. In this respect, self-deception should be understood as a *moral* concept, since, as my analysis of the cases that I have considered show, the context for deceiving oneself is moral. It is because of one's values, concerns, commitments and expectations on oneself to live up to them – that one has the potential and therewith the obligation to do so[380] – that one sometimes cannot face the fact that one fails to think and act according to these values.

When self-deception is understood as a flight from anxiety where anxiety arises because one senses that one fails to act rightly, that one's values and desires are in conflict, or because of other personal failures or flaws, the opening lines of Davidson's paper "Deception and Division" are almost startling:

> Self-deception is usually no great problem for its practitioner; on the contrary, it typically relieves a person of some of the burden of painful thoughts, the cause of which are beyond his or her control. But self-deception is a problem for philosophical psychology. For in thinking about self-deception, as in thinking about other forms of irrationality, we find ourselves tempted by opposing thoughts [...][381]

[380] Or, at least one *thinks* that one has that potential and therefore *requires* oneself to live up to it.
[381] Davidson, "Deception and Division" in *Problems of Rationality*, p. 199. See Chapter One, p. 59.

Self-deception is most definitely a problem for the practitioner when it arises in a moral context, although it does also, at least temporarily, relieve a person of the burden of painful thought. Self-deception, insofar as it consists in the avoidance of facing up to one's moral failures, is not only a real problem in real life, but a very deep one at that. It implies a kind of ethical and intellectual lethargy with regard to one's values and responsibilities as a human being. Further, and more concretely, self-deception can keep one from achieving one's goals. Take Carlos as an example: because he deceives himself of his chances of passing the driving test, he might not study much as he would need to in order to have good chances of passing. To be caught up in wishful thinking and self-deception regarding one's own abilities can be to fail to take control of something that one can control. Surely this, if nothing else, is a problem for the self-deceiver. I object to the claim that "self-deception is usually no great problem for its practitioner", since I take the moral context to be essential for many – or even most – cases of self-deception. Davidson and Gardner fail to see self-deception as a moral concept, as arising out of the context of values, principles, obligations to oneself and others, etc. which makes being human so difficult. To fail to acknowledge the moral context of self-deception, I argue, is not just to fail to describe a certain *dimension* of self-deception, but to fail to account for self-deception as a real problem, and not just a theoretical one.

Concluding Remarks *& Summary*

Remarks on Guiding Assumptions and Aims

In this last chapter, I wish to comment upon two sets of problems that I take to be present in all three accounts that I have discussed in this book. My aim is to accentuate certain central issues in previous discussions, and further analyze them. First, I will claim that problems arise in all three accounts because of certain presuppositions and purposes implicit in the framework in which they are developed, which influence the description of certain phenomena. Second, I will argue that Davidson, Gardner and sometimes Freud, misrepresent self-knowledge and self-deception in various ways because they fail to acknowledge that first-person knowledge (awareness) of one's mental states is *different* from knowledge of objects or of other peoples' mental states. Here I will rely heavily on Richard Moran's analysis of the difference between self-knowledge and knowledge of things. I will start with the first topic.

How do theoretical presuppositions and aims of the theorist give rise to problems in the analysis of self-deception and the paradoxes associated with it? In the first chapter, I presented Mark Johnston's critique of Davidson's interpretive view, i.e. the view that there is nothing more to being in a mental state than being apt to have that state attributed to one within an adequate interpretive theory. On this view, someone's behavior is regarded as evidence for his mental states, that is, a certain behavior is taken as evidence that someone is intentionally misleading himself about beliefs that he holds. I argued, in line with Johnston, that Davidson's account of self-deception is largely informed by the over-arching theory of radical interpretation. The cornerstones of the *Principle of Charity* – correspondence and coherence – are seemingly threatened by self-deception, and Davidson's account is intended to explain self-deception in such a way that the idea of correspondence and coherence can be maintained. These presuppositions are reflected first in Davidson's assumption that the self-deceiver holds the belief which he has best reason to hold, i.e. that his initial belief corresponds to the facts, and, second, that the two contradictory beliefs which go into self-deception are propositional attitudes, each connected to other propositional attitudes, which, taken together, make up a coherent, rational whole. I claimed that the paradox of self-deception, to which Davidson responds with the hasty suggestion that the mind of the self-deceiver must be viewed as divided, is a consequence of assumptions made in his theory of self-deception. Further, I argued that these assumptions emanate from his account of radical interpretation.

In Chapter Two, I showed that Gardner's account of self-deception is equally infused with his initial assumption that self-deception is a form of ordinary irrationality that ought to be distinguished from forms of irrationality treated by psychoanalysis. I argued that Gardner's characterization of self-deception as an intentional action directed at obtaining a further goal, involving

the manipulation of desires and beliefs as means, follows from his thesis that self-deception ought to be accounted for in rational terms; it is the result of his way of *mapping* self-deception, and it gets in the way of the attempt to provide an unbiased description. I have argued that Davidson's and Gardner's accounts of self-deception are imbued with implicit and explicit presuppositions, such as, for instance, that self-deception is analogous to deception, that it is a manipulation of beliefs, that it is directed at obtaining a further goal, etc., and I have discussed the problems these presuppositions give rise to in their accounts.

The main ambition in Chapter Three was to present *another* view of self-deception, in which self-deception is not portrayed as a lie to oneself, and where there is no general, or qualitative distinction made between the normal and the pathological. In further exploring how one can understand that about which one deceives oneself, if it is not a belief, a propositional attitude, I closely examined passages in which Freud discusses how we are to understand that which is repressed. He suggests that we understand it as an idea rather than a full-fledged belief, that is, as something that is not clearly articulated; as lacking in psychical organization. In search of ways to analyze self-deception other than self-deception as an intentional action directed at obtaining something or fulfilling a wish, I presented Freud's study of defensive reactions. I argued that they are not intentionally directed at obtaining something further, but that they are ways of escaping anxiety, and thus that self-deception ought rather to be seen as a flight from anxiety than as a lie to oneself.

Although Freud's texts reveal many important aspects of self-deception and related issues, some of the passages that are of relevance to self-deception are problematic. His analysis of the neurotic woman's repetition of the scene from the wedding night is a case in point. Freud claims that there is in her act an unconscious intention present, to correct her husband's impotence so as to exculpate him and avoid embarrassment. Freud is not satisfied with understanding the woman's actions as repetition of the event of the wedding night in which she is trying to come to grips with her sense of shame and the unpleasant facts, but argues further that an unconscious intention that motivates her act, is present all along. In my view, Freud interprets her act as being more rational and strategic than it really is. I argue that Lear's critique of Freud's interpretation of the Rat Man's fearful behavior also applies to this case: just as the Rat Man does not have a reason for his behavior, the woman's motivation is not an intention directed at achieving something.

What gives rise to the inclination to over-rationalize the act of the self-deceiver or neurotic in the cases of Davidson and Gardner, I have argued, is the underlying presupposition that it must be possible to give a rational interpretation of such cases. In Gardner's explanation, self-deception *is* a "rational" form of irrationality. Freud's motivation is different. Freud's agenda is not to offer a rational description of the act or to claim that the woman's behavior is a rational act. His motivation is rather, I have suggested, to identify *the real cause* of her behavior, that is, the intention on which she acted. Freud's

interpretation of the case suggests that analysis *reveals* the intention that was there all along; helping the patient to articulate the intention does not add anything to it. In my view, Freud's claim that he has identified the real cause, or, laid bare the intention that was there all along, is an expression of his aim to *prove* that his interpretations are correct. He claims that he has *discovered* the hidden intention. In other places, as I have showed, Freud describes the becoming conscious of something, the articulation, as making a difference to the feeling, thought, intention, etc., that is made conscious. No such suggestion is made in the analysis of the neurotic woman's action.

To summarize, all three accounts are heavily influenced by the theoretical apparatus in which they are formulated. I have discussed some of the ways in which the assumptions and aims involved leave their mark on the object described.[382]

Self-Knowledge and Morality

In his book *Authority and Estrangement: An Essay on Self-Knowledge*, Richard Moran asks: in what ways, if any, do attitudes towards oneself exhibit any systematic difference from attitudes towards someone else? He emphasizes that there is a basic difference between how a person may know his own mind and how he may know the mind of another:

> while a person may learn of somebody else's beliefs or other attitudes from what she says and does, he may arrive at knowledge of his own attitudes in a way that is *not* based on evidence or observation of himself. In this sense, a person may know his own mind "immediately", yet nonetheless declare his belief with an authority that is lacking in anyone else's observation-based description of him.[383]

Moran claims that immediacy of self-knowledge is a *fundamental* form of self-apprehension, and that it belongs to the concept of a person that he should be able to achieve knowledge of his attitudes in this immediate way.[384]

This asymmetry between knowledge of someone else's attitudes and knowledge of one's own is commonly discussed under the topics of first-person knowledge or first-person authority as distinct from knowledge of things or of other people's mental states. Moran discusses the different ways in which accounts of self-knowledge fail to pay attention, or do justice, to this distinction: "*prominent accounts of self-knowledge often end up either describing*

---

[382] Interpretation, whether in the context of intellectual theorizing or in personal reflection, lends itself to being influenced by one's guiding aims and assumptions. An example of the latter is the Rat Man's interpretation of his cringing as a fear of being beaten by Freud, discussed by Lear.
[383] Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001), Preface, xxix.
[384] Ibid. Preface, xxx.

something that could just as well be a third-person phenomenon, or transposing an essentially third-person situation to some kind of mental exterior."[385] Moran indicates a problem that I will argue attaches to all three accounts investigated here. A most striking example of transposing an essentially third-person situation to the mental exterior of a single person is to account for self-deception as analogous to deceiving someone else, as Davidson does. In so doing, Davidson ascribes to self-deception the characteristics of deceiving someone else, such that deceiving is an intentional action with a purpose or goal of making the other (oneself) believe what is false. I have already discussed the problems that arise in his description of self-deception as analogous to deception.

In my discussion of Davidson, I argued that evidence does not come into the picture since the self-deceiver evades any possibility that there might be to apprehend threatening facts and take them as evidence. In taking Moran's discussion on first-person authority into account, we see that not only do we have reason to question that self-deception, as a *failure* of self-knowledge, involves evidence; Moran argues that evidence typically does not play a part in self-knowledge at all. It is because one is led astray by the superficial analogy between self-knowledge and knowledge of objects, or knowledge of others, that one takes a person to judge from evidence that he is happy, in love, etc., while in reality it is uncommon that one becomes aware of one's feelings (etc.) in this way. As Moran notes, knowledge of one's own attitudes is not typically observational or reached by inference from observational knowledge or other people's statements; it is not based on evidence at all, but it is immediate experience. Still, although knowledge of one's own mental states is *normally* immediate, one *can* fail to know one's own feelings, desires etc. and can become aware of them in much the same way as one becomes aware of someone else's attitudes and views. I will argue that Gardner and, at times, Freud misunderstand what failure of self-knowledge is, and that Moran is helpful in showing in what way they do so. Davidson, for whom self-deception is *not* a failure of self-knowledge, judging at least from the examples provided in "Deception and Division", falls out of the picture. For Davidson, self-deception is to deceive oneself about a fact. Carlos deceives himself about *the fact* that he is not likely to pass the test, and the man who is going bald (Davidson) deceives himself about this fact. These are examples of misleading oneself about a fact that threatens one's view of oneself and one's self-esteem, but not examples of misleading oneself about one's own mental states.

Moran is concerned to shed light on the misconceptions that are inherent in what he calls the "Cartesian picture" of self-knowledge. He argues that although

[385] Ibid. p. 2. For a deeper discussion of the problems that arise in certain modern, philosophical accounts when the language and thought-forms appropriate to third-person observations about states of affairs is applied to first-person expressions, see Sharon Rider's dissertation *Avoiding the Subject*: *A critical Inquiry into Contemporary Theories of Subjectivity* (Stockholm: Thales, 1998).

it aims as accounting for first-person authority, this picture is misleading because it construes the first-person perspective as no different from the spectator's stance, except insofar as I am taken to have a *privileged position* with respect to my own mental life. In this picture, introspective awareness is thought of as an eye that gazes inwards. Self-knowledge is portrayed as analogous to knowledge of objects in being observational and evidence-based, but, because of the privileged position one is assumed to have with respect to one's own mental states, introspection is thought to be infallible. This picture of self-knowledge, however, has been challenged, as Moran points out. Freud's theory of the unconscious, for example, presupposes that we are *not* aware of all of our mental states.[386]

I agree with Moran that Freud's theory of the unconscious shows that one can be ignorant of, or mistaken about, one's own mental states. As regards the distinctiveness of self-knowledge, I have directed my attention mainly to passages in Freud's works where, in my view, he makes important contributions towards dismantling the picture of self-knowledge as analogous to knowledge of objects and knowledge of other people's feelings, beliefs, etc. Nevertheless, there are remains of "the Cartesian picture" in Freud's thinking. As we saw in the last chapter, Freud even uses the image of self-awareness as an inner eye in attempting to formulate the difference between conscious and unconscious material. By necessity, the "inner eye" is directed at some particular mental content at any given point in time; more importantly, some mental content is *prevented* from catching the glance of this "eye of attention" since unconscious mental content is "in a different room" than the eye, and thus beyond its reach.

In Chapter Three, I argued that this picture is confused and gives rise to problems. Here I will return to Freud's case study of the woman who repeats the wedding night scene to argue that Freud is led astray by the picture referred to above in accounting for her failure of self-knowledge. Freud claims that there is an unconscious intention in the woman's act, namely, to correct her husband's mishap and rehabilitate him in the eyes of the maid. I have argued that what is misleading in Freud's claim that he has discovered the intention of the act is that he describes the woman's motivation as an intention that was there all along, seemingly *in just the same form or articulation* as it is when it has become conscious, as if the only difference is that it was not observed by her before. In this picture, failure of self-knowledge is like failure of perceptual knowledge. The idea is that one fails to see what is, in a literal sense, there, and not that one fails to makes sense of it. To speak of unconscious intentions is not necessarily problematic in itself, but what is problematic here, I hold, is that Freud minimizes the differences between unconscious and conscious intention. Freud's inclination to sometimes think of introspective awareness as observation

[386] Ibid, p. 3 and p. 13. "My interpretation carries with it the hypothesis that intentions can find expression in a speaker of which he himself knows nothing but which I am able to infer from circumstantial evidence." (Freud, *Introductory Lectures*, SE, vol. 15, pp. 64).

of objects (as an inner eye) partly explains why he sometimes fails to distinguish between conscious and unconscious motivation.

The failure to appreciate the differences between knowledge of my own feelings, on the one hand, and knowledge of things and other people's feelings, on the other, is present also in Gardner's account of self-deception. Like Freud, in his analysis of the neurotic woman, Gardner argues that Anna Karenina had an intention all along, but that she was not aware of it while in a state of self-deception. Gardner says that the evidence that she deceived herself in order to make time for her to cultivate her relationship with Vronsky was there all along but was postponed.[387] Gardner's account clearly assumes that to become conscious is to discover something that was there all along. The intention is "the object" that was there all along, but which Anna failed to see, in Gardner's version of the "Cartesian picture".

In the articulation of self-knowledge exemplified by Freud's analysis of the case of the neurotic woman and by Gardner's analysis of Anna Karenina's self-deception, self-knowledge is taken to be like knowledge of external objects, where becoming aware of something, e.g. one's motivation or one's feelings regarding someone, makes no difference to this "something". The object remains the same. The fine distinctions and nuances that Freud captures rather well, in my view, by describing how something can go from being an unconscious idea to a belief which one holds, are not captured in his account of the neurotic woman's act. This analysis does not bring out how the subject's awareness of her motivation makes a difference to the quality of the motivation.

Moran claims, "self-consciousness has specific *consequences* for the object of consciousness."[388] Further, when we speak of the "consciousness" of a conscious belief, we do not specify a theoretical relation that the subject has to this belief. Rather, "a conscious belief enters into different relations with the rest of one's mental economy and thereby alters its character";[389] consciousness thereby "informs and qualifies the belief in question".[390] Moran quotes Charles Taylor:

> Formulating how we feel, or coming to adopt a new formulation, can frequently change how we feel. When I come to see that my feeling of guilt was false, or my feeling of love self-deluded, the emotions themselves are different [...]. We could say that for these emotions, *our understanding of them or the interpretation we accept are constitutive of the emotion*. The understanding helps shaping the emotion. And that is why *the latter cannot be considered a fully independent object* […][391]

---

[387] "What we can however do instead is to postpone the evidence: at some later time Anna Karenina will think, or be able to think, 'so that's why I told myself… And there is of course an explanation for why the intention does not show itself in the present tense: to do so would gain nothing for it, and risk its extinction." (Gardner, p. 22)
[388] Moran, p. 28.
[389] Ibid. pp. 30.
[390] Ibid. p. 30.
[391] Charles Taylor, "The Concept of a Person" in *Human Agency and Language: Philosophical Papers*. Vol. 1 (Cambridge University Press, 1990). My italics. Quoted in Moran, p. 39.

Anna Karenina's insight that she loves Vronsky is a good example. In allowing herself to understand her feelings as love, her feelings are altered. Being in love with Vronsky is now a part of her self-conception. We should also understand the motivation for action in this way; i.e. the neurotic woman's motivation is not a fully independent object, but her understanding helps *shape* it. In Moran's words: "one's state of mind is in some sense conceptually dependent on how one interprets it."[392] This is why it is misleading to describe a motivation of which one is unconscious or unaware as a *conscious* intention but which the subject has not observed, i.e. as an independent object. Awareness or consciousness of a belief (idea) or attitude alters its character. And, in so doing, it alters the mental life of the individual as a whole by integrating it into the rest, which, in the case of important insights, might mean a radical psychological and existential transformation.

But, one might ask, does not the view that in becoming conscious of a feeling (for instance) one alters the character of that feeling, suggest that one cannot be self-deceived? Is it right to say that Anna Karenina was in love before she became aware of her feelings *as* love? If not, would it not be wrong to say that she deceived herself? Moran asks if it is possible to be mistaken about one's own mental states and replies in the affirmative: "Even though introspective awareness does not base itself on observation of behavior, and even after we have weaned ourselves from the picture of observation directed to an interior, there remains the sense that one's reflection is answerable to the facts about oneself, that one is open to the normal epistemic risk of error, blindness and confusion".[393] Thus, although first-person knowledge is privileged, it does not mean that it is always authoritative: one can be mistaken about one's own mental states. Given the possibility of being mistaken, Moran asks, "[d]oesn't this require the idea of a 'fully independent object'?"[394], that is, a feeling, intention etc., which remains the same whether or not its "owner" is conscious of it? Here Moran's answer is negative: we can admit to the possibility of being mistaken about one's own mental states without assuming that the feeling, intention etc. remains the same. He accounts for this in terms of first- and second-order beliefs: "If a person is at all rational, his first-order beliefs will indeed be sensitive to his second-order beliefs about them, and they will change accordingly. He may, for instance, discover that some set of his beliefs is inconsistent, or suspect that a particular belief of his is the product of prejudice or carelessness […]. His first-order beliefs will then normally change in response to his interpretation of them."[395]

Let me apply this to the example of Anna Karenina. Anna holds the first-order belief that her feelings for Vronsky are innocent (she might think of him

[392] Moran, p. 40.
[393] Ibid.
[394] Ibid.
[395] Ibid. p. 55.

as her soul-mate, and regard the flirtation between them as secondary and irrelevant). But when Anna is overwhelmed with emotion and Vronsky foists his confession of love upon her, she realizes that her understanding of their relation as innocent is the product of wishful thinking that she has maintained despite all the things that should have made her realize the truth (second-order belief). Her first-order belief changes with this insight; she realizes that she is in love with Vronsky. What makes this a case of self-deception is not that "love" was present as a constant, independent object in Anna's mind all along, which she kept herself from seeing. Rather, what makes this self-deception is that it has been a fact for some time that Anna bears feelings for Vronsky which are not innocent, that she has made rearrangements in her life so that she can see him more, etc., and she has kept herself from understanding this as love.

I have said that when one becomes conscious of a mental attitude, i.e. a feeling, it changes the character of that feeling, and it also re-shapes one's conscious mental life as a whole. It is in this sense that it can be said that becoming conscious of something also makes a difference to one's moral life. Moran writes:

> The special features of first-person awareness cannot be understood by thinking of it purely in terms of epistemic access (whether quasi-perceptual or not) to a special realm to which only one person has entry. Rather, we must think of it in terms of special responsibilities the person has in virtue of the mental life in question being *his own*. In much the same way that his action cannot be for him just part of the passing show, so his beliefs and other attitudes must be seen by him as expressive of his various and evolving relations to his environment, and not as a mere succession of representations (to which, for some reason, he is the only witness). And in both the cases of actions and attitudes, self-consciousness makes a difference to what the person's responsibilities and capacities are, with respect to his involvement in their development. It is modeling self-consciousness on the theoretical awareness of objects that obscures the specifically first-person character of the phenomenon, whether or not this theoretical perspective takes the specific form of the perceptual model of introspection.[396]

When we think of first-person awareness of one's mental states as taking possession of one's desires, feelings, beliefs, etc., as part of one's conscious mental life, and so part of that which one *can* question, decide about, take responsibility for, and so forth, we see that, in a moral person, this also gives rise to the *requirement* that one should act responsibly. We can thus see why one is unconsciously motivated not to become aware. Self-deception is both a flight from anxiety and a flight from responsibility.

---

[396] Ibid. p. 32. Sebastian Gardner offers a critical notice of Moran's book, which discusses, for example, Moran's view that our relation to our psychological states is practical in the sense that it *does* something to our states. (Sebastian Gardner, "Critical Notice of Moran's *Authority and Estrangement: An Essay on Self-Knowledge* in *Philosophical Review*", vol. 113, no. 2, April, 2004, p. 259-262.)

Short Summary

My aim in this investigation has been to break with a conception of self-deception that I find implausible as a general picture, and which gives rise to great problems and drastic solutions.[397] I have been concerned with identifying and discussing problems that arise in the rationalist and intentionalist accounts of Davidson and Gardner. In carrying out this task I have presented another way of viewing self-deception, another context in which to understand it, than the one that Davidson and Gardner propose. I have said: "I want you to look at self-deception in this way". I do not only intend for the picture of self-deception that I have in part found in, in part constructed from the works of Freud and others, to be seen as an illuminating analogy which can reveal some problems with Davidson's and Gardner's account of self-deception while, perhaps, introducing a range of other problems. I believe that many of the problems that arise in Davidson and Gardner accounts arise because they remove self-deception from the contexts in which we are acquainted with it and in which it makes sense, and assign to it a place in an artificial context, their respective theories. It has been my ambition to present a picture that is more true to the phenomenon of self-deception and to our experience of it (which amounts to the same thing). Having said that, I recognize that in presenting my view I do, perhaps, in places, slip into making claims about self-deception that are stronger or more general than should be. The reader must be the judge of that.

In rounding up this investigation, I will not recapitulate all the central arguments. Rather, I wish to make clear the point and purpose of the point and project as a whole. I have argued against the standard approach to self-deception in which self-deception is seen as analogous to deceiving someone else, using Davidson's account as my prime example. I have argued that his construal of the problem of self-deception, as well as his proposed solution, rests on problematic assumptions: for example, that it involves holding two contradictory beliefs, that the self-deceiver knows what he has best evidence to believe, that self-deception is an intentional action strategically directed at inducing the opposite belief in oneself from that which one takes to be true so as to avoid pain, and, more specifically, that the avoidance of pain is the result of judging that, all things considered, it is better to avoid pain. I have argued against these assumptions. I hold the conception of self-deception as analogous to deceiving someone else to be problematic as such, and I find many of its central problems to be displayed in Davidson's account.

I turned to Gardner's account, a main task of which is to separate "ordinary irrationality", including self-deception, from forms of irrationality treated by psychoanalysis. I argued that the distinction that Gardner makes is problematic,

[397] Davidson's account of self-deception starts with a paradox, a seemingly unsolvable problem, and Davidson's account of self-deception consists in solving this puzzle.

and that the problem lies in his conception of self-deception as rational, even hyper-rational, and in his view of human beings as essentially "rational unities". Gardner takes himself to have found support in Freud for his view that self-deception involves preference and can be fitted into his view of persons as essentially rational. I have argued that this support is lacking. In Gardner's account as well as in Davidson's, self-deception is portrayed as an intentional action directed at obtaining something further. Anna Karenina deceives herself about her love for Vronsky in order to make time for their love to grow. I argue against the conception of self-deception as directed at obtaining something further or fulfilling a wish.

In the third chapter, I presented a context for understanding self-deception that is very different from Davidson's and Gardner's, by looking at what the term connotes, and turning to Freud's discussions of illusion, (psychiatric) delusion, defensive reactions, the Ego as an organization, etc., all of which I take to be exploring phenomena that are central to a discussion of self-deception. The chapter was an exploration of these themes in Freud's writings. I argued that self-deception is a motivated failure of self-knowledge in which motivations affect already what we apprehend. Thus, self-deception is not typically to mislead oneself about what one knows, but to prevent oneself from coming to know or becoming aware. In response to Davidson's conception of self-deception as a lie to oneself, I argued that self-deception is better seen as a "flight from anxiety", where anxiety arises because one's conception of oneself as a whole is threatened, especially when the moral demands that one places on oneself are in conflict with the desires. I also paid attention to passages in Freud's investigations that both are problematic in themselves and pose difficulties for the Freudian conception of self-deception. I contrasted these with other passages in which Freud gives a more careful and better account of these issues.

It is my view that Freud's texts open up for an insightful and fruitful approach to self-deception. Davidson and Gardner construe self-deception as primarily a theoretical problem for which they propose theoretical solutions. At his best, Freud shows us both how and why it is natural that we should at times deceive ourselves, but also why this inclination is a very real and very deep *human* problem.

Bibliography

Anscombe, Elisabeth, *Intention* (Oxford: Basil Blackwell, 1976).

Augustine, *Confessions*, transl. John K. Ryan (Image Book, 1960).

Bortolotti, Lisa, "Intentionality without Rationality", published in the *Proceedings of the Aristotelian Society*, vol. 105, 2005, p. 369-376 (Blackwell Publishing).

Bouveresse, Jacques, *Wittgenstein Reads Freud: The Myth of the Unconscious* (Princeton: Princeton University Press, 1995).

Cavell, Marcia, "Introduction" to Donald Davidson, *Problems of Rationality* (Oxford: Clarendon Press, 2004).

—— *The Psychoanalytic Mind: From Freud to Philosophy* (Harvard: Harvard University Press, 1996).

Demostenes, *Olynthiacs, Phillippics, Minor Public Speeches*, trans. J. H. Vince, Loeb Classical Library (Harvard: Harvard University Press, 1930).

Davidson, Donald, "Deception and Division" in *Problems of Rationality* (Oxford: Clarendon Press, 2004).

—— "Expressing Evaluations" in *Problems of Rationality* (Oxford: Clarendon Press, 2004).

—— "Incoherence and Irrationality" in *Problems of Rationality* (Oxford: Clarendon Press, 2004).

—— "Introduction" to *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984).

—— "Mental Events" in *Actions and Events* (New York: Oxford University Press, 1980).

—— "On Saying That" in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984).

—— "Paradoxes of Irrationality" in *Problems of Rationality* (Oxford: Clarendon Press, 2004).

—— "Three Varieties of Knowledge" in *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press, 2001).

—— "Truth and Meaning" in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984).

—— "Who is Fooled?" in *Problems of Rationality* (Oxford: Clarendon Press, 2004).

Dupuy, Jean-Pierre (ed.), *Self-Deception and Paradoxes of Rationality* (Stanford: CSLI Publications, 1998).

Evnine, Simon, *Donald Davidson* (Stanford: Stanford University Press, 1991).

Fingarette, Herbert, "Self-Deception Needs no Explaining" in *Self-Deception* (Berkeley: University of California Press, Ltd., 2000).

Fink, Bruce, *Lacan to the Letter* (Minneapolis: University of Minnesota Press, 2004).

Freud, Sigmund, *Gesammelte Werke*, ed. Anna Freud (Frankfurt am Main: Fisher Verlag, 1952), *Gesamtregister* (1972), Band 18.

—— *Studien über Hysterie*, Band 1, („Zur Psychotherapie der Hysterie").

—— *Gesammelte Werke*, ed. Anna Freud (London: Imago Publishing, 1947-48), „Bemerkungen über einen Fall von Zwangsneurose", Band 7.
—— *Das Ich und das Es*, Band 13.
—— *Jenseits des Lustprinzips*, Band 13.
—— „Fetischismus", Band 14.
—— *Hemmung, Symptom und Angst*, Band 14.
—— *Zur Psychopatologie des Alltagslebens*, Band 4.
—— „Der Realitätsverlust bei Neurose und Psychose", Band 13.
—— *Das Unbehagen in der Kultur*, Band 14.
—— „Das Unbewusste", Band 10.
—— „Die Verdrängung", Band 10.
—— „Über die Psychogenese eines Falles von weiblicher Homosexualität", Band 10.
—— *Vorlesungen zur Einführung in die Psychoanalyse*, Band 11.
—— *Die Zukunft einer Illusion*, Band 14.
—— *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, ed. and transl. James Strachey (London: The Hogarth Press and the Institute of Psycho-Analysis, 1978), *Beyond the Pleasure Principle*, vol. 18.
—— *Civilization and its Discontents*, vol. 21.
—— *The Ego and the Id*, vol. 19.
—— "Fetishism", vol. 21.
—— *The Future of an Illusion*, vol. 21.
—— *Inhibitions, Symptoms and Anxiety*, vol. 20.
—— *Introductory Lectures on Psycho-Analysis, part I* (1915-16), vol. 15.
—— *Introductury Lectures on Psycho-Analysis, part II* (1916-17), vol. 16.
—— "The Loss of Reality in Neurosis and Psychosis", vol. 19.
—— "Notes upon a Case of Obsessional Neurosis" (The Rat Man), vol. 10.
—— "The Psychogenesis of a Case of Homosexuality in a Woman", vol. 18.
—— *Psychopathology of Everyday Life*, vol. 6.
—— "Repression", vol. 14.
—— Editor's Note on "Repression", vol. 14.
—— *Studies on Hysteria*, vol. 2, ("The Psychotherapy of Hysteria").
—— "The 'Uncanny'", vol. 17.
—— "The Unconscious", vol. 14.
—— Editor's note on "The Unconscious", vol. 14.
Gardner, Sebastian, "Critical Notice of Moran's *Authority and Estrangement: An Essay of Self-Knowledge*", in *Philosophical Review*, vol. 113, no. 2 (April, 2004), pp. 249-267.
—— *Irrationality and the Philosophy of Psychoanalysis* (Cambridge: Cambridge University Press, 1993).
—— "The unconscious", in *The Cambridge Companion to Freud*, ed. Jerome Neu (Cambridge: Cambridge University Press, 1991), pp. 136-160.
Guttman, Samuel A., *The Concordance to the Standard Edition of the Complete Psychological Works of Sigmund Freud*, 2:d ed. (London: International Univ. Press, 1984).
Hållén, Elinor, "Självbedrägeriet i den analytiska filosofin" in *Tankar tillägnade Sören Stenlund*, Niklas Forsberg, Sharon Rider, and Pär Segerdahl (eds.), Uppsala Philosophical Studies 54, p. 233-248 (Västerås: Edita Västra Aros: 2008).
Ibsen, Henrik, *The Wildduck*, transl. Frances E. Archer, 2nd edition, (Massachusetts : Digireads.com, 2008).
Johnston, Mark, "Self-Deception and the Nature of Mind", in *Perspectives on Self-Deception*, ed. Brian. P. McLaughlin and Amélie Oksenberg Rorty (Berkeley and Los Angeles: University of California Press, 1988).

Kant, Immanuel, *Critique of Pure Reason*, transl. Norman Kemp Smith, (London: Macmillian Press Ltd.).
Koffka, Kurt, "On the Structure of the Unconscious" in *The Unconscious: A Symposium* (New York: Alfred A. Knopf, 1928).
Lacan, Jacques, *Écrits: The First Complete Edition in English*, transl. Bruce Fink (New York: W. W. Norton & Company, 2006).
Landweer, Hilge, „Selbsttäuschung", in *Deutsche Zeitschrift für Philosophie* (2001), 49/2, s. 209-227.
Landy, Joshua, *Philosophy as Fiction: Self, Deception, and Knowledge in Proust* (Oxford: Oxford University Press, 2004).
Lazar, Ariela, "Deceiving Oneself Or Self-Deceived? On the Formation of Beliefs 'Under the Influence'" in *Mind*, Vol. 108. April 1999 (Oxford University Press, 1999).
—— "Division and Deception: Davidson on Being Self-Deceived" in *Self-Deception and Paradoxes of Rationality*, ed. Jean-Pierre Dupuy (Stanford: CSLI Publications, 1998).
Lear, Jonathan, *Freud: An Introduction* (New York: Routledge, 2006).
—— *Open Minded*: *Working out the Logic of the Soul.* (Harvard University Press, 1999).
McLaughlin, Brian P. and Oksenberg Rorty, Amélie (eds.), *Perspectives on Self-Deception*, (Berkeley and Los Angeles: University of California Press, 1988).
Mele, Alfred, *Irrationality: An essay on Akrasia, Self-Deception, and Self-Control* (New York: Oxford University Press, 1987).
Alain de Mijolla (ed.), *International Dictionary of Psychoanalysis* (Macmillan Library Reference, 2004).
Moran, Richard, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001).
Murdoch, Iris, *The Sacred and Profane Love Machine* (Penguin Books Ltd, 1976).
Nietzsche, Friedrich, *Beyond Good and Evil: Prelude to a Philosophy of the Future* (London: Penguin Classics, 1990).
—— *Human, All Too Human: A Book for Free Spirits*, transl. R. J. Hollingdale (Cambridge: Cambridge University Press, 2002).
Plato, *Cratylus*, G. P. Goold (ed.), transl. H.N. Fowler Loeb Classical Library (London: Harvard University Press, 1977).
—— *Phaedrus*, in *Plato in Twelve Volumes*, vol. 9, transl. Harold N. Fowler (London: William Heinemann Ltd., 1925).
Rider, Sharon, *Avoiding the Subject: A Critical Inquiry into Contemporary Theories of Subjectivity* (Stockholm: Thales, 1998).
Ritter, J., Gründer, K. (eds.) *Historisches Wörterbuch der Philosophie* (Darmstadt: Wissenschaftliche Buchgesellschaft). Band. 9 (Se/Sp), 1995.
—— Band. 10 (St-T), 1998.
Russell, Bertrand, *The Conquest of Happiness* (London: The Unwin Brothers Ltd., 1930).
Sartre, Jean-Paul, *Being and Nothingness*, transl. Hazel E. Barnes (London and New York, 2005).
Sjöbäck, Hans, *Psykoanalysen som livslögnsteori: Läran om försvaret* (Lund: Bo Cavefors Bokförlag, 1977).
Stenlund, Sören, *Det Osägbara* (Stockholm: Norstedts förlag, 1980).
—— "Ethics, Philosophy and Language" in *Commonality and Particularity in Ethics*, Lilli Allanen, Sara Heinämaa och Thomas Wallgren (eds.) (Ipswich: Macmillian Press Ltd, 1997).

Taylor, Charles, "The Concept of a Person" in *Human Agency and Language: Philosophical Papers*, vol. 1 (Cambridge University Press, 1990).
Tolstoy, Leo, *Anna Karenina* (London: Penguin Books Ltd, 2006).
—— *Anna Karenina* (Harmondsworth, 1978).
Wittgenstein, Ludwig, *The Blue and Brown Books*, (Oxford: Blackwell, 1958).
—— *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief* (Oxford: Basil Blackwell, 2007).
—— *Philosophical Investigations* (Cornwall: Blackwell Publishing, 2001).
Wollheim, Richard, *Sigmund Freud* (Cambridge: Cambridge University Press, 1995).

Appendix

A Comparison with Fetishism

Although there are a number of doubtful assumptions at play in Freud's theory of fetishism, such as that fetishism arises as a way of protecting the subject from the threat of castration, I find Freud's portrayal of the fetishist's evasive behavior, as well as Freud's attempts at describing to what extent the fetishist is aware of that of which he tries to keep himself oblivious, illuminative of what is going on in self-deception. Disregarding the truth of Freud's account of fetishism, I will briefly present Freud's analysis and use it as an analogy, in hope that it can be an illustrative picture of some aspects of self-deception.

According to Freud's theory, fetishist fantasies and ceremonies are typically borne in childhood when the little boy sees a woman's genitals for the first time and becomes frightened when he sees that she lacks a penis. Having been threatened with castration as a form of punishment, seeing the woman's genitals makes the boy apprehend the threat as a real danger: the woman appears to have been castrated. In effect, the sense perception of her genitals is rejected

from his memory. What remains from the exciting experience of peering under the woman's skirt is something that he saw just before or at the same time as he had the frightening sight of her genitalia, for example, the shoes or garter. This object becomes a substitute for what is the natural object for the boy's curiosity, and later the natural object for his sexual desire. Freud writes:

> when the fetish is instituted some process occurs which reminds one of the stopping of memory in traumatic amnesia. As, in this latter case, the subject's interest comes to halt half-way, as it were; it is as though the last impression before the uncanny and traumatic one is retained as a fetish. Thus the foot or shoe owes its preference as a fetish – or part of it – to the circumstance that the inquisitive boy peered at the woman's genitals from below, from her legs up […] pieces of underclothing, which are so often chosen as a fetish, chrystallize the moment of undressing, the last moment in which the woman could still be regarded a phallic.[398]

Isolation is key to the avoidance of anxiety in the case of fetishism, and, moreover, to how the boy, later the man, can both have his sexual desires fulfilled *and* avoid recalling what was traumatic in the experience.

Let us now compare this analysis of fetishism to the central questions of the thesis regarding what it is that the self-deceiver deceives himself about. Is it a belief? Is it conscious? Or, is self-deception rather a prevention of coming to awareness of something; does it disrupt the formation of a belief? Freud says of the effect that seeing the woman's genitals had on the little boy's beliefs and behavior: the boy doesn't recognize that females have no penis – a fact which is extremely undesirable to him since it is a proof of the possibility of his being castrated himself. Instead he disavows his own sense perception, which showed him that female genitals lack a penis, and holds fast to the contrary conviction. The disavowed perception does not, however, remain entirely without influence: he removes his interest from the genitalia and takes hold of something else – a part of the body or some other object – and assigns it the role of the penis. It is usually something that he saw at the moment at which he saw the female genitals.[399]

Freud says that the boy's sense perception showed him that female genitalia lack a penis, but that he still does not let go of his conviction that they do have a penis. How is this possible? He continues, it is not the case that the sense perception is *erased*. "'Scotomization'[400] seems to me particularly unsuitable, for it suggests that the perception is entirely wiped out, so that the result is the same as when a visual impression falls on the blind spot in the retina."[401] The sense perception continues to have an effect on the subject: "In the situation we

---

[398] Freud, "Fetishism" (1927), SE, vol. 21, p. 155.
[399] Ibid. pp. 153.
[400] Scotomization is a psychoanalytic term for the mind's ability to delete of forget a trauma or overwhelming event.
[401] Freud, "Fetishism", p. 154.

are considering, on the contrary, we see that the perception has persisted, and that a very energetic action has been undertaken to maintain the disavowal. It is not true that, after the child has made his observation of the woman, he has preserved unaltered his belief that women have a phallus. He has retained that belief, but he has also given it up."[402] The boy's belief that the woman has a penis (motivated by his avoidance of anxiety) still remains, in spite of the perception that should undermine it. But maintaining the belief gives rise to the fetish as well as other "symptoms", such as aversion to the female genitals. The fetish remains, as Freud says, a *stigma indelebile*[403] of the repression that has taken place. This means that the belief no longer is what it was before.[404] The direction of attention away from the female genitals to the fetish object and the intense desire for the fetish that keeps the attention there, is the effect of the anxiety provoked by seeing the female genitals as indicating the danger of castration. The desire and attention are turned to the substitute, the fetish, and it is kept there by the (substitutive) intense desire for the fetish.

Freud writes that the sense perception shows the boy that the female genitals lack a penis, yet he says that the boy "doesn't recognize that females have no penis." Thus there is a gap between what the boy sees and what he acknowledges: although he has a sense perception which will affect his direction of attention as well as his behavior and reflection deeply, he doesn't acknowledge this sense perception and what it entails. Many cases of self-deception can be analogously understood, although it is not typically sense perceptions that one deceives oneself about. The case of Carlos serves well as an example. Carlos hears his driving instructor's negative comments, yet he does not fully grasp the sense: that his instructor's words is a judgment which is to be taken seriously that he (probably) will not pass the test. The instructor's comment nevertheless continues to affect Carlos, perhaps so much so that he avoids reflecting upon what the instructor said as well an exposure to situations in which he could receive another negative comment. The example of Anna Karenina shows that she was confronted with the wonderful yet anxiety-provoking "feeling of animation" at her very first meeting with Vronsky, later she realizes that ever since then she has been deceiving herself. This can be compared with the boy's exciting yet frightening experience of seeing the female genitals. But, as the little boy's sense perception did not change his view that women have a penis, sensing the "feeling of animation" and suppressing it did not mean that Anna's view of herself changed: she did not acknowledge the strength of her feelings for Vronsky, nor did she consider how they might

---

[402] Ibid.

[403] Ung. "a mark not easily erased".

[404] "Yes, in his mind the woman *has* got a penis, in spite of everything; but this penis is no longer the same as it was before. Something else has taken its place, has been appointed its substitute, as it were, and now inherits the interest which was formerly directed to its predecessor." (Freud, "Fetishism", p. 154.)

change her life. Being over-whelmed by desire for Vronsky continued to affect her behavior without altering her beliefs, values and views.

Anna's dismissive response to Vronsky's proclamations of love is in some ways similar to the little boy's aversion to the sight of the female genitals. If Anna accepts that what Vronsky and herself share is love, a whole world of worries, anxiety, shame etc. opens up to her. Her self-deception has been quite successful in keeping this "knowledge", this insight, out of awareness. To play on the analogy with fetishism, one could say that so far she has amused herself with the shoes or garters when enjoying Vronsky's company. She has received great enjoyment from their talks, from dancing together and, from receiving his full attention whenever they have met and, although her desire cannot be completely fulfilled by these few moments with Vronsky, their relation does not cause her any worries as long as she thinks of it in a way which does not stand in conflict with her life as wife and mother. She has kept herself from appreciating the seriousness in their feelings for each other and thereby been protected from the strong anxiety which appreciating that she loves him would evoke, and from the choice that insight would seem to make inevitable. Ignorance is bliss.

According to Freud, by instantiating the fetish and acting out fetish behavior, the sense perception is kept buried and the castration anxiety is kept on a tolerable level. The development of the fetish is motivated, but the motivation is unacknowledged. This is also the case with self-deception: self-deception is motivated, since it protects the subject from acknowledging painful or hurtful truths. In the fetishist-case, the disavowal of the sense perception is not a choice. In fact, it sets in before a choice is possible, since a choice presupposes awareness what one chooses between; it presupposes judgment. In the fetishist case, the sense perception does not reach articulation as a belief, but remains disavowed. Thus there can be no question of choice. Likewise, the self-deceiver remains unaware of that of which she deceives herself and, while motivation is critical for her remaining in ignorance, her ignorance is not something that she chooses.[405]

[405] I am grateful to Tuomas Nevanlinna for having pointed out that Freud's discussion of fetishism could be helpful in my attempts to come to grips with the way in which one might "know" that which one is deceiving oneself of.