The Apologetic Implications of
Self-Deception

Greg L. Bahnsen

Table of Contents

Chapter One: Introduction

1.1 The Familiarity of the Notion

Describing or portraying persons as self-deceived is far from uncommon in the history of human literature, even though it has of late become uncommonly difficult to explain how self-deception is possible. "The easiest thing of all," said Demosthenes, "is to deceive one's self; for what a man wishes he generally believes to be true" (*Olynthiaca* iii.19). We find a classic portrayal of a man-in the grip of self-deception in the Greek tragedy by Sophocles, *Oedipus Rex*. Plato is the first philosopher to mention the phenomenon, when he represents Socrates as saying, "For there is nothing worse than self-deception--when the deceiver is always at home and always with you" (*Cratylus* 428d). Elsewhere Plato exposits the same as "the true lie" which, in contrast to a mere lie "in words," is a lie "in the soul" or a matter of lying to oneself (*Republic* 382a).

Self-deception likewise plays a part in the traditional literature of the Bible. The Old Testament clearly portrays King David as a man who knew the moral standards of God according to which adultery and murder are condemned; yet he is depicted as terribly dull to the character of his relation with Bathsheba and his behavior toward her husband, Uriah, until later convicted by the parable of Nathan the prophet (II Samuel 11-12; cf. Psalm 51). The New Testament explicitly mentions self-deception in the context of false profession of faith: "If anyone seems to be religious, and bridles not his tongue but deceives his own heart, this one's religion is vain" (James 1:26); "If we say that we have no sin, we deceive ourselves and the truth is not in us" (I John 1:8). Ancient literature thus manifests a familiarity with the notion of self-deception and supplies us with recognized illustrations of its working.

The writers, philosophers, sociologists, and psychologists of the modern age equally manifest that the notion of self-deception is a common

one. Sir Philip Sidney, the outstanding man of letters in Queen Elizabeth's Court, once remarked, "It many times falls out that we deem ourselves much deceived in others because we first deceived ourselves." His biographer, Fulke Greville, added his own aphorism, "No man was ever so much deceived by another as by himself." And Shakespeare borrowed a plot from Sidney's *Arcadia* in creating his superb *The Tragedy of King Lear*, in which we find the masterful picture of Lear's self-deceptive denial of the death of his daughter Cordelia. Another seventeenth-century writer making direct reference to the phenomenon was Matthew Prior, as in this couplet from his poem *Solomon*: "Hoping at least she may herself deceive/against experience willing to believe" (Bk. iii, 1. 223).

However, the most extensive discussion of "self-deception" to be found prior to the present day comes from the pen of a Puritan preacher; in 1617 the fifth edition of Daniel Dyke's *The Mystery of Selfe-Deceiving* appeared in London and ran over four hundred pages. The "mystery" mentioned in his title is not identical with the paradox which is central to the current discussion in philosophical circles, for Dyke's interest was primarily religious and ethical--as was also Bishop Butler's, a century later, when he published a collection of his famous sermons. In the "Preface" to these sermons Butler wrote in a way which anticipated the tendency of some analysts today to liken intrapersonal deception to interpersonal deception; Butler also recognized that, despite the obscurity which might attend its philosophical analysis, self-deception was a personal reality:

> The Sermon upon the character of Balaam, and that upon Self-deceit, both relate to one subject. I am persuaded that a very great part of the wickedness of the world is, one way or other, owing to the self-partiality, self-flattery, and self-deceit, endeavored there to be laid open and explained. It is to be observed amongst persons of the lowest rank, in proportion to their compass of thought, as much as amongst men of education and improvement. It seems, that people are capable of being thus artful with themselves, in proportion as they are capable of being so with others. Those who have taken notice that there is really such a thing, namely, plain falseness and insincerity in men with regard to themselves, will readily see the drift and design of these Discourses: and nothing that I can add will explain the design of them to him, who has not beforehand remarked, at least, somewhat of the character. And yet the admonitions they contain may be as much wanted by such

a person, as by others; for it is to be noted, that a man may be entirely possessed by this Unfairness of mind, without having the least speculative notion what the thing is.[1]

About a decade later a similar moralistic interest in the subject was expressed by Benjamin Franklin in his *Poor Richard's Almanac* (for 1738): "Who hath deceived thee so often as thyself?" Such a point about human nature came to assume the virtual status of a popular, cynical platitude, judging from its pithy recurrence in a series of subsequent writers. Göethe wrote, "We are never deceived; we deceive ourselves" (*Sprüche in Prosa*, iii)--which is quite a contrast to one opinion current today that, unlike being deceived by others, self-deception is not literally possible at all. Agreeing with Göethe's remark, Rousseau said, "Nature never deceives us; it is always we who deceive ourselves" (*Emile on Education*, III). Without being so exclusivistic, Schopenhauer nevertheless laid the stress on intrapersonal deception when he said, "We deceive and flatter no one by such delicate artifices as we do our own selves" (*The World as Will and Idea*, book i). The obvious reality of something matching up to the term 'self-deception' is attested by Edward Bulwer-Lytton: "The easiest person to deceive is one's self" (*The Disowned*, ch. 42), as well as by the saying attributed to the Unitarian clergyman, William R. Alger: "Every man is his own greatest dupe." Similar sentiments often appeared in Nietzsche's discussions of human nature; he said that, despite their protestations, it is not clear that men really do want the truth because, after all, it is often something with which we find it hard to live (see, e.g., *Beyond Good and Evil*, sections 25, 35, 264). Consequently, Nietzsche could call philosophers "you strange actors and self-deceivers," and he could suggest with respect to Kant's alleged discovery of a faculty in man for synthetic *a priori* judgments that "he deceived himself in this matter" (sections 9, 11).

A long list of talented novelists have provided insightful descriptions of the self-deception into which human beings can treacherously fall. Charles Dickens gives expression to the startling character of self-deception in this soliloquy from *Great Expectations*:

All other swindlers upon earth are nothing to the self-swindlers, and with such pretences did I cheat myself. Surely a curious thing. That I should innocently take a bad half-crown of somebody else's manufacture is reasonable enough; but that I should knowingly

---

[1] Joseph Butler, *Sermons* (New York: Robert Carter & Bros., 1870 [1729]), p. xv.

reckon the spurious coin of my own make as good money! An obliging stranger, under pretence of compactly folding up my bank-notes for security's sake, abstracts the notes and gives me nutshells; but what is his sleight of hand to mine, when I fold up my own nutshells and pass them on myself as notes![2]

Flaubert rendered a pitiless and detailed expose of the self-delusion arising from romantic fantasy in his realist novel, *Madame Bovary*, where we read of the gradual downfall of Emma--from her early sentimental daydreaming, through self-deceptive intrigues with lovers, to the final degrading affair initiated so as to raise money: for her husband's creditors. The same psychological realism characterizes Proust's portrayal of the reaction of Marcel to the disappearance of Albertine in *Remembrance of Things Past* and Henry James' description of the way in which Strether remains oblivious for so long to the relationship that has developed between Chad and the Contesse de Vionnet in *The Ambassadors*. James writes, "He almost blushed, in the dark, for the way he had dressed the possibility in vagueness. . . . He recognized at last that he had really been trying all along to suppose nothing."[3] This reference to an effort at thinking nothing of the available evidence is noteworthy, and the analysis of such a thing will eventually come to occupy a critical position in our discussion. Tolstoy was another author who was keenly aware of the human penchant to hide one's head from what one does not want to see. Writing of Count Rostov's return home from a business trip, only to discover that something had happened to his daughter, Tolstoy says this in *War and Peace*:

The count saw clearly that something had gone wrong during his absence; but it was so terrible for him to imagine anything discreditable occurring in connection with his beloved daughter, and he so prized his own cheerful tranquility, that he avoided asking questions and did his best to persuade himself that there was nothing very much wrong or out of the way….[4]

---

[2] Charles Dickens, *Great Expectations* (New York: New American Library, 1963 [1861]), p. 245.

[3] Henry James, The Ambassadors, vol. 2 (New York: Harper and Brothers, 1902), p. 238

[4] Leo Tolstoy, *War and Peace*, vol. 1 (New York: Penguin Books, 1972), p. 698.

The author speaks with ease of Rostov clearly seeing something, yet doing his best to persuade himself that it was not the case. Similarly Dostoevsky's underground man says that if a man does not want to "see" something, no matter how obvious a belief it would be, he can always "purposefully go mad in order to be rid of reason and have his own way."[5] Pictures of people caught in self-deception are not difficult to come by in modern literature; indeed self-deception amounts to a predominant theme in novels and plays of the last century. We see it at work in Kitty Scherbatsky's unhappy attempt to emulate the altruistic lifestyle of Varenka in Tolstoy's *Anna Karenina*. Tolstoy also portrays it well in his shorter pieces, like *Father Sergius* and *The Death of Ivan Illych*. Self-deception is made manifest in an arresting fashion when, at the end of O'Neill's *The Iceman Cometh*, Hickey stops himself short upon realizing that he has expressed to his friends the rationalization he used and has exposed the fact that he actually killed his own wife. It is illustrated as well in Andre Gide's *Pastoral Symphony*, where a middle-aged and married pastor falls in love with Gertrude, the blind girl he has cared for and educated, but steadfastly and self-deceptively hides that damning fact from himself, even as he frantically works to hinder his grown son's interest in the young lady. The subtlety and details of self-deception are graphically discussed in such literary pieces, but the depiction reaches something of a climax of excellence in *The Fall* by Camus, where Jean-Baptiste Clamence's confession of former egoism and self-deception is itself portrayed as a continuing manifestation of his very self-deception.

The interest in and approach to self-deception in the previously mentioned philosophers, theologians, and writers was of an ethical-religious nature; seeing it in these terms continues to be characteristic of existentialist philosophers, like Kierkegaard and Sartre. For Kierkegaard the double-mindedness of self-deception is the ultimate personal sin which keeps men from attaining purity of heart. It is a universal condition:

> If it were true--as conceited shrewdness, proud of not being deceived, thinks--that one should believe nothing which he cannot see by means of his physical eyes, then first and foremost one ought to give up believing in love. If one did this and did it out of fear of being deceived, would not one then be deceived? Indeed, one can be deceived in many ways; one can be deceived in believing

---

[5] Fyodor Dostoevsky, *Notes from Underground*, trans. R. E. Neatlow (New York: Dutton, 1960), p. 28.

what is untrue, but on the other hand, one is also deceived in not believing what is true; one can be deceived by appearances, but one can also be deceived by the superficiality of shrewdness, by the flattering conceit which is absolutely certain that it cannot be deceived….Which sight is more sorrowful, that which immediately and unrestrainedly moves to tears, like the sight of one unhappily deceived in love, or that which in a certain sense could tempt laughter, the sight of one who is self-deceived, whose foolish conceit of not being deceived is ludicrous.[6]

Kirkegaard taught that men deceive themselves by ignoring what is in their hearts, thus leading inevitably to a clash between one's private and public selves; in such a state men are unable to will without conflict and frustration. He writes of this at length in his *Purity of Heart is to Will One Thing*, for instance:

There is an ignorance about one's own life that is equally tragic for the learned and for the simple, for both are bound by the same responsibility. This ignorance is called self-deceit. There is an ignorance that by degrees, as more and more is learned, gradually changes into knowledge. But there is only one thing that can remove the other ignorance which is self-deception. And to be ignorant of the fact that there is one and-only one thing, and that only one thing is necessary, is still to be in self-deception….The ignorant man can gradually acquire wisdom and knowledge, but- the self-deluded one if he won 'the one thing needful' would have won purity of heart.[7]

For Sartre self-deception is that "bad faith" by which men attempt to escape personal responsibility for what they are and do. Sartre sees an inherent duplicity in human consciousness because human existence stands between being and nothingness--between facticity and the freedom to transcend factual existence (creating the kind of person one shall be). Consciousness is always of something and as such is "being for-itself," and what one is conscious of always involves choice on one's part. Yet as self-conscious beings, human

---

[6] Soren Kierkegaard, *Works of Love*, trans. Howard and Edna Hong (New York: Harper Torch Books, 1962), p. 23.

[7] Soren Kierkegaard, *Purity of Heart Is to Will One Thing* (New York: Harper and Row, 1956), p. 52.

beings have an ambiguous reality, for they also are a given something of which they can be aware (thus "being in-itself"). In bad faith one denies this inner ambiguity and thus the role of choice in making him what he is to be. Finding the anxiety unbearable which results from the realization that one is free to be as he chooses, one seeks security (immunity from responsibility) by thinking of himself as in some way determined. On the other hand one can attempt to avoid responsibility for choices he has already made by denying his facticity and identifying himself purely with consciousness and choice. Either way, bad faith or self-deception arises in the attempt to escape the incoherence of nothingness (freedom) and being (facticity) in man himself:

> Good faith seeks to flee the inner disintegration of my being in the direction of the in-itself which it should be and is not. Bad faith seeks to flee the in-itself by means of the inner disintegration of my being. But it denies that it is itself in bad faith. Bad faith seeks by means of "not-being-what-one-is" to escape-from the in-itself which I am not in the mode of being what one is not. It denies itself as bad faith and aims at the in-itself which I am not in the mode of "not-being-what-one-is-not." If bad faith is possible, it is because it is an immediate, permanent threat to every project of the human being; it is because consciousness conceals in its being a permanent risk of bad faith. The origin of this risk is the fact that the nature of human consciousness simultaneously is to be what it is not and not to be what it is.[8]

In bad faith one lies to himself, and since all acts--whether premeditated or spontaneous--are intentional for Sartre, and since a person's commitments (his "fundamental project") are always chosen even when he is not "reflectively conscious" of having them, therefore in bad faith one intentionally chooses to lie to himself about himself. He both knows and is ignorant of the same thing at the same time. Sartre says that this impasse cannot be rejected, even though it cannot be comprehended.

The ethical-religious approach to self-deception which has been the focus of our survey up to this point must now, if we are to be fully aware of how common the notion has been in human thought and writing, be supplemented with the sociological and psychological approaches to self-

---

[8] Jean-Paul Sartre, *Being and Nothingness*, trans. with Introduction by Hazel E. Barnes (New York: Washington Square Press, 1966 (1956)), p. 116.

7

deception that have profoundly affected Western culture in the last century or so.[9] The theories of "unhappy consciousness" in Hegel's *Phenomenology of Spirit*, of distortion through "ideology" and "false consciousness" in Marx and Mannheim, and of "repression" and "the unconscious" in Freud have fostered conceptions of the mind which deepen our awareness of the human capacity for self-deception. Approaching the phenomenon of self-deception in a broad sociological framework, Marx spoke of false consciousness in the sense that a man's thoughts do not truly reflect that they have unconscious, material determinants and do not let on that they unwittingly express--under the guise of unbiased thinking--one's own economic interests. To use the words of Engels,

> Ideology is a process accomplished by the so-called thinker consciously, it is true, but with a false consciousness. The real motive force impelling him remains unknown to him; otherwise it simply would not be an ideological process. Hence he imagines false or seeming motive forces. Because it is a process of thought, he derives its form as well as its content from pure thought, either his own or his predecessors. He works with mere thought material, which he accepts without examination as the product of thought, and does not investigate further for a more remote source independent of thought.[10]

More importantly, however, false consciousness applies to that ideological thought which misrepresents the present economic situation, unconsciously serving to perpetuate a waning economic condition which is in the interest of the one propounding the ideology. The thinking of a whole class of people may hereby reflect collective illusion. The rise of new, realistic beliefs about the economic situation will always bring class struggle and finally the overthrow of worn-out ideologies which are self-deceptive.[11] Whereas Marx applied the idea of false consciousness in a limited way, Mannheim came to view all consciousness as inherently deceptive. By nature man's mind is self-

---

[9] Cf. Daniel Shapiro, "Self-Deception," Ph.D. dissertation, City University of New York (Ann Arbor: Xerox University Microfilms, 76-1491, 1975), pp. 118-130. An extensive discussion of moral self-deception, especially in reference to "the meaning of one's life," can be pursued in Ilham Daman and D. Z. Phillips, *Sense and Delusion* (New York: Humanities Press, 1971).

[10] Friedrich Engels, "Letter to Franz Mehring," *Marx and Engels Basic Writings on Politics and Philosophy*, ed. Lewis S. Feur (Garden City: Doubleday Anchor, 1959), p. 408.

[11] Marx elaborates these views especially in *The German Ideology*, "A Contribution to the Critique of Political Economy," and "Manifesto of the Communist Party," ibid.

deceptive:

> The particular conception of ideology is implied when the term denotes that we are skeptical of the ideas and representations advanced by our opponents. They are regarded as more or less conscious disguises of the real nature of a situation, the true recognition of which would not be in accord with his interests. These distortions range-all the way from conscious lies to half-conscious unwitting disguises; from calculated attempts to dupe others to self-deception.
>
> The particular conception of ideology merges with the total. . . . Herewith a new and perhaps most decisive stage in the history of modes of thought has been reached. . . . The total conception of ideology raises a problem which has frequently been adumbrated before, but which now for the first time acquires broader significance, namely the problem of how such a thing as "false consciousness". . . could ever have arisen. It is the awareness that our total outlook as distinguished from its details may be distorted, which lends to the total conception of ideology a special significance and relevance for the understanding of our social life.[12]

For Mannheim all beliefs are the combined result of the "cooperative process of group life" as well as our personal interests and strivings. When a belief continues to be held even though the underlying external, social-determinant of it-has changed, then it has become pragmatically deceptive (i.e., acting on it will prove ineffective). In. self-deception we continue to hold such deceptive beliefs because they play a central role in organizing our thoughts and forming our general perspective on ourselves and others. Men have a general reluctance to examine thoroughly their theoretical formulations lest it have a disquieting effect on their positions; they are too intimately identified with their beliefs to open their eyes to the unrealistic aspects of them. In a sense Mannheim's sociology of knowledge is an epistemological analogue to Freud's theory of defense, which in turn is the functional equivalent of Freud's own account of self-deception.

Turning from the sociological to the psychological approach to the phenomenon of self-deception, we naturally look at the work of the father of psychoanalysis, Sigmund Freud. The self and its deceits receive a full

---

[12] Karl Mannheim, *Ideology and Utopia* (New York: Harcourt, Brace, and World, 1946), pp. 55, 70-71

treatment in Freud's works, although he apparently never used the term 'self-deception.' Freud very well knew the reality denoted by the term, however, as is clear from his autobiographical story of unsuccessfully searching for a train connection which, if found, would help him arrive in time at his brother's, but prevent him from fulfilling the desire to see a display of Rembrandt's paintings; after a frantic search, Freud missed the train, only later to remember (after seeing the art display) that the sign and appropriate train had been clearly in sight at the outset of his search at the train station.[13] Freud knew that he was not alone in such self-deceiving types of experience:

> It is astonishing (and this generally meets with too little acceptance) how-easily and frequently intelligent people give reactions of partial feeble-mindedness under psychological constellations; anyone who is not too conceited may observe this in himself as often as he wishes, and especially when some of the thought-processes concerned are connected with unconscious or repressed motives. . . It is rather an everyday occurrence, even in normal people, that they are deceived about the motives of their actions and do not become conscious of them until afterward, when a conflict of several emotional occurrents establishes for them the causality of such confusion.[14]

Psychoanalysis can be viewed as an attempt at a systematic study of self-deception and its motivations. It sets forth a complex and elusive view of a divided self which is irrational at base and always opaque to itself; the rational processes of man are secondary to his primitive processes shared with the animal world. Freud maintained that much of what we think, feel, and do can only be explained by unconscious forces within us--especially the unconscious attempt to protect ourselves from inner impulses which are deemed unacceptable in society. Thus men were said to need to deceive themselves about certain unacceptable psychic realities, which they accomplished through defense mechanisms like repression, sublimation, isolation, inhibition, and rationalization. Such willful yet unconscious maneuvers were portrayed by Freud as natural and necessary since he thought men needed their illusions in order to live securely and happily. This outlook allowed for the use of skill

---

[13] Sigmund Freud, *The Psychopathology of Everyday Life*, trans. A. Tyson, ed. J. Strachey (New York: Norton, 1965 [1901]), pp. 227-228.

[14] Sigmund Freud, *Delusion and Dream*, trans. H. Zohn, ed. P. Rieff (Boston: Beacon Press, 1968 f19071), pp. 94, 88.

and cleverness in one's hiding from himself what he found unacceptable in himself; he could both recognize it and successfully fail to see it. One's beliefs, especially about himself, would not need to be integrated into a coherent whole, given Freud's new perspective, and it would be tolerable that a man was unable to see his own self-deceptive behavior and devices. In broad strokes, such was Freud's now famous attempt to analyze psychologically a wide range of self-deceptive behavior in man.

Our preceding survey of ancient and modern literature, found in a number of localities and throughout a variety of fields, has demonstrated by way of illustration that describing and portraying persons as self-deceived is far from uncommon. The history of human literature gives ample evidence that men identify something in their experience as self-deception. The notion is quite a familiar one. Accordingly we are inclined to think that the notion of self-deception must make sense. After all, not only in traditional literature, but also in common life the concept is well known. The vocabulary of self-deception is recognizable, mastered by people, and even taught to others. In addition to professional scholars in various fields, even men with little advanced education can, and do, speak readily of self-deception. When the son of Mr. Jones has been caught red-handed in stealing lunch money out of the desks of fellow students at school, and Mr. Jones continues to protest his son's innocence ("the school officials have a vendetta against little Johnny; they are framing him…"), nobody finds it awkward to say that Mr. Jones is "deceiving himself." Prior to reflecting seriously on just what self-deception could be and how it could be possible, we show little if any inclination to dismiss the notion as muddled, incoherent, or senseless. The literature utilizing it is both vast and diverse. Self-deception is part of our common experience and conversation. Familiarity breeds acceptance.

1.2 The Apparent Paradox Elaborated

In observing the familiarity of the notion of self-deception we have referred to popular writers, novelists, theologians, philosophers and other scholars whose overlapping interests in and approaches to self-deception have been roughly categorized as ethical-religious, sociological, or psychological. However, notwithstanding the value and insights these approaches can have, a fourth approach to the subject will be our present concern. This might generally be characterized as the analytic-epistemological approach to the subject of self-deception; it takes a philosophical interest in certain conceptual questions pertaining to self-deception which arise in the theory of knowledge

and the philosophy of mind. The recent addition of such a philosophical interest in self-deception dates from the 1960 article by Raphael Demos, "Lying to Oneself."[15] Nevertheless, it is adumbrated in Bertrand Russell's (likely misapplied) criticisms of Freud in *The Analysis of Mind* and in Gilbert Ryle's challenge to the traditional mind-body dualism in *The Concept of Mind.* Both Russell and Ryle attempted to analyze mental phenomena in behavioral terms. Russell distinguished between primary desires, which were bodily and natural for man, and secondary desires, which were in some sense caused by beliefs a person held. Ordinarily the satisfaction of a secondary desire would not completely remove discomfort for a person, unless all of his primary desires were also satisfied. However, there are some secondary desires which are not only caused by a belief, but also completely satisfied by a belief. Genuine self-deception pertains to such grand desires (e.g., vanity, religion, optimism) that can be satisfied simply by believing that they are satisfied. In short, self-deception is simply a matter of desire-motivated belief or wishful thinking; in self-deception one satisfies a desire through holding a belief.

> What may, with some propriety, be called self-deception arises through the operation of desires for beliefs. We desire many things which it is not in our power to achieve: that we should be universally popular and admired, that our work should be the wonder of the age, and that the universe should be so ordered as to bring ultimate happiness to all, though not to our enemies until they have repented and been purified by suffering. Such desires--are too large to be achieved through our own efforts. But it is found that a considerable portion of the satisfaction which these things would bring us if they were realized is to be achieved by the much easier operation of believing that they will be realized. This desire for beliefs, as opposed to desire for the actual facts, is a particular case of secondary desire, and, like all secondary desire its satisfaction does not lead to a complete cessation of the initial discomfort. Nevertheless, desire for beliefs, as opposed to desire for facts, is exceedingly potent both individually and socially. According to the form of belief desired, it is called vanity, optimism, or religion.[16]

---

[15] Raphael Demos, "Lying to Oneself," *The Journal of Philosophy* 57, no. 18 (September 1, 1960):588-594.

[16] Bertrand Russell, *The Analysis of Mind* (London: George Allen and Unwin, 1968), pp. 74-75.

Ryle viewed the phenomenon of self-deception as an acute problem 1 for the adherents of the traditional mind-body dualism ("the dogma of the Ghost in the machine"), according to which a person has direct knowledge of the workings of his mind--an introspective ability that is free from illusion and doubt. Such a concept is difficult to hold when we see people gulled by their own hypocrisies:

> People are actuated by impulses the existence of which they vigorously disavow; some of their thoughts differ from the thoughts which they acknowledge; and some of the actions which they think they will to perform they do not really will. They are thoroughly gulled by some of their own hypocrisies and they successfully ignore facts about their mental lives which on the official theory ought to be-patent to them.[17]

Yet, ironically, the attempt to understand self-deception from the standpoint of Ryle's own teaching proves equally problematic.[18] Ryle denied that reflection on one's current conscious occurrences is possible, and he limited the data which were available to one's "retrospection." Accordingly one's discovery of his motives for acting in a particular way would not be immune from bias.

> The way in which a person discovers his own long-term motives is the same as the way in which he discovers those of others. The quantity and quality of the information accessible to him differ in the two inquiries, but its items are in general of the same sort. He has, it is true, a fund of recollections of his own past deeds, thoughts, -fancies and feelings; and he can perform the experiments of fancying himself confronted by tasks and opportunities which have not actually occurred. He can thus base his appreciations of his lasting inclinations on data which he lacks for his appreciations of the inclinations of others. On the other side, his appreciations of his own inclinations are unlikely to be unbiased and he is not in a favourable position to compare his own actions and reactions with those of others. . . .[19]

---

[17] Gilbert Ryle, *The Concept of Mind* (New York: Barnes and Noble, 1969 [1949]), p. 14.

[18] Janice Auritt Oser, "Invitations to Self-Deception: An Application of Jean-Paul Sartre's Analysis of Self-Deception to Views of Self-Knowledge, Self-Identity, and Self-Deception Advanced, Respectively, by Gilbert Ryle, Terence Penelhum, and Herbert Fingarette." Ph.D. dissertation, New York University (Ann Arbor, MI: Xerox University Microfilms, 76-19,531, 1976), pp. 5-51.

[19] Ryle, *Concept of Mind*, p. 90.

We are to understand in a case of personal self-deception that one <u>will</u> not recognize his motives, but rather systematically avoids them:

> . . . At least part of what is meant by describing jealousy, phobias or erotic impulses as 'unconscious' is that the victim of them not only does not recognise their strength, or even existence, in himself, but in a certain way <u>will</u> not recognise them. He shirks a part of the task of appreciating what-sort of a'person he is, or else-he systematically biases his appreciations.[20]

By means of this systematic effort, a person can deceive himself about his motives, just as he can deceive another person about them. In such a case, though, self-deception would be an intelligent strategy, which for Ryle implies that the agent of the systematic avoidance would be quite able to say, without research or conjecture, what he is engaged in and is trying to accomplish.[21] Having learned to detect the insincerities of others, a person could readily apply the techniques of detection to his own present insincerity.[22] At this point Ryle's account begins to baffle us, intimating that there is some bifurcation within the person after all. For how can the self-deceiver simultaneously avoid and detect his insincerity? How can a person act purposely (systematic avoidance) and not know or be able to acknowledge his purpose (as biased and deceived)? How does the self-deceiver hide his motives as well as his *hiding* of them?

Questions such as these reflect, and provide a suitable introduction to, the kind of analytic-epistemological discussion of self-deception which Demos inaugurated in 1960 with his article, "Lying to Oneself." Normally people have not seriously reflected upon and critically questioned either the common literary references to self-deception or their own personal mentions of it. Apart from Sartre, even those writers who have given some special attention to the phenomenon of self-deception in an ethical-religious, sociological, or psychological context have not inquired as to just what this kind of deceiving must involve and whether someone could actually accomplish this feat. Yet it has turned out that pioneering philosophical analyses of self-deception over the last two decades have frequently ended in, or been stumped by, some form of paradox. Just as Aristotle was puzzled over <u>akrasia</u>, contemporary analytic-epistemological examination of self-deception has left many a philosopher puzzled.

---

[20] Ibid., p. 157.
[21] Ibid., pp. 74, 147.
[22] Ibid., pp. 193-195.

There are cases where a man's mistaken belief is of his own making as such; to uncover instances of this self-deception we need not look far. Yet the idea that a man could perpetrate a deception upon himself should strike us as peculiar. How could S hold a belief which is not only erroneous (or poorly supported) but which also appears styled to help him evade facts of which he can hardly be unaware? Demos expressed the quandary in this fashion:

> Self-deception exists, I will say, when a person lies to himself, that is to say, persuades himself to believe what he <u>knows</u> is not so. In short, self-deception entails that B believes both p and not-p at the same time. Thus self-deception involves an inner conflict. . . . Believing and disbelieving are pro and con attitudes; they are contraries and therefore it is logically impossible for them to exist at the same time in the same person in the same respect.[23]

As any number of philosophers have observed, "It is easy to understand how a man can deceive another and difficult to understand how a man can deceive himself."[24] The difficulty in self-deception is that S must play two roles, deceiver and deceived, although in ordinary cases of interpersonal deception the two roles are incompatible. "This makes self-deception sound about as difficult as presiding at one's own funeral."[25] Being modeled after other-deception, deceiving oneself would require a duality which is precisely precluded by the term '<u>self</u>-deception.'

> Only what changes everything is the fact that in bad faith it is from myself that I am hiding the truth. Thus the duality of the deceiver and the deceived does not exist here. Bad faith on the contrary implies in essence the unity of a <u>single</u> consciousness…. It follows first that the one to whom the lie is told and the one who lies are one and the same person, which means that I must know in my capacity as deceiver the truth which is hidden from me in my capacity as the one deceived. Better yet I must know the truth very exactly <u>in order</u> to conceal it more carefully--and this not at two

---

[23] Demos, "Lying to Oneself," pp. 588, 591.

[24] Frederick A. Siegler, "Self-Deception," *Australasian Journal of Philosophy* 41, no. 1 (May 1963):29.

[25] T. S. Champlin, "Self-Deception: A Reflexive Dilemma," *Philosophy* 52, no. 201 (July 1977):28l.

different moments, which at a pinch would allow us to re-establish a semblance of duality--but in the unitary structure off-a single project. How then can the lie subsist if the duality which conditions it is suppressed?… We must agree in fact that if I deliberately and cynically attempt to lie to myself, I fail completely in this undertaking; the lie falls back and collapses beneath my look; it is ruined *from behind* by the very consciousness of lying to myself which pitilessly constitutes itself well within my project as its very condition.[26]

If we understand the deception involved in interpersonal deception and intrapersonal deception to be the same, then apparently a paradox arises:

In a typical instance, when Jones deceives Smith about some assertion P, it is true that:

(i) Jones is aware that P is false,
(ii) Jones intends to make Smith believe that P is true,
(iii) Jones succeeds in making Smith believe that P is true. When Jones deceives himself about P, substituting 'Jones' for 'Smith' in the above, we get:

(A) Jones, aware that P is false, intends to make himself believe that P is true, and succeeds in making himself believe that P is true.

The puzzle here is that, since people sometimes deceive themselves, and since (i), (ii), and (iii) mirror standard features of the use of 'deceive', we seem committed to asserting sentences of the form of (A); yet (A) is an extremely odd-sounding, if not contradictory, statement. How, for example, can a person come to believe (at time t) what he is aware of (at t) as false?[27]

Can S intentionally persuade himself to believe what he recognizes as false—to believe what he disbelieves?

The kind of problem sometimes said to arise here can be brought out by comparing what is supposed to happen when somebody is self-deceived with standard cases of deception proper, or "other-

---

[26] Sartre, *Being and Nothingness*, p. 89.
[27] John V. Canfield and Don F. Gustavson, "Self-Deception," *Analysis* 23 (December 1962):32.

deception.!' If a man sets out to deceive another, he tries to make him believe something which he himself knows, or believes, not to be so; such an attempt commonly takes the form of lying, although it need no—there are other methods, such as behaving (or arranging things) in certain ways that may be, depending upon the circumstances, appropriate. In so far as he is successful, the other will have been caused to accept something incompatible with what he, the deceiver, holds to be the case. It has been claimed, however, that when this is used as a model whereby to explicate-the notion of self-deception, unwelcome paradoxes ensue. For it is surely odd to suggest that somebody could try to make, and succeed in making, himself believe something which he, *ex hypothesi*, at the same time believes not to be true. In so far as lying, e.g., is a deliberate attempt to misinform, or conceal the truth from, the person lied to, it is essential that the liar should know and accept what it is that he is trying to hide from his victim; it is also a presupposition of this type of undertaking that the intended victim should not be aware of the deceiver's aims. But the transposition of these conditions to cases where the roles of deceiver and deceived are allegedly occupied by one and the same individual might lead one to conclude that self-deception is a contradictory or incoherent enterprise, incapable of fulfillment.[28]

"The problem remains as to whether there could be a coherent account of how one can hold a belief which one is aware is unsatisfactory."[29] When we say that S has convinced himself that p is true, although he believes (or ought to believe) that p is false, it then seems that S has not succeeded in deceiving himself after all. Already believing that p is false, how could S get himself to think simultaneously that p is true? The very mark of a successful effort at self-deception (viz., perpetrating a false belief upon oneself) seems to presuppose the very circumstance which would constitute failure in such an effort (viz., one's believing that the belief is false to begin with).[30]

Mention was made previously of the hypothetical case-of Mr. Jones, who continues to protest his son's innocence even though the boy was caught

---

[28] Patrick Gardiner, "Error, Faith and Self-Deception," *Proceedings of the Aristotelian Society* 70, n.s. (1969-1970):224-225.

[29] David Pugmire, "'Strong' Self-Deception," *Inquiry* 12, no. 3 (Autumn 1969):343.

[30] James Marvin Shea, "Self-Deception," Ph.D. dissertation, Cornell University (Ann Arbor, MI: Xerox University Microfilms, 67-1411, 1966), pp. 1-2.

red-handed in stealing money out of fellow students' desks at school. At the time there seemed nothing awkward about saying that Mr. Jones was deceiving himself. Things appear otherwise now. Is there really such a thing as perpetrating a deception on oneself? We want to ask how it could be possible. This question may be asking two different things: how can such a notion avoid incoherence and thus be logically possible?, or what forms (mechanisms, practical procedures) might the application of the concept take?[31] In the current philosophical literature on self-deception these two facets of the question (coherence and capacity) are sometimes distinguished and sometimes combined. For instance, John Turk Saunders begins his article, "That we engage in self-deception is a datum. How we manage to pull it off is what needs explanation."[32] However, Gardiner asserts: "The question round which [the difficulties of self-deception] revolve is not, How-does self-deception, as a familiar psychological phenomenon, occur? Rather it is the question, How, given that a certain view of what constitutes self-deception is correct, could it occur?"[33] The precise "question of self-deception" to be answered is somewhat broader when Stanley Paluch asks, "Can there be a substitution-instance for "I know p but believe not-p' which would (a) be logically coherent and (b) be 'compossible' with a charge of self-deception?"[34]

Returning to the example of Mr. Jones, we want to ask how and if he could really be deceiving himself regarding his son's innocence. As deceiver, Mr. Jones believes that Johnny is guilty but wants to convince himself otherwise; as deceived, Mr. Jones comes to believe (falsely) that Johnny is innocent and persecuted. As self-deceiver, then, Mr. Jones must simultaneously believe that John is guilty and innocent. He must be knowledgeable and ignorant of the facts at the same time. But that clearly appears paradoxical to us on the face of it, presenting "the paradox of a person persuading himself to believe

---

[31] D. W. Hamlyn, "Self-Deception," *Proceedings of the Aristotelian Society, Supplemental Volume* 35 (1971):57.

[32] John Turk Saunders, "The Paradox of Self-Deception," *Philosophy and Phenomenological Research* 35, no. 4 (June 1975):561.

[33] Gardiner, "Error, Faith, and Self-Deception," p.224.

[34] Stanley Paluch, "Self-Deception," *Inquiry* 10, no. 3 (Autumn 1967):269. "Things are possible when they are not self-contradictory; two or more things are compossible when they belong to one and the same possible world, i.e., when they may coexist. All possible worlds have general laws, analogous to the laws of motion; . . . Hence two or more things which cannot be brought under one and the same set of general laws are not compossible" (Bertrand Russell, *A Critical Exposition of the Philosophy of Leibniz*, 2nd ed., [London: George Allen and Unwin, 1937], p. 66).

what he knows isn't so."[35] Two kinds of difficulty are actually packed into this problem, one pertaining to the voluntariness, and the other to the rationality, of belief.

> The view that we can induce ourselves into holding a belief appears to be an impossible undertaking. The idea that we can persuade ourselves to believe what we know isn't so is doubly odd. Even if we could induce ourselves into holding a belief, how can we do this when we already know better?[36]

We observe a third facet of the problem of self-deception in the paradox created by the element of mendacity (or conscious misrepresentation) which people often take to be an unavoidable concomitant of purposeful deception. It is because of the "purposeful mis-appreciation of some matter"[37] that self-deception puzzles us:

> It is self-contradictory to say that M purposefully gets himself to believe something that he <u>all along</u> knows is incompatible with something he believes . . . . The paradox arises because of the purposefulness of self-deception. . . . Self-deception is not a matter of mere stupidity or carelessness in thinking. It is a craftily engineered project, and this is why it seems pointless and self-contradictory.[38]

In this regard two kinds of self-deception are dealt with in the current literature, as indicated by the way in which Fingarette specifies the focus of his book: "When I speak of self-deception, I shall not mean to include the 'innocent kind,' i.e., the cases where the belief is not induced purposely and with a knowledge that it is false."[39] Paluch distinguishes "weak models" of self-deception which take the sting out of the paradox (that a man believes the opposite of what he knows) by holding that the agent is in some sense unconscious of, or does not occurrently know, what he is up to.[40] Pugmire

---

[35] Herbert Fingarette, *Self-Deception* (New York: Humanities Press), p. 18.

[36] Shapiro, "Self-Deception," p. 6.

[37] James Michael Russell, "Self-Deception," Ph.D. dissertation, University of California at Santa Barbara (Ann Arbor, MI: Xerox University Microfilms, 73-8105, 1971), p. 1.

[38] Saunders, "The Paradox of Self-Deception," p. 561.

[39] Fingarette, *Self-Deception*, p. 20.

[40] Paluch, "Self-Deception," pp. 271, 272.

defends the possibility of a person intentionally bringing himself to believe contrary to what he is aware is true (thereby paralleling the mendacity of other-deception) and speaks of this as "strong self-deception."[41] The problem involved in self-deception which is not innocent and weak, that is self-deception which is purposeful, can be variously expressed, depending on the response one gives to the question whether intentional activity entails that the agent is conscious of what he is attempting. It is either the problem of explicating "unconscious purposefulness" or the problem of how S can be successful in consciously deceiving himself. In the case of unconscious motivation,

J. M. Russell says,

> . . . there arises the puzzling notion of making a mentalistic ascription to a person when you don't suppose that the matter in question occurs to the person, when you think he would disavow that ascription, and where you think it inappropriate to call him a liar.[42]

In the case of conscious deception of oneself, the question becomes:

> How is self-deception possible, since for a man to deceive himself it would seem he would have to think something such that, if he did think it, he would not be deceived? . . . Are there mental occurrences which must take place in persons who seek to deceive others, which would make a similar—i.e., purposeful—deception of (and about) oneself impossible? For 'common sense' seems to urge that a person could not set out to deceive himself without giving thought to what he was doing in a way which would make deception of himself impossible.[43]

In examining the syndrome of self-deception it will prove-helpful and 1 necessary to bear in mind the various facets of, or distinctions to be drawn with respect to, the general problem as it is taken up for 1 analysis by different philosophers.

Returning again to Mr. Jones, we are now quite aware of how awkward it

---

[41] Pugmire, "'Strong' Self-Deception," passim.
[42] J. M. Russell, "Self-Deception," pp. 6-7.
[43] Ibid., pp. 1-2.

would be to say that he is "deceiving himself" about his son's innocence. The very suggestion swarms with questions. Could self-deception be possible? Is the notion coherent? How could it be carried out? Can Jones simply believe what he wants? Can he believe what he disbelieves or knows to be false? Can he do this on purpose? Could he do it with an unconscious intention, thus ending up with knowledge and ignorance of the same thing? Could he consciously deceive himself, thus rendering the roles of deceiver and deceived compatible in practice? The addition of the analytic-epistemological approach to the literature on self-deception over the past two decades makes us hesitant to speak of it confidently and clearly. Does it make sense to say that Jones is deceiving himself? What account can we offer for the notion of self-deception? The various analyses given of it are problematic and might very well discourage us from referring to self-deception at all. At least that might be our inclination after confronting the maze of conflicting philosophical treatments of self-deception currently available. Our common understanding seems to lead us into confusion, and we are no longer sure that men can genuinely lie to themselves--until we reflect again on our all-too-familiar experience.

> Were a portrait of man to be drawn, one in which there would be highlighted whatever is most human, be it noble or ignoble, we should surely place well in the foreground man's enormous capacity for self-deception. The task of representing this most intimate, secret gesture would not be much easier were we to turn to what the philosophers have said. Philosophical attempts to elucidate the concept of self-deception have ended in paradox--or in loss from sight of the elusive phenomenon itself. . .   We are beset by confusion when once we grant that the person in question is in self-deception. For as deceiver one is insincere, guilty; whereas as genuinely deceived, one is the innocent victim. What, then, shall we make of the self-deceiver, the one who is both the doer and the sufferer? Our fundamental categories are placed squarely at odds with one another. . . . 'The one who lies with sincerity,' who convinces himself of what he even then knows is not so, who lies to himself and to others and believes his own lie though in his heart he knows it is a lie--the phenomenon is so familiar, the task so easy, that we nod our heads and say, 'of course.' Yet when we examine what we have said with respect to its inner coherency, we are tempted to dismiss such a description as nonsense.[44]

---

[44] Fingarette, *Self-Deception*, p. 1, 5.

21

1.3 Requirements of an Adequate Solution

As familiar as we may be with self-deception, the attempted analysis of it has still proven problematic and paradoxical in various ways. Some answer to the problems and some solution to the paradox must be found. Since we so commonly refer to self-deception and find its occurrence to be frequent, any analysis of it which suggests that it is impossible is itself to be deemed suspect. Many people are surer that self-deception occurs than they would be of any explanation which concludes that self-deception is only apparent and not genuine. We resist such a conclusion because people do not merely play at self-deception; they engage in it in tragic ways, and very often they will come to realize this fact later.[45]

> We must remember that when we speak of self-deception we refer to a concept of which very many people have made use through many centuries. Now it is not inconceivable that whenever these people have, through the centuries, made use of this concept they have been contradicting themselves without noticing it. On the other hand, though conceivable, it is not likely. Consequently when in philosophy we are confronted by an account of self-deception which makes' it appear contradictory, our first assumption should be that the confusion lies not in the notion of self-deception but rather in the philosophical account.[46]

 "Since it is a datum that we engage in self-deception, the notion of self-deception cannot be more than apparently self-contradictory."[47] Announcing this assumption means that our work is cut out for us. We will deny the occurrence of self-deception and call the notion incoherent only if unavoidably driven to such a position. It is the analysis of self-deception which has generated recent perplexity. Therefore, although it has proven elusive, the rendering of an adequate account of self-deception is the task to which we are here committed.

In the most general sense, an adequate analysis of self-deception would attain the ideal originally set down by Demos: "Such an analysis 'saves' the phenomena while at the same time conforming to the requirements of the

---

[45] Stanley Hauerwas and David Burrell, "Self-Deception and Autobiography: Theological and Ethical Reflections on Speer's *Inside the Third Reich*." *Journal of Religious Ethics* 2, no. 1 (Spring 1974): 99-117.

[46] H. O. Mounce, "Self-Deception," *Proceedings of the Aristotelian Society, Supplemental Volume* 45 (1971):6l.

[47] Saunders, "The Paradox of Self-Deception," p. 559.

law of contradiction."[48] It would explain self-deception in a way which rids of incoherence: "One is puzzled not because one is unfamiliar with the phenomenon of self-deception. One's puzzlement arises from the apparent contradictions in the grammar of 'self-deception'."[49] An adequate analysis would also be descriptively correct:

> Any philosophical treatment of self-deception must meet the phenomenon head on. Certain facts about a number of human beings make it plausible to describe what they are engaged in as self-deception. The philosopher's task is, given the plausibility of that description, to examine the question, how, if at all, that description can be a possible one.[50]

Being true to the phenomenon would include not departing radically from ordinary language with respect to it and not confusing it with similar phenomena in human experience. In elucidating the notion of self-deception we should see what it is about the notion that invites a paradoxical construction and then aim to remove the perplexity,[51] which certainly discourages us from appealing to concepts which are even less clear than self-deception might be. When self-deception is put in a true light, it will be seen as a complex yet non-contradictory notion. However, in our attempt to clear up the paradoxical air around the notion, we must not dismiss the disturbing element of self-deception, so that after analysis it is unclear why anyone would have been perplexed in the first place. The disturbing quality of the phenomenon must remain and be displayed, even though the analysis of it is non-paradoxical.[52]

With these general guidelines in mind we can formulate the standards-of an adequate solution to the apparent paradox of self-deception. These criteria explicitly and implicitly agree with those that are commonly put to service in the contemporary philosophical literature on self-deception.[53] If

---

[48] Demos, "Lying to Oneself," p. 594.

[49] Béla Szabados, "Self-Deception," *Canadian Journal of Philosophy* no. 1 (September 1974):52.

[50] Hamlyn, "Self-Deception," p. 45.

[51] Charles Dawson Bruce, "An Investigation of Self-Deception," Ph.D. dissertation, Michigan State University (Ann Arbor, MI: Xerox University Microfilms, 75-20,815, 1975), p. 42.

[52] Shea, "Self-Deception," pp. 172-173.

[53] Clear examples of the application of these various criteria can be found especially in the different series of articles authored by Frederick Siegler and Béla Szabados which are listed in the Bibliography. Most of the criteria are explicitly announced, e.g., in. Frederick A. Siegler, "An Analysis of Self-Deception," *Nous* 2, no. 2 (May 1968): 147-164; Szabados, "Self-Deception."

an analysis of self-deception is acceptable and adequate:

> (1) It must supply the truth conditions for 'S deceived himself into believing that p.'
> (2) It must be true to the ordinarily recognized, paradigm examples of self-deception (in both traditional literature and current philosophical treatments) and be able to account for the ordinary language of 'self-deception.'
> (3) It must avoid logical contradiction and paradox.
> (4) It must avoid confusing self-deception with related conditions and reducing it to one or more of them.
> (5) It must not depend on appeal to notions which are even more puzzling or paradoxical.
> (6) It must account for the fact that 'deception' is used both in cases of interpersonal and intrapersonal deception.
> (7) It must be amenable with, or incorporate, the credible insights of alternative solutions without falling prey to their defects.

1.4 Short Survey of Basic Approaches to the Problem

The enigma of self-deception has recently entertained the attention of philosophers in both the analytical and existential schools, the most prominent writers of the latter tradition who have dealt with the issue being Sartre and Fingarette. Fingarette has authored the only major book on the subject, and his challenging study has gained a sympathetic hearing in some journal articles[54] and has been influential in the outlook of a number of dissertations.[55] Because Fingarette's approach to self-deception challenges a central aspect of the thesis to be developed herein, it will be discussed at the end of chapter 2. However, Sartre's discussion of self-deception ("bad faith") will receive little further attention. I will purposely omit a consideration of Sartre's fascinating treatment of bad faith--with its famous illustrations--for a number of reasons. One reason that is not a factor here is that Sartre's study is devoid of value. His critique of the psychoanalytic theory of unconsciousness and his persistent stress on unavoidable personal

---

[54] E.g., Hauerwas and Burrell, "Self-Deception and Autobiography"; Ronald B. de Sousa, "Review Discussion: Self-Deception," *Inquiry* 31, no. 3 (Autumn 1970):308-321; Joseph F. Fell, Review of *Self-Deception* by Herbert Fingarette, *Philosophical Quarterly* 20, no. 8 (July 1970): 290.

[55] E.g., Bruce, "Investigation of Self-Deception"; J. M. Russell, "Self-Deception"; Shapiro, "Self-Deception."

responsibility (especially if one is to escape self-deception) are well worth pursuing in another setting. However, Sartre does pay attention to a notion of "self-deception" which is more than a bit esoteric and to the side of the more popular conception. Indeed, it would require a dissertation in itself to consider the alternative interpretations of Sartre's works and pin down a proper understanding of those varied writings and many dark sayings. Even the available literature which touches on Sartre's specific perspective regarding self-deception is characterized by interpretive argument[56]; major studies conflict with each other, some viewing Sartre as generally congruous with Freud[57] or Fingarette,[58] and some pitting him strongly against Freud[59] and Fingarette.[60] I cannot enter into the hermeneutical dispute here. Moreover, it is equally noteworthy that Sartre's aim in discussing bad faith is to assert, rather than to resolve, the paradox of self-deception. Bad faith is possible, he says, because human reality must be what it is not, and not be what it is.[61] Accordingly bad faith indicates to us that nothingness (and thus absolute freedom) is essential to consciousness and exists within the for-itself. The present aim and philosophical context or commitments of this study are not identical with the difficult, if not obscure, metaphysic propounded by Sartre. While he wishes to accept the incomprehensible impasse that in bad faith someone knows and is ignorant of the same thing at the same time,[62] my own presupposition would be closer to that announced by Saunders: "If the notion of self-deception were really self-contradictory, there would be no such thing as self-deception: for there cannot be any instances of a self-contradictory notion."[63] Rather than asserting that the incoherence in self-deception is not merely apparent and drawing out philosophical conclusions from that, the present examination is guided by the desire to resolve apparent paradox. The fundamental difference with Sartre cannot be resolved here. Finally, it is noteworthy that Sartre himself interprets his famous anecdotes

---

[56] E.g., Canfield and Gustavson, "Self-Deception"; cf. John King-Farlow, "Self-Deceivers and Sartrian Seducers," *Analysis* 23, no. 6 (n.s. no. 96) (June 1963):131-136.

[57] Shapiro, "Self-Deception," chapter 5.

[58] J. M. Russell, "Self-Deception," chapter 4.

[59] John Buchanan Cochran, "An Investigation into Self-Deception: Three Dialogues," Ph.D. dissertation, University of Oregon (Ann Arbor, MI: Xerox University Microfilms, 76-19,379, 1976), chapter 4.

[60] Oser, "Invitations to Self-Deception," chapter 3.

[61] Sartre, *Being and Nothingness*, p. 100.

[62] Gardiner, "Error, Faith, and Self-Deception," pp. 225-226; cf. Amelie Oksenberg Rorty, "Belief and Self-Deception," *Inquiry* 15, no. 4 (Winter 1972):398.

[63] Saunders, "The Paradox of Self-Deception," p. 559.

of bad faith as involving the postponing of a moment of decision, the fluctuating of an agent's conception or representation of his situation, and the pretense of acting out a role. As such his illustrations would not count as genuine cases of self-deception, for reasons that will become apparent in the subsequent discussion.

In the analytic tradition a variety of treatments have been given to the familiar notion of self-deception with the aim of removing the apparent difficulties in it.[64] In light of the way in which we commonly arrive at perplexity over self-deception these strategies can be conveniently outlined. People often speak of self-deception and seem to understand what they mean. Yet when they try to spell out what is involved in self-deception, we normally observe a rather tacit compulsion to model self-deception on cases of interpersonal deception (other-deception). That is, when S deceives R, he does so by getting R to believe something contrary to what S himself believes. S may believe, for instance, that this new sports car belongs to his roommate, but he purposely gets his female companion, R, to believe that it belongs to himself. In abort, S believes that p ('the sports car does not belong to S'), while R has been made to believe that not-p. The beliefs held by deceiver and deceived are incompatible. Having this kind of case in mind, people tend to construe self-deception as the same sort of activity, with the exception that S and R are the same person. Hence while S believes that p, he persuades himself to believe that not-p. He lies to himself and comes to believe that the sports car is really his after all. That strikes us as peculiar and leads to apparent paradox.

In response, different philosophers have travelled alternate routes) in detouring the conceptual jam. First, some altogether reject the parallel between self-deception and other-deception, 'resisting the common compulsion to model intrapersonal deception on purposeful interpersonal deception. Among them there are those who come to the skeptical view that there is no such thing as actually perpetrating a deception on oneself; strictly speaking there is no such thing as self-deceit, for deceit is inherently other-regarding.[65] What we commonly call self-deception requires a more

---

[64] Various categorizations of these approaches can be found in: Shea, "Self-Deception," pp. 2-3; Bruce, "Investigation of Self-Deception," pp. 34-35; Alan R. Drengson, "Self-Deception," Ph.D. dissertation, University of Oregon (Ann Arbor, EM:, Xerox University Microfilms, 72-14,723, 1971), pp. 2-3; Shapiro, "Self-Deception," pp. 16ff., 115ff., 254ff.; Szabados, "Self-Deception," pp. 52-59; de Sousa, "Review Discussion: Self-Deception," pp. 308-309.

65 E.g., Paluch, "Self-Deception"; A. E. Murphy, *The Theory of Practical Reason* (La

accurate description. Others contend that words such as 'deceive,' 'know,' or 'believe' are used in a somewhat non-standard fashion; when used in an account of self-deception, such words have a different intended sense than in descriptions of other-deception.[66] Finally, among those who deny the parallel between self-deception and other-deception we find the recommendation that we "look and see" what conditions hold when people use the term 'self-deception.' The conclusion is that self-deception does not involve incompatible beliefs (as it does in other-deception, S and R believing contrary things) but is only a particular kind of (single) belief held under peculiar circumstances. Accordingly 'self-deception' locutions are viewed as figures of speech which we use as a reprimand when someone is responsible for holding an unwarranted belief,[67] or self-deceived beliefs are seen as those held in belief-adverse circumstances,[68] those concerning which one has become unsure yet irrationally refusing to reflect on the evidence,[69] and those which are held simply because they are desired,[70] weak-willed dishonesty has permitted desire-generated blindness,[71] or some emotion has irrationally obscured the contrary evidence.[72]

A second major route taken by philosophers has been, over against the first approach, to accept the model of other-deception as appropriate for understanding self-deception; self-deception is naively like other-deception with only one person involved, and thus there actually is such a thing as perpetrating a deception on oneself. Accordingly they maintain that self-

Salle: Open Court Publishing Co., 1965); Champlin, "Self-Deception: A Reflexive Dilemma." However, Champlin is willing to speak of 'self-deception' (rather than 'self-deceit'), which he sees as unintentional and associated with being "dishonest with oneself" in some unexplained and unclear sense

[66] E.g., Mounce, "Self-Deception."

[67] E.g., Siegler, "Demos on Lying to Oneself," *The Journal of Philosophy* 59, no. 16 (August 2, 1962):469-475; "Self-Deception," *Australasian Journal of Philosophy* 41, no. 1 (May 1963):29-43.

[68] E.g., Canfield and Gustavson, "Self-Deception"; Terence Penelhum, *Philosophy of Mind*, ed. by Stuart Hampshire. Sources in Contemporary Philosophy Series (New York: Harper and Row, Publishers, 1966).

[69] E.g., Richard Reilly, "Self-Deception: Resolving the Epistemological Paradox," *The Personalist* 57, no. 4 (Autumn 1976).

[70] E.g., Shea, "Self-Deception"; Eugene Valberg, "Rationality and Self-Deception," Ph.D. dissertation, State University of New York at Buffalo (Ann Arbor, MI: Xerox University Microfilms, 73-29,146, 1973).

[71] E.g., Drengson, "Self-Deception"; he likens this to some forms of unintentional other-deception.

[72] E.g., Charles B. Daniels, "Self-Deception and Interpersonal Deception," *The Personalist* 55, no. 3 (Summer 1974).

deception is a conflict state wherein S holds incompatible beliefs of some nature; he aims to induce in himself a belief contrary to some other belief already held and succeeds. In order to resolve the apparent paradox of believing contrary things, various kinds of distinctions are then introduced. In a strategy similar to one seen above, some writers have said that, while S does believe something he knows (or believes) to be false, nevertheless this involves a special kind of knowledge (e.g., "as-it-were-knowledge")[73] or slightly different kinds of belief (e.g., full belief and half-belief)[74]; these different senses for the epistemic vocabulary render the paradox only apparent. Other philosophers treat self-deception as though it were just like other-deception, that is, a case of two different persons. In this vein some posit a duality within the person himself (i.e., the self is not united),-So that one part of the soul deceives another part of the soul (or the person believes in one part of the soul but disbelieves in another).[75] On the other hand, some see self-deception as involving a temporal distinction between S-the-deceiver and (later) S-the-deceived.[76] In contrast to such solutions which take the model of other-deception so literally, there are a host of philosophers who propose that self-deception is a conflict state wherein one must draw some distinction centering on the notion of awareness (in one sense or another). Accordingly we read of the difference -between two levels of awareness,[77] two kinds of consciousness, general and explicit,[78] general awareness and detailed awareness,[79] conscious purpose and unreflective purpose,[80] conscious and unconscious knowledge,[81] or strong and weak consciousness.[82] By use of these kinds of distinctions writers have hoped to maintain that self-deception can involve a conflict of incompatible beliefs, but not one which precludes a coherent and feasible account of how S can successfully perpetrate a

[73] E.g., Hamlyn. "Self-Deception."

[74] E.g., Eric Joseph Lerner, "The Emotions of Self-Deception," Ph.D. dissertation, Cornell University (Ann Arbor, MI: Xerox University Microfilms, 75-27,038, 1975); Siegler, "Analysis of Self-Deception."

[75] E.g., King-Farlow, "Self-Deceivers and Sartrian Seducers"; Rorty, "Belief and Self-Deception."

[76] E.g., Béla Szabados, "Rorty on Belief and Self-Deception," *Inquiry* 17, no. 4 (Winter 1974):464-473; "Self-Deception," *Canadian Journal of Philosophy* 4, no. 1 (September 1974):51-68.

[77] E.g., Demos, "Lying to Oneself."

[78] E.g., Bruce, "Investigation of Self-Deception."

[79] E.g., Pugmire, "'Strong' Self-Deception."

[80] E.g., Gardiner, "Error, Faith, and Self-Deception."

[81] E.g., Hamlyn, "Self-Deception."

[82] E.g., Saunders, "Paradox of Self-Deception."

deception on himself; the self-deceiver is not fully aware that he believes something he knows to be false.

Finally, in addition to schools of thought which have rejected and accepted the model of other-deception in explaining self-deception, a third proposed route out of the conceptual jam is to utilize a completely different model altogether which avoids appeals to such epistemic terms as 'knowledge' or 'belief.' .Such a new analysis might instead utilize a volition-action model of self-deception, wherein S fails to "spell-out" his engagements in the world. It is thought that in this way we can preserve the elements necessary to an adequate account of the phenomenon (e.g., purposiveness, motivation, culpability) and yet avoid the paradoxes inherent in the epistemic accounts of self-deception.[83]

Although the main purpose of this study is not to interact systematically with every one of these proposals as such, the various views found in the recent literature will be taken up for discussion as they bear on particular aspects of the development of my own proposed thesis. It will turn out, I believe, that none of these three major routes described above will pass the tests of adequacy which were previously listed. In some cases we find necessary, but not sufficient, conditions Of self-deception set forth (e.g., adverse evidence, influence of desire). In other cases necessary conditions will be dismissed altogether (e.g., belief, incompatible beliefs). Some of the proposals merely restate the need for a resolution to the problem (e.g., new senses for the epistemic vocabulary used) or else reintroduce the paradox at a different point (e.g., having a policy of not spelling-out an engagement in the world). Further suggestions reduce self-deception to something else (e.g., change of belief, ignorance, cognitive error, pretending), thereby rendering the notion dispensable. Another group of attempted resolutions rely on notions which are even more obscure or problematic than self-deception itself (e.g., diverse kinds of consciousness), and they escape the appearance of paradox at the price of equivocation on just what the self-deceiver believes (or is aware of). Often self-deception is reduced to one of many related states or actions: e.g., wishful thinking, ignorance, error, delusion, simple trust, vacillation, obstinance, motivated belief. While helpful insights are contributed by virtually all of the authors, I am not convinced that they have been fully true to the phenomenon or escaped paradox.

---

[83] E.g., Fingarette, Self-Deception.

1.5 The Aim and Significance of This Study

The object of this study is the resolution of the apparent paradox of self-deception. Its significance, if it can claim any, is that it accomplishes this end and passes the tests of adequacy laid out before.

In searching for a basic analysis of self-deception we are fully cognizant of the fact that a wide variety of forms and types of self-deception overlap and crisscross. Self-deception can be about many things (circumstances, thoughts, feelings, emotions, desires, character traits, personality, capabilities, talents, plans, motives, personal relations, facts, life's meaning, etc.), pursued in various ways (perception, memory, reasoning, etc.), and engaged for various general reasons (to blind one to the painful, to help one feel good, to enable one to refuse the distressing truth, etc.).[84] To be sure, "the family of self-deceivings is large and heterogeneous,"[85] and "it is highly misleading to speak of a 'typical' case of deception when this involves overlooking the very many different kinds of case there in fact are."[86] These are useful warnings against reductionism and concentrating on a narrow range of cases. We are duly chastened by Wittgenstein's words about a "craving for generality" which is a "contemptuous attitude toward the particular case,"[87] and we can sympathize with the consternation over the way in which philosophers can impose preconceived pictures or draw boundaries for special purposes, which triggered Wittgenstein's preference for the "family resemblance" approach to general words.[88]

However, it is pressing the diversity of self-deceivings too far when some writers go on from this observation of variety and this obvious warning about reductionism to say: "To try to fit all the ways in which we deceive

---

[84] Cf. Shapiro, "Self-Deception," pp. 3-4; Shea, "Self-Deception," chapter 10; Lerner, "The Emotions of Self-Deception," chapter 8; Rorty, "Belief and Self-Deception," passim; Bruce, "Investigation of Self-Deception," pp. 147-149, 178-186; Valberg, "Rationality and Self-Deception," pp. 186-206; J. M. Russell, "Self-Deception," chapter 4; Dilman and Phillips, *Sense and Delusion*, passim; Drengson, "Self-Deception," chapter 6; Cochran, "Investigation into Self-Deception," passim.

[85] John King-Farlow, Review of *Self-Deception*, by Herbert Fingarette, in *Metaphilosophy* 4, no. 1 (January 1973):84.

[86] Drengson, "Self-Deception," p. 92; cf. pp. 81-82.

[87] Ludwig Wittgenstein, *The Blue and Brown Books* (New York: Harper Torch Books, 1965 11958]), p. 18.

[88] Ludwig Wittgenstein, *Philosophical Investigations* 3rd ed., trans. G. E. M. Anscombe (New York: Macmillan Co., 1968 [1958, 1953]), sections 65-71.

ourselves into one broad definition is also probably impossible"[89]; "no single list of necessary and sufficient conditions for the existence of self-deception can be formulated."[90]

> Thus, it is concluded that to more fully understand self-deception, we must first identify concrete cases and then examine carefully the contexts from which these cases emerge. No unified account of self-deception, no recipe, no definition, no formula, no essential rules to account for all cases of self-deception-suffice—we are left with concrete cases and their contexts. . . . We must use an <u>ad hoc</u> strategy-in dealing with cases of self-deception and try to understand each case on its own merits. We want to bring in good sense. [91]

But just how are we to be able to "identify concrete cases" and "bring in good sense" if we have nothing but an "ad hoc strategy"? There must be some standard by which genuine cases are picked out. Surely there is some explanation for the fact that the same word, 'self-deception,' is predicated of these diverse cases and for the fact that philosophers (among others) distinguish between phenomena which do and phenomena which do not belong to the class of self-deception. There parallels with the typical case can be drawn). Thus the proposed elucidation and description of self-deception in its typical, or full-fledged, or paradigmatic sense will not amount to an all-encompassing account in some uniquely competent, ideal language; yet it will be entirely adequate, provided it passes the tests laid out previously and can account reasonably for the atypical cases which arise. A sufficient and acceptable analysis will have been found if it descriptively conforms to the ordinarily recognized phenomenon and is conceptually clear and consistent.

There are several good reasons for engaging in a philosophical analysis of the notion of self-deception. First, the notion is a familiar but perplexing one, as noted already. It seems-that it cannot make sense, and yet it must make sense. The resolution of a paradox which touches on much of what we say and think, especially about human nature and behavior, is surely interesting in its own right for a philosopher. Seeing through the paradox of self-deception is a worthwhile goal, then, for this study.

---

[89] Shapiro, "Self-Deception," p. 4.
[90] Rorty, "Belief and Self-Deception," p. 408.
[91] Cochran, "Investigation into Self-Deception," pp. 83, 89.

But secondly, the Investigation of this notion involves giving attention to related matters of perennial philosophical interest-- concepts like belief, desire, rationalization, motivation, consciousness, intention, etc. The proper analysis of self-deception could shed light to some extent on these other topics as well.

Thirdly, although it is not the purpose of this study to do so, many philosophers who have addressed the question of self-deception have done so as a vehicle for advancing significant philosophical doctrines pertaining to such matters as: the explanation of conduct,[92] consciousness,[93] emotions,[94] desire and will,[95] rational belief,[96] thought,[97] the self and the mind,[98] motives and intentions,[99] freedom and determinism;[100] man's basic moral nature as good or evil,[101] etc. The notion of self-deception carries with it a great deal of philosophical baggage for many writers; it is an intellectually potent concept for influencing one's fundamental perspective on man and ethics. That is why authors point out that, although one should not make too many moralistic evaluations in advance, [102] self-deception cannot easily be explained in morally neutral terms,[103] and one's attitude toward the self-deceiver's responsibility will determine the account he offers of self-deception.[104] With so much argued on the basis of self-deception it is certainly requisite that we begin with a clear idea of what it is and is not; from that perspective a better

---

[92] E.g., Fingarette, *Self-Deception*, chapters 3-4.

[93] E.g., Bruce, "Investigation of Self-Deception," chapter 4; Oser, "Invitations to Self-Deception," chapter 2.

[94] E.g., Lerner, "Emotions of Self-Deception," chapters 7-9.

[95] E.g., Shapiro, "Self-Deception," chapters 3-4, 6.

[96] E.g., Valberg, "Rationality and Self-Deception," passim.

[97] E.g., J. M. Russell, "Self-Deception," chapter 2.

[98] E.g., Bruce Wilshire, "Self, Body, and Self-Deception," *Man and World* 5, no. 4 (November 1972):442; Fingarette, "Self-Deception," passim; Rorty, "Belief and Self-Deception," pp. 404-406; Shapiro, "Self-Deception," chapters 1, 7.

[99] E.g., Gardiner, "Error, Faith, and Self-Deception," pp. 237-243; Saunders, "Paradox of Self-Deception," passim; Shapiro, "Self-Deception," chapter 4.

[100] E.g., James N. Jordan, "On Comprehending Free Will," *The Southern Journal of Philosophy* 11, no. 3 (Fall 1973):184-201; Oser, "Invitations to Self-Deception," passim; Cochran, "Investigation into Self-Deception," chapter 4.

[101] E.g., Shapiro, "Self-Deception," chapters 5-7; Chaplin, "Self-Deception: A Reflexive Dilemma," pp. 297-298; Szabados, "Morality of Self-Deception," passim; Gardiner, "Error, Faith, and Self-Deception," pp. 221-224; Hamlyn, "Self-Deception," pp. 58-60; Mounce, "Self-Deception," pp. 70-72; Hauerwas and Burrell, "Self-Deception and Autobiography," passim.

[102] King-Farlow, "Review of *Self-Deception*," pp. 79-84.

[103] Dilman and Phillips, *Sense and Delusion*, p. 72.

[104] Cochran, "Investigation into Self-Deception," p. 83; cf. p. 17.

evaluation of these further doctrines may be gained.

Finally, if nothing else, a study of self-deception can serve as a negative starting point for the project of self-knowledge (e.g., understanding one's aims, personality, relationships with others, significance of actions and decisions). "Whatever is obscure about self-deception infects our understanding of what it is to be a person, what it is to know oneself, and what it is to act responsibly."[105] The present study of self-deception, then, holds promise for further application to the endeavor of attaining self-knowledge. If Socrates is to be believed, this would be prerequisite to self-improvement and thereby to the social beatitude of mankind.

1.6 A Sketch of the Proposed Solution

In the subsequent study I will maintain, in short, that self-deception involves an indefensible belief about one's beliefs.'[106] That is, S perpetrates a deception on himself when, because of the distressing nature of some belief held by him, he is motivated to misconstrue the relevant evidence in a matter and comes to believe that he does not hold that belief, although he does. When he holds a belief that is discomforting, the self-deceiver simultaneously brings himself to believe that he does not hold it, and toward the end of maintaining that unwarranted second-order belief he presses 'into service distorted and strained reasoning regarded the evidence which is adverse to his desires. Be not only hides from himself his disapprobated belief, but when he purposely engages in self-deception he hides the hiding of that belief as well.

This thesis will be developed according to the following synopsis of chapters. Chapter two asserts that self-deception involves a person's beliefs and that these beliefs are not themselves somehow odd or defective in character. Contrary to Fingarette, one cannot avoid a belief-analysis of self-deception, and the analysis given need not use the word 'belief' in some special or "twilight" sense. Moreover, the analysis can be given simply in terms of belief, without construing self-deception in terms of a mixture of knowledge and belief. The self-deceiver sees p as evidenced (whether it actually is or not); p presents itself to him in some way as the truth. That S believes p is indicated by his behavior (in much the same way as the

---

[105] Fingarette *Self-Deception*, p. 1.

[106]   Ronald B. de Sousa, in his "Review Discussion: Self-Deception also approaches self-deception with the distinction between first-order and second-order beliefs in mind, although he develops the helpful suggestion in a somewhat different direction.

dispositional analysis of belief would suggest). However, his assent to p is thwarted because S induces in himself an unwarranted 1 belief that he does not believe p. The avowal of this second belief often functions for S as the "taking of a stand" on his identity as a person and for the protecting of his interests. The appreciation that people can believe things without assenting to them, and that what people give assent to believing (i.e., what they explicitly profess about their beliefs) is fallible, is made crucial to understanding the phenomenon of self-deception.

Chapter 3 argues for the appropriateness of modeling self-deception on other-deception. I will contend that there is a common sense for the word 'deception' in both other-deception and self-deception. This does not commit us to going to the extreme of making self-deception out to be a literal case of other-deception or like it in every detail; rather, self-deception will be seen as a general parallel to other-deception in certain specifiable ways. For instance, elements of deception which are shared by both other-deception and self-deception are the deceiver's responsibility for causing the deceived to believe falsely, the fact that *the deceived holds (at least implicitly) an erroneous belief about the deceiver's beliefs, and the rationalization maneuvers taken in the face of evidence brought to the attention of the deceiver by others.

Chapter 4 explains why incompatible beliefs need to be attributed to the self-deceiver on the basis of his behavior. Accordingly self-deception is portrayed as a conflict state wherein S holds incompatible beliefs. In essence the self-deceiver holds a first-order belief (viz., the belief that p) which is, far from being a matter of personal indifference to S, somehow distressing to S; he has a personal stake in (or against) it. It is a special kind of belief: one which S dreads, cannot face up to, or wishes were otherwise, since it brings some unpleasant truth (as he perceives it) before him. Thus S comes to deny- that belief. While believing that p, he comes to hold additionally a (false) second-order belief about this belief, namely, that he does not believe that p. Due to the pain it would otherwise cause him, S cannot believe that he believes p. Although it need not do so, this denial of his belief can take the form of, or be facilitated by, S coming to believe not-p as a way of counteracting his belief that p. The behavioral symptoms of believing that one does not believe p and the behavioral symptoms of believing not-p, it will be observed, often overlap extensively; in the examination of one's actions, both may be easily taken as disbelieving p. The incompatibility of beliefs which is found in self-deception, where S believes that p and yet believes that he does not believe that p, is not logical in nature but behavioral and practical; these beliefs are not formally contradictory. For that matter, even when self-deception leads S

to believe that p and to believe that not-p, this is not formally contradictory. The two beliefs held by the self-deceiver are incompatible because they call for conflicting types of behavior. In particular, S believes p, but his assent to it is blocked by acquiring the second-order belief that he does not believe p.

S has an obvious interest at stake in maintaining the rationality of his second-order belief (contrary to the first-order belief that p); thus chapter 5 contends that this state of mind comes about when, in the face of evidence adverse to his cherished second-order belief, S engages in contrived and pseudo-rationality in his treatment of the evidence. That is, he manipulates, suppresses, and rationalizes the evidence so as to support a belief incompatible with his belief that p; he ignores the obvious, focuses away from undesirable indicators, twists the significance of evidence, goes to extreme measures to enforce his policy of hiding the belief that p. This rationalizing activity, in order to count as self-deception and not something else (e.g., a cavalier disagreement), must be given a motivational explanation. S distorts the evidence in order to satisfy a desire-- namely, the desire to avoid the discomfort or pain associated with believing that p. By means of it he enters into and maintains self-deception, believing that he does not believe that p. Actions which have the effect of achieving the special state of incompatible beliefs which has been traced above are referred to in statements like 'S is deceiving himself regarding p.'

As human actions they may be purposively engaged (or done intentionally), but they need not be. Chapter 6, thus, takes up the vexed questions of awareness and purpose in self-deception, addressing what is perhaps our underlying perplexity in making sense of self-deception. It will be maintained that, while the self-deceiver is aware of the truth of p, he does not entertain it before his mind (as it were, internally perceiving it) and give it assent. Accordingly the self-deceiver is not aware that he holds incompatible beliefs; after all, he does not believe that he has a belief that p, but only that he does not believe that p. He should recognize the incompatible situation, but the strategy of hiding his dreaded belief prevents it; if he did recognize the incompatibility and did not resolve it, he would simply be irrational or vacillating. Further, the self-deceiver is not aware that his professed (and cherished) belief about his beliefs is false, even though as a rational man he should. The critical question, however, is whether one can <u>try</u> to deceive himself and not be aware of these things. Can one engage in self-deception on purpose? The common assumption is that, if S purposes to do something, then he must be aware of its character. Thus if S can purposely engage in the activity of self-deception (i.e., rationalizing the evidence so as to hide

a dreaded belief), it would seem that he would be aware of its character; he would believe that he is attempting to deceive himself, and that would foil his effort--just as much as if R realized that S was intending to mislead him from the truth in a case of other-deception. I will argue that S's awareness of his aim to make the belief that p covert (by believing something incompatible with it) will not undermine the success of his deception effort. What S thinks about in his purposeful attempt at self-deception need not be deception-defeating, for the intention to deceive oneself can be self-covering; that is, S can purpose not only to hide his belief that p, but also to hide his hiding of it. The self-deceiver conceals his intention from himself, or in short, deceives himself about his intentions. To avoid an infinite regress of self-deceptions in the case of purposeful self-deception, then, it must be possible for self-deception to be self-covering: obscuring itself in the process of obscuring S's belief that p, and yet without calling for a further intention regarding itself in this matter. This will be likened to the intention to go to sleep.

Chapter 7 summarizes the analysis of self-deception which has been developed in chapters 2-6. The truth conditions for 'S deceived himself' can be listed as these at that point:

1. S believes that p.
2. S is motivated to ignore, hide, deny (etc.) his belief that p.
3. By misconstruing or rationalizing the evidence, S brings himself to believe that he does not believe that p (or alternatively, to believe that not-p).

Finally, it is shown how this account of self-deception passes the various tests of adequacy which were previously formulated and specified. It refrains from contradiction, does not reduce to related conditions, and is satisfactory in explaining common illustrations of self-deception. Therefore, we may conclude that the paradox has been resolved, even though what is disturbing to us about the phenomenon itself has not been eliminated.

Chapter Two
Self-Deception and Belief

2.1 Self-Deception Involves False Belief

Deceived people believe false propositions. This should seem obvious, and there are few if any plausible grounds for disputing the point. However, before our analysis can begin with this elementary assumption, a few clarifying remarks are necessary in order to prevent misunderstanding and discourage misconceived disagreement. When S is deceived, he is engaged in some kind of cognitive error: he has been misled, deluded, beguiled, or somehow mistaken in what he believes or expects about something. R promises to meet S at the library but deceives him thereby, for when R does not show up it turns out that S's belief that R will be there is wrong. While S is waiting, R is enthralled with the speech of a political propagandist across campus, being taken in by his lies and thereby deceived. Most writers, in conformity with ordinary usage, treat deception in such a fashion as a form of erroneous belief, especially in recent studies of self-deception; however, Fingarette's novel belief-less approach to self-deception will need to be given a reply at the end of this chapter.

There are secondary uses of the word 'deceive' that might give the impression that believing false propositions is only contingently related to deception. A Freudian psychologist might speak of someone in the grip of unconscious motivations as self-deceived, and yet not necessarily ascribe any cognitive processes to the person at all. Fingarette portrays self-deception as a matter of volition, of failing to be what we are, and thereby leading false, inauthentic, or meaningless lives; this reflects a general existentialist conception of self-deception and its significance. However, these uses of the term are either figurative (being based on the central, cognitive sense of 'deception') or have the cognitive sense implicit in them. As Shapiro notes:

Without being able to cash in talk of unconscious motives or true selves into descriptions of ourselves that we can believe, the notion of being false to ourselves or leading untrue lives really doesn't make sense. . . . If it makes sense to say that our lives are false and our existence empty it also makes sense to say that we believe wrongly or fail to understand "real life" and "true self."[1]

Kierkegaard spoke of deception in the sense of a failure to be ethically consistent with oneself. In a similar vein T. S. Champlin, following up on the illustration used by D. W. Ham1yn, points out that a man who has been unfaithful to his wife, even when she is not ignorant of the fact, is said to have "deceived her."[2] In its definition of 'deceive' *The Oxford English Dictionary* includes such obsolete senses as "to betray into mischief, to deal treacherously with, to be (or prove) false to, to play false, to cheat (out of)." Thus Champlin finds philosophical significance in the fact that the verb 'deceive'

. . . leads a double life: either it functions on its own or else it enters into a variety of syntactic structures using prepositions and subordinate, clauses. . . . You can be deceived <u>tout court</u>, or you can be deceived <u>about</u> etc., <u>in</u> (e.g., 'one's belief, expectation') etc., and <u>into</u> etc. The wife in Hamlyn's story was deceived. tout court but she was not deceived about, in or into anything at all.[3]

Such ethical uses of 'deceive' (<u>tout court</u>) may be seen as metaphors built up from a familiar form of vice (e.g., in being unfaithful to his wife, it is as though the husband lied to her about his promise of fidelity) or seen against a cultural background where people were thought to stand in various relations of moral obligation to each other (in which case they could expect certain .kinds of behavior from others, the form of social interaction being implicitly promised). In the first case it is as if the wife were led to believe falsely that her husband would honor his vow; in the second case any moral misdeed could falsify our belief that the offender would behave in the prescribed manner. So then, when we speak of 'deception' it appears that the sense of cognitive error is always present, either directly or waiting in the wings.

Apart from these linguistic considerations, there are certain situations

---

1   Shapiro, "Self-Deception," pp. 31, 32.
2   T. S. Champlin, "Double Deception," *Mind* 85, no. 337 (January 1976):100-102.
3   Ibid., p. 101.

which suggest that self-deception might appear to be exemplified without the presence of false belief. These lead Shapiro to assert more generally that "what is essential to self-deception is failure to know or believe what the available evidence or conditions warrants."[4] For instance, he suggests that a man is "deceived" not only when he has a false belief, but also when he has a true belief which is not justified by the available evidence (e.g., a patient who disbelieves his doctor's studied diagnosis of cancer and turns out to be free of the disease after all). Here we have, not false belief, but unwarranted belief.[5] However, this is confused. We would not, after the fact, say of a man who actually held a true belief all along that he was deceived; in light of the adverse evidence we would more likely say that he had strong faith.[6] Conceivably such a man could be deceived about the status of the evidence, giving evidence in his conduct and discussion that he really did not think that there were any serious indicators of failing health with which to reckon. (Although unmentioned, such an added condition to Shapiro's example might explain why he felt it to be plausible.) However, that would again be a case of false belief--about the evidence, rather than about his health.

A second alleged example of self-deception without false belief is given by Shapiro; whereas the former one attempted to illustrate the absence of falsity in a self-deceiver's belief, this one emphasizes the omission of any belief that one should have on the basis of the evidence. "The cuckold who fails to believe that his wife is unfaithful, when the neighborhood is loud with rumors and his head is full of shady late night excuses, is very apt to be self-deceived."[7] Shapiro goes on to stress that he does not mean that this man holds a belief contrary to the evidence to which would, after all, destroy the point of his illustration), but that he prevents himself from thinking anything at all. He has no belief whatsoever on the matter. Apart from the air of unreality about this example, the fact remains that it is not likely to be deemed a case of self-deception after all; there is to be found no assent, no assertion, no awareness, no avowal on this man's part, and thus it is hard to see in what sense he is deceived. He is simply ignorant--precisely like the person used in an earlier example posed, then rejected,-by Shapiro. Previously he mentioned someone who refuses to read a book or someone who chooses to ignore sources of adverse evidence, and his comment then was: "A person who avoids holding beliefs and doesn't make any claims as a result of his ignorance, may be someone who is reprehensible . . . but he

4 Shapiro, "Self-Deception," p. 40.
5 Ibid., pp. 37-38.
6 Cf. Penelhum, "Pleasure and Falsity," p. 259.
7 Shapiro, "Self-Deception," p. 39.

isn't a self-deceiver. In general, where the truth knowingly goes unsought or lays unclaimed, deception is not possible." Not until the person mistakenly describes his situation *vis a vis* the evidence or is by default led to hold false beliefs is there self-deception.[8] Similar comments are in order for Shapiro's later example which, due to its formal parallel to the earlier, can be advanced only at the price of inconsistency.

Therefore, to return to the observation at the opening of this chapter, and putting aside misconceived linguistic usage or misleading illustrations, we can reiterate that people who are deceived are said to be such because they (minimally) believe false propositions. ,They have an affirmative epistemological attitude which falls short of knowledge (since knowledge cannot be of false propositions).[9] The role of belief in self-deception is simply assumed in virtually all of the literature treating the subject, with the exception of Fingarette. This is not to say that people do not deceive themselves about their hopes, attitudes, emotions, etc. (e.g., false pride, false security); Fingarette is correct to see such matters as important in a full exposition of the phenomenon of self-deception.[10] However, even these "objects" of self-deception in common parlance have a cognitive core.[11] Without the element of belief, such emotions and attitudes would not obtain. The parent who is inappropriately proud of his son's report card from school must at least believe that the card has high marks. Even when emotion appears as the object of self-deception, cognitive language is natural:
"He may think that he is sorry about Smith being fired, but he knows quite well that he is delighted over it." We will, then, assume from henceforth, and we will make it foundational to an analysis of self-deception, that mistaken belief is essential in being deceived.

> At the core of our lives, there is, no doubt, something of a mess. We are incoherent in our choices, inconsistent in our beliefs, falsely persuaded that we know ourselves. But at the core of all such self-deprecation, there is the concept of Belief: and it too is something of a mess.[12]

---

8   Ibid., pp. 32-33.
9   Cf. Lerner, "Emotions of Self-Deception," p. 131.
10   Fingarette, *Self-Deception*, p. 33.
11   Szabados, "Self-Deception," p. 57. The close connection between belief and emotion states is discussed by H. H. Price, *Belief*, Muirhead Library of Philosophy Series (London: George Allen and Unwin, 1969), pp. 267-275. See also Lerner, "Emotions of Self-Deception," chapter 3.
12   Ronald B. de Sousa, "How to Give a Piece of Your Mind: or, The Logic of Belief and Assent," *The Review of Metaphysics* 25, no. 1 (September 1971):52.

2.2 A Characterization of Belief for Analyzing Self-Deception

2.2.1 Difficulty and Significance of the Question

If self-deceived people believe false propositions, then an understanding of self-deception will be facilitated by a proper conception of belief itself. Yet a thorough discussion, critical interaction, and analysis of belief is beyond the scope of the present study. In the immediately preceding quote from Ronald de Sousa we are told that, not only is (false) belief at the core of our failure at self-knowledge but the notion of belief is itself troublesome as well.

> It is connected with so much of our lives, that we are inclined to fragment it. One is tempted to view it in relation to action alone; or, on the contrary, merely as a mental act of disposition to such acts; or else, to change the subject altogether and discuss some idealization.[13]

There are numerous facets to, or questions about, the concept of belief; there are many and conflicting views of the nature of belief. The concept plays several contrasting roles: with doubt, with action, with knowledge. And there are several ways in which philosophers speak of belief itself: conscious belief, behavioral belief, unconscious belief, rational belief,[14] occurrent belief, dispositions' belief, half-belief, opinion, conviction, commitment, assent, being under an impression, etc.[15] People speak of degrees of belief in various senses: degrees of assent, degrees of evidential support, degrees of emotional conviction, degrees of determinateness (from vague to specific).[16] Notorious questions surround the notion of belief: what is it for someone to believe that p? Is believing p something we do, or something we have? What are the truth conditions for 'S believes that p'? What does the sentence mean? What kind of concept is the concept of belief?[17] Belief is a central and pervasive

---

13   Ibid.

14   Robert J. Ackermann, *Belief and Knowledge* (Garden City, New York: Anchor Books of Doubleday and Co., 1972), pp. 1-10.

15   Cf. Price, *Belief*, passim; also Paul Helm, *The Varieties of Belief*, Muirhead Library of Philosophy Series (New York: Humanities Press, 1973), passim.

16   D. J. O'Connor, "Beliefs, Dispositions and Actions," *Proceedings of the Aristotelian Society* 69, n.s. (1968-1969):1-2.

17   John Stafford Peal, "The Theory of Belief," Ph.D. dissertation, University of North

aspect of human experience and philosophical reflection:

> Belief or believing is one of the most pervasive and-familiar aspects
> of human experience. We have all sorts of beliefs. Some of these
> are deeply ingrained in us. Some are of importance; some are even
> cherished by us. Others are fleeting and are held self-consciously
> but without much conviction or care. Some are acquired and held
> by us unconsciously or implicitly. Belief or believing not only is an
> aspect of human experience, but its emphasis is a central problem
> in epistemology, metaphysics, and the philosophy of mind.[18]

So then, it is not hard to see how vast and complex would be an adequate
study of belief. It would likewise be philosophically controversial, as Bertrand
Russell's remark hints: ". . . on the view we take of belief our philosophical
outlook depends."[19] The subject of belief is broad, many-faceted, subject to
conflicting opinions, philosophically important, and pervasive in our lives.

It is not surprising, therefore, that different approaches to belief affect
one's outlook on self-deception. For instance, it is possible to define 'belief'
in such a way as to preclude self-deception.[20] Some who may disagree with
the analysis of self-deception which will finally be given here will likely do
so because they more fundamentally disagree with something maintained
regarding belief. What that would indicate, most likely, is that they are
employing a specialized notion of belief, one which sets boundaries on it by
philosophical choice. Whereas the corresponding disagreements over self-
deception would "save" some philosopher's extraordinary notion of belief,
the account offered herein will have the advantage of saving the ordinary
notions of belief and self-deception--rather than dooming the latter to
incoherence in advance. Quine has rightly observed that when we assess
beliefs we assess several in combination, rather than one at a time.[21] And
this is undoubtedly true of our beliefs about the notions of belief and self-
deception; what we think about the one will greatly affect our treatment of
the other, and vice versa. In the long run the adequacy of anyone's analysis of

---

Carolina at Chapel Hill (Ann Arbor, MI: Xerox University Microfilms, 75-15,687, 1974), p. 6.

18   Ibid., p. 1.

19   Russell, *Analysis of Mind*, p. 231.

20   E.g., Bernard Mayo, "Belief and Constraint," in *Knowledge and Belief*, ed. A. Phillips
Griffiths (London: Oxford University Press, 1967), p. 156.

21   E.g., W. V. Quine and J. S. Ullian, *The Web of Belief* (New York: Random House,
1970), pp. 8ff.; cf. Quine, "Two Dogmas of Empiricism," *From a Logical Point of View*, 2nd ed.
(New York: Harper Torchbooks, 1963 (1953)0, pp. 41ff.

belief or self-deception will probably be judged jointly. For all that, we must still settle for less than a thorough treatment of the subject of belief at this point since an entire volume cannot be inserted here.

However, not only would it be impractical to attempt a full-fledged analysis of belief at this juncture, it might also be unwise in view of the fact that such an analysis appears impossible to give. Numerous able philosophers have put their hand to the task of laying down necessary and sufficient conditions of belief, but without satisfaction. When they have been descriptively accurate, they have dealt with but one aspect of belief or believing, but when they have rendered an account intending to give a complete analysis, they have strayed from descriptive accuracy. I would suggest that the reason for this is that the notion of belief is primitive to explanations in the philosophy of mind, and as such belief cannot be traditionally defined in terms of anything more basic.

2.2.2 Preliminary Survey and Critique

Let us briefly survey some of the main accounts offered of belief and the drawbacks of them. Two views explicitly concern the grammar of 'belief.' The first holds that the statement 'I believe' is not used to make an autobiographical report or to describe the believer; rather, by means of such an utterance, one expresses an attitude or takes a. stand on something. That is, first person, present tense, belief sentences are performative in function.[22] But this can only be an aspect of the total picture, for it does not equally apply to second and third person uses of the verb (i.e., 'I believe' can be used to assert a proposition, but 'he believes' cannot), there surely is a self-ascriptive use of 'I believe' which is autobiographical and must be taken into account (e.g., "Like you, I believe at present that Nixon was duped"), and finally 'I believe' is sometimes used when one does not believe at all (cf. "To make a lying promise is to promise; to make a lying statement of belief is not to believe").[23] A second account of the grammar of 'believe' holds that this verb has no present continuous sense; moreover, such a verb has no descriptive sense in the first person present tense (since it can be used in a number of places in an indicative sentence). Further, the first person use of belief sentences carries an implied claim for the truth and reasonableness of the statement associated with the verb.[24] However, it is implausible that second

---

22   E.g., Price, *Belief,* pp. 29-37; cf. Peale, "Theory of Belief," pp. 14-22.

23   A. Phillips Griffiths, "On Belief," in *Knowledge and Belief,* ed. A. Phillips Griffiths (London: Oxford University Press, 1967), p. 132.

24   J. O. Urmson, "Parenthetical Verbs," *Essays in Conceptual Analysis,* ed. Antony Flew

and third person belief sentences signal us how the believer understands the statement, and they do not naturally carry an implied claim for the truth and reasonableness of what is believed. Thus at best this is a limited account of the grammar and use of 'believe.' But more fundamentally, about such analyses as the one 'before us, I find myself continually believing that they are shortsighted and mistaken. The claim that first person belief sentences are not descriptive and have no continuous present tense is controverted by the immediately preceding sentence in this paragraph. Therein I am described as holding a particular belief for a continuous period of time, and this report is made by means of the verb 'believe' itself. Turning from grammatical analyses of the verb, we can look briefly at the two major analyses of the nature of belief as an epistemic notion in general. These are, of course, the occurrent and dispositional views of belief which have been debated back and forth, the former pointing to belief as a datable mental event and the latter seeing it as a readiness or inclination (perhaps latent or unobserved) Ito respond in certain ways when an appropriate issue or circumstance arises.[25] Concern with <u>one's own</u> beliefs has generated the mentalist occurrence theory of belief as an introspectible event which happens in a person, whereas from the point of view of others' beliefs we come to the dispositional account of belief as something an agent has and which is manifested in his behavior (whether accompanied by a conscious formulation of a sentence in mind or not).[26]

Various definitions of belief have been offered by advocates of the occurrence analysis: to entertain p with a feeling of conviction, to entertain p while having evidence for it,[27] to have a mental attitude of favor,[28] to give reasoned assent to an entertained proposition.[29] Basically, belief is here an inner state of mind, directly accessible to introspection, wherein a believed sentence is explicitly formulated; the believer is aware of this sentence (or at least can become aware of it when asked an appropriate question), and special emphasis in this analysis is usually placed on the act of assenting to such a sentence. For example, H. H. Price offered an analysis of belief earlier

---

(New York: St. Martin's Press, 1956), pp. 192-212; cf. Peale, "Theory of Belief," pp. 22-25.

25   E.g., C. A. Campbell, "Towards a Definition of Belief," *Philosophical Quarterly* 17, no. 68 (July, 1967):204-220; Murdith 'McLean, "Episodic Belief," *Philosophical Quarterly* 20, no. 81 (October 1970):389; Quine and Ullian, *Web of Belief*, pp. 3-4.

26   Cf. Price, *Belief*, pp. 19-21; Ackermann, *Belief and Knowledge*, pp. 5-8; Peale, "Theory of Belief," pp. 9-10, 141.

27   Cf. Griffiths, "On Belief," pp. 132-133.

28   E.g., Campbell, "Towards a Definition of Belief."

29   E.g., H. H. Price, "Some Considerations About Belief," in *Knowledge and Belief*, ed. A. Phillips Griffiths (London: Oxford University Press, 1967).

in his career which saw it as the entertainment of a proposition accompanied by the mental act of assenting to it; this assent was made up of two elements in turn: volitional (preferring p and dissenting from its alternatives) and emotional (feeling a conviction, in some degree, about p).[30] In favor of an occurrent or mentalist account of belief are certain facts about self-ascription of belief. In attributing beliefs to ourselves we do riot always do so in any ordinary observational or inductive fashion, as we do in ascribing beliefs to others; moreover, self-ascription of belief is often immediate, performed without deliberation, and yet made with a very high degree of accuracy (when checked against behavioral indicators).[31]

Nevertheless, difficulties can be seen in the occurrent account of belief. Without modification or qualification it would suggest that when we are not thinking consciously about the objects of our beliefs (e.g., when we are sleeping or thinking about some other subject matter) we cease believing them. It appears that this account confuses belief with coming to or acquiring, along with subsequent reflection on, a belief. Many versions, moreover, automatically preclude unreasoned beliefs or unconscious beliefs from their account even though such beliefs are ordinarily countenanced in common discourse. Descriptive accuracy is also compromised regarding the feeling of conviction which allegedly accompanies belief; the fact is that such a feeling is not always present even with our consciously entertained beliefs (e.g., as when they are taken for granted unless challenged by someone). Some elements in the occurrent analysis are only obscurely explained. For instance, what is this act of preferring (in contrast, say, to an emotional preference)? And what is the mental act of entertaining a proposition? Price characterizes it as understanding the proposition (being aware of what it would be like if the proposition were true), and then later as thinking of something <u>as</u> something else; yet these two descriptions are not at all coextensive. A further stumbling block in the occurrent analysis is found in its portrayal of belief as a unique, logically private, mental experience. If belief were such, then it would teem that this experience could not be described by reference to anything else, in which case there would be no way of telling---apart from surreptitiously employing ordinary behavioral criteria--when anyone, including oneself, believed anything. For whatever its strengths, the mentalist account of belief ends up underplaying the explanatory role of belief sentences; beliefs are commonly, if not usually, ascribed to ourselves and others in order to explain

---

30   Ibid., pp. 41-59.

31   Robert C. Coburn, "Believing Things," *Canadian Journal of Philosophy* 1, no. 1 (September 1971):101-102; Mitchell Ginsberg, *Mind and Belief: Psychological Ascription and the Concept of Belief* (New York: Humanities Press, 1972), pp. 32, 47-48.

observed behavior.[32] Later in his career H. H. Price came to be skeptical of acts of-belief and finally adopted a tempered dispositional analysis of belief; while remaining critical of some expositions of dispositional belief (e.g., where to believe p is to "act as if p were true"), Price was willing to advance a compromise between the dispositional and occurrent analyses of belief. In a major volume on the subject[33] he continued to say that mental events of assenting do occur, but now they became categorized as one of the relevant manifestations of the complex disposition called 'belief.' One of the multiform manifestations of belief is the occurrent and introspectible experience of assenting to the object of the belief, especially at the initiation of the disposition.[34] Nevertheless, at base belief is a disposition.

Various versions of the behaviorist or dispositional approach to belief have been maintained; belief is analyzed as: to behave in some way (regardless of an agent's reasons), to behave in specific ways in specific situations, to have a disposition to behave,[35] to act as if p were true, to have a disposition to act as if p were true (given that p is thought of), or to entertain p and have a tendency to assert it,[36] etc. Basically, believing is not an activity or a mood, but is the possession of a disposition to respond in particular ways when an appropriate issue arises. To analyze the sentence 'S believes that p' calls for something like an inductive hypothesis, a complex series of conditional statements about what S would do if certain delineated circumstances arose; such a conjunctive sentence would include counterfactual conditionals. The disposition which is called 'belief' is seen to be actualized in many different kinds of behavior and action (which may be deemed symptoms of belief). Often it is specified that the proposition which is said to be believed must at some time or another be entertained by the believer, a period of time must be specified during which the disposition is possessed, and the given analysis of a belief (i.e., the series of conditional statements about S's behavior) applied

---

32   de Sousa, "How to Give a Piece of Your Mind," pp. 53-54.

33   Price, *Belief*, pp. 130-301; cf. Antony Flew, review of Belief, by H. H. Price, in Mind 79, no. 315 (July 1970):454-460; A. Phillips Griffiths, review of Belief: *The Gifford Lectures at the University of Aberdeen in 1960*, by H. H. Price, in Philosophy 46, no. 175 (January 1971):63-68; Konrad Marc-Wogau, review of *Belief*, by H. H. Price, in *Philosophical Review*, 81, no. 2 (April 1972):246-250; Alan R. White, review of *Belief*, by H. H. Price, in *Philosophical Books* 10, no. 3, (October 1969):21-23.

34   For a similar compromise see John Perry Casey, "Knowledge, Belief, and Evidence," Ph.D. dissertation, University of Iowa (Ann Arbor, MI: Xerox University Microfilms, 75-23,022, 1975), pp. 235-239.

35   Cf. Casey, "Knowledge, Belief, and Evidence," chapter 9.

36   Cf. Griffiths, "On Belief," pp. 128-129.

to one individual only.[37] R. B. Braithwaite offers an example of a dispositional analysis of belief when he says that the nature of believing is to entertain a proposition along with a dispositional readiness to act as if the proposition were true (even in one's verbal behavior), given the particular desires that the agent wishes to satisfy.[38] The dispositional approach to belief is favored by the fact that it avoids the problems of accounting for a completely private object (as we find in the mentalist analysis), and by the fact that others can often check our self-ascriptions of belief against our behavior (just as we ourselves can) and overrule thereby our claims to believe when appropriate behavior does not accompany them. Observation of behavior, rather than private introspection, comes to justify any and all ascriptions of belief. An agent is not in a privileged position to find out what and if he believes, for as the proverb says, "Acts speak louder than words."[39]

However, the dispositional understanding of belief is not trouble-free. Of such an account we must ask, as we did in the case of the occurrent analysis of belief, just what entertaining a proposition amounts to and whether it is not dispensable in many instances of belief. It seems relatively rare when we attend consciously to the many propositions we believe. Braithwaite's explanation of entertaining is particularly confusing since he wants to say that it is not a disposition, but the understanding of a linguistic sign. Yet understanding is in some sense a disposition to act in appropriate days under relevant circumstances, and one can surely understand many things (e.g., how to do something, questions) without making use of a linguistic symbol like a sentence in the mind. But even putting aside problems with the notion of entertainment, the dispositional account of belief has a hard time rendering a convincing and helpful translation of belief sentences into behavioral hypotheses. In the case of first person belief sentences it is quite unlikely that one's claims about one's beliefs are understood by him as inductive conjectures about what he would do under certain circumstances. And in the case of second and third person belief sentences it is very difficult to specify what a claim to belief amounts to. Shall the various condi! tional statements about the believer's appropriate behavior be related to each other by conjunctions or disjunctions? If the former, then one single action of a person could not fully count as the manifestation of his belief, and if the latter, then one insincere action of a person (e.g., his acting as if he believed something) could nevertheless count as indicating a genuine belief.

---

37 E.g., Price, *Belief*, pp. 246-266; Quine and Ullian, Web of Belief, pp. 3-4, 6.

38 R. B. Braithwaite, "The Nature of Believing," in *Knowledge and Belief*, ed. A. Phillips Griffiths (London: Oxford University Press, 1967), pp. 28-40.

39 Coburn, "Believing Things," pp. 93-94; Price, *Belief*, pp. 248249, 254-256.

The dispositional understanding of belief would also, if not sufficiently qualified by mentalist terms, justify us in attributing a belief that p to S just in case he had appropriate behavior, even though he had never heard of p nor deliberated consciously about mot. Thus the dispositional analysis of belief cannot provide an account of what belief sentences mean in general because of the obvious fact of human variability in other beliefs, aims, desires, emotions, etc.[40] For that matter, the fact of variation in mere physical abilities is a major problem for a dispositional analysis of belief; a man who is conscious, yet completely paralyzed, would have no tendency to any kind of behavior in the ordinary behaviorist sense, and yet he surely can have beliefs nonetheless (e.g., about overheard comments of loved ones, about his chances for recovery, about impending death). Finally, the facts of <u>akrasia</u> ("weakness of will") and self-deception (acting in a way contrary to one's sincere belief-ascriptions) are both very resistant to integration into a dispositional account of belief.[41] Many men who genuinely believe that they are mortal die without making wills--either not doing what they believe they ought to do or being somehow misled to believe that they could accurately prophesy the time when they would need to have their affairs in order prior to death.

Attempts to analyze belief, then, have turned out to be less than (fully satisfactory; while each effort has something to commend itself, all of them have failed to be descriptively accurate with respect to the entire range of believings that are familiar to us. Proposals for defining belief, it will be noted, have not achieved an analysis which incorporates necessary and sufficient conditions <u>except</u> at the price of <u>circularity</u> or arbitrary <u>restriction</u> on the concept. For instance, when the advocate of a dispositional analysis tells us that 'S believes that p' means that in a particular Circumstance S would do some Action, it should be rather obvious that the analysis will prove to be a true one only if S believes that he is in that particular Circumstance, believes that his behavior is properly designated as that Action, and believes that this Action is in fact appropriate to this Circumstance. That is, a true dispositional analysis of belief can be given only by mentioning 'belief' in the analysis itself, thus becoming circa tar. This can be illustrated in terms of a man who believes that toadstools are poisonous and yet, when served toadstools, readily eats them (without any suicidal motivation, etc.). Since the man has not acted appropriately in terms of what would be a natural dispositional analysis of 'S believes that toadstools are poisonous,' the analyst must salvage his account by

40   Ackermann, *Belief and Knowledge*, pp. 18-19.
41   Price, *Belief*, pp. 27, 257.

systematically justifying the man's behavior in terms of false beliefs: e.g., the man does not believe that what he has been served is a toadstool (confusing it, for instance, with a mushroom). This will render a successful analysis of the belief in question, but only in terms of belief itself.[42] Likewise, circularity is evident when Price looks upon belief' as a disposition which has various manifestations, among which he lists (for instance) hope; yet Price elsewhere explicates the nature of hope in terms of entertaining p, along with certain feelings toward it and certain beliefs about its possibility.[43] So also, analyses of belief which use the word 'sincere' in their account of belief will prove to be circular and unhelpful. Ronald de Sousa sets forth a theory of belief which sees it as a "bet on the truth" along the lines suggested by Bayesian Decision Theory; the notion of assent is central to his account of belief, and about it he says that "in so far as assertion is sincere or candid, we can abstract from it a bet on truth alone, solely determined by epistemic desirabilities."[44] What trivializes this attempt to explain belief is simply that a sincere assertion is precisely one which accords with the agent's beliefs; thus belief is explained in terms of assertions, but only those which the person actually believes.

When discussions of belief are not running in the kind of circles observed above, they are often pursuing a very specialized version or notion of belief, restricted in advance so as to be amenable to credible analysis. Accordingly, in Price's earlier account of belief he makes assent crucial to belief, but then narrowly defines assent so as not to countenance unreasoned beliefs as genuine beliefs; he draws a distinction between believing something and taking something for granted, but this is quite arbitrary and manifests the philosopher's predilection to impose preconceived boundaries for a special purpose.[45] Likewise, when some philosophers declare that statements embodying Moore's paradox (e.g., 'p, but I don't believe that p,' or 'I believe p, but don't believe that I believe p') are inconsistent, it is not unusual to find that in advance they have restricted their discussion to the narrow and idealized boundaries of "conscious, rational belief."[46] Many analyses of belief display this very character: they analyze not belief in general, but a very special and circumscribed subclass of beliefs. The predominant restriction placed on the

---

42   Griffiths, "On Belief," p. 238; Casey, "Knowledge, Belief, and Evidence," pp. 225-232; Price, *Belief*, pp. 253-254, 260.

43   Price, *Belief*, pp. 295-296, 267-272; Griffiths, "Review of Belief," pp. 63-64.

44   de Sousa, "How to Give a Piece of Your Mind," p. 59.

45   Price, "Some Considerations About Belief," pp. 48ff.; cf. Casey, "Knowledge, Belief, and Evidence," pp. 191-192.

46   E.g., Ackerman, *Belief and Knowledge*, pp. 22, 27.

beliefs to be analyzed is the one we have seen illustrated here; the restriction would preclude any responsibility to take account of beliefs which are not fully consistent or arrived at rationally. The commentary of de Sousa is apt:

> Nothing seems to follow strictly from 'X believes that p.' But if we reinterpret it to mean: 'X can consistently be described as consistently believing p'--which roughly renders, I think, Hintikka's notion of "defensibility" [in his *Knowledge and Belief*]--we can get on with the subject, freed from the inhibitions of descriptive adequacy. But defensibility is neither necessary nor sufficient for truth: it tells us little, therefore, about the concept of belief on which it is based. It cannot, in particular, specify necessary conditions for the consistent ascription of *belief*--as opposed to *rational* belief.[47]

In a related matter Robert Audi maintains his thesis that disavowals of knowledge (or belief) are corrigible and defends it against a particular criticism, saying:

> We should not multiply senses beyond necessity; and the fact that there is both confident knowledge and knowledge whose belief component falls short of "full confidence" no more justifies positing two senses of "know" than the fact that there are both upright and grand pianos justifies positing two senses of "piano."[48]

It is simply unprofitable to give a supposed analysis of belief for philosophical purposes when one is really giving a sense to 'believe' which is suited to his idealized philosophical outlook. The concept of belief then gets counterfeited and is no longer currency which can be traded in a wide variety of philosophical contexts. It is worth noting here that, <u>even</u> in the restricted context of conscious and consistent belief, there is no completely satisfactory philosophical analysis of the narrow notion of "rational belief" since certain paradoxes (e.g., the lottery paradox) force us to abandon at least some of our intuitions about the criteria of rational belief. So then, some idealized agent could possibly satisfy one philosopher's account of rational belief, only to appear irrational to at least some other philosophers who have constructed a different account of rational belief.[49] This is all the more reason to refrain

---

47   de Sousa, "How to Give a Piece of Your Mind," p. 52.

48   Robert Audi, "Epistemic Disavowals and Self-Deception," *Personalist* 57, no. 4 (Autumn 1976):380.

49   Ackermann, *Belief and Knowledge*, pp. 40, 41, 50.

from artificial boundaries when we are attempting to analyze the common notion of belief. But the underlying and relevant fact is that because it is simply wrong to say that absurd beliefs (for instance) are not genuine beliefs, all efforts to restrict beliefs to rational beliefs I must fail.[50] "Human nature is capable of more-things and stranger things than common sense philosophers suppose," for like it or not the fact is that "it is beyond our power to be wholly rational all the time."[51] Indeed, even when one is interested in being reasonable and recognizes an inconsistency in his beliefs, it is still possible for him to hold to--and wittingly confess--the irrational set of beliefs: for instance, the neurotic leaves his psychiatrist, saying "It's terrible; I want to kill my father, but I don't believe it"; or the philosopher who recognizes that he has granted that two conflicting arguments are of sufficient force to establish contradictory conclusions and depends alternately on them in different contexts, yet without denying the conclusion contradictory to the one he presently uses, might say "p and q are incompatible, and nevertheless I genuinely believe them both." We could, of course, stipulate that such mental states cannot exist and that the word 'believe' will be ruled out of such cases, but what would that ultimately accomplish? As Keith Graham has said:

> I believe, on the contrary, that such mental states exist not only in the possible worlds explored in philosophy journals but also in the real world of imperfectly rational human beings. . . 'If you ignore it, it will go away' is a common piece of advice. The philosophers whose views I have opposed argue, on the contrary, that in the case of false or inconsistent belief if you pay attention to it, it will go away. My reason for opposing them is the belief that one stands a far smaller chance of ridding the world of falsehood and irrationality if one takes the short cut of defining it out of existence.[52]

The reason why some philosophers have been hesitant to concede the psychological possibility of people holding irrational and wittingly inconsistent beliefs is explained in basically the same way by Graham and Price: these philosophers overlook the fact that belief is neither completely passive nor

---

50   Griffiths, "On Belief," p. 136; Peale, "Theory of Belief," p. 130.

51   Price, Belief, pp. 227, 238; cf. pp. 215, 261, 298.

52   Keith Graham, "Belief and the Limits of Irrationality," *Inquiry* 17, no. 3 (Autumn, 1974):323, 325. The examples used immediately above are taken from Graham; they are similar to those offered by Arthur W. Collins, "Unconscious Belief," *The Journal of Philosophy* 66, no. 20 (October 16, 1969):667-680.

completely active, thereby overemphasizing either its involuntary or voluntary character.[53] They "have not always been clear about the distinction between the mental processes which just go on in us automatically, and those which are rationally and consciously controlled."[54] We will return to this insight below in characterizing belief in such a way as to facilitate an analysis of self-deception.

The preceding survey and critique of various attempts to analyze belief (or 'believe') has been belabored for a few reasons. First, it illustrates why a full-fledged analysis of belief is not possible here and why thorough answers to the complex questions pertaining to belief as an epistemic notion cannot be worked out now. The subject is too extensive and challenging. Second, the preceding survey has alerted us to particular errors we should avoid in understanding belief as we use the notion in our analysis of self-deception; in particular we have noted the descriptive inadequacies, circularities, and arbitrary restrictions of many preferred accounts of belief, while at the same time observing various features about the conflicting analyses which commend themselves to us. Finally, the above survey has been rehearsed and the particular kinds of criticisms delineated in order to point out that belief is very likely a primitive notion which cannot be analyzed in terms of necessary and sufficient components more basic than itself.

2.2.3 General Characterization: A Positive Propositional Attitude

Rather than maintaining that 'S believes that p' describes the goings-on in S's head, a disposition of the person, that something is being actively entertained, or any one particular feeling or active experience,[55] it might be far better to admit that the question 'What is belief?' is not answerable in a general description that covers all and only cases of belief[56] Relevant to a difficult question about belief, Quine admits that some philosophical problems are better deflected than met head on.[57] The best advice available to us at this juncture is perhaps that offered by John Peale in his doctoral dissertation, which argues at length that if belief cannot be formally defined we can at least give it an accurate characterization (responsive to the valuable

---

53   Graham, "Belief and the Limits of Irrationality," pp. 323-324.
54   Price, *Belief*, p. 230.
55   Such assumptions are questioned by Helm, *Varieties of Belief*, p. 60; and Collins, "Unconscious Beliefs," p. 676.
56   Cf. Griffiths, "On Belief," pp. 138-139...
57   Quine and Ullian, *Web of Belief*, p. 4.

points of the various analyses which have been offered heretofore).[58] Such a characterization can serve us well here. By it we can-adequately account for self-deception, using the postulation of belief in order to explain various aspects of a person's behavior, and yet heeding Price's advice against maintaining specialized concepts that are untrue to their common usage and thereby hinder a natural explanation of our experience.

> If we insist on such a narrow and rigid terminology, we deprive ourselves of insight into the complex facts of human nature, by depriving ourselves of the linguistic tools which are the most handy ones for describing them and discussing them.[59]

What follows, then, is a general characterization of belief intended to reflect the way in which the notion is ordinarily used. The characterization makes no claim to being complete. I will comment only on aspects of belief which are particularly relevant to the final end in mind of giving an analysis of self-deception; related topics of interest will not be explored or elucidated at any length, and some will not even receive mention. Objection to this incomplete characterization of belief in order to gain a basic framework within which to offer a philosophical understanding of self-deception will only be devastating here if what I have to say about belief is not only a failure to answer all of the questions about it, but a failure to answer any. My goal is not a systematic account of belief; but of self-deception; some things said about the former may turn out to be problematical (due to the narrow focus I bring to it), but I trust not in any way that will substantially affect the discussion of self-deception as a whole. Because enough philosophers have defended the various elements of my characterization of belief, my remarks will remain programmatic and not be given a thorough discussion and defense in this place.

We may characterize belief as a positive, intellectual, propositional attitude (not excluding false propositions), expressed in a large variety of symptoms, some of which are subject to degrees of strength; such belief may be, but is not necessarily, achieved consciously and rationally (i.e., constituted by a preference among alternatives which have been weighed against relevant evidence).[60] We can take up aspects of this characterization one by one.

In the first place, belief is a propositional attitude. Truth and falsity are

---

58 Peale, "Theory of Belief," pp. 11-13, 116ff.

59 Price, *Belief*, p. 300.

60 This characterization is similar to that offered by Peale, "Theory of Belief," pp. 7, 44ff., 76, 121ff.; it also has been largely affected by Price, *Belief*, pp. 130-131.

essential elements in the assessment of beliefs, whether they are our own or others' beliefs; beliefs are thought of as having some discernible content. A condition of believing something is that the predicate 'is true' may be applied to it; adding this to a description of belief is always legitimate and yet supplies no new-distinction (e.g., "S believes that p" and "S believes that p is true assert the same thing). Therefore, the object of a belief is whatever one considers the bearer of truth within his conceptual Outlook.[61] Here we will treat propositions as the bearers of truth and thus the object of beliefs. If on the other hand we took sentences to be the objects of belief, it would imply that, when S believes 'Die Preise werden erhot werden' and R believes 'The prices will be raised,' they do not believe the same thing--which is clearly wrong. What S and R believe is the proposition which is customarily asserted by the German and English sentences here. It is this proposition that is either true or false, regardless of whatever S and R think about it. Propositions facilitate the translation of one natural language into another, and they may be viewed as that which is asserted when a declarative sentence in any natural language is used to assert a truth. All propositions, and only propositions, can be either true or false. As beliefs are said to be either true or false, the content or object of a belief will be taken to be a proposition. Thus beliefs are propositional attitudes and not, as others have suggested,[62] predicates expressing properties that are exemplified by either propositions or persons. In the first case it hardly seems that the alleged property expressed by 'S believes' is exemplified by the proposition asserted in 'the prices will rise,' in the same way that the property expressed by 'is spherical' is exemplified by the object denoted by 'a baseball.' In the second case, if belief is a complicated property of a person (denoted, e.g., by 'believes that-the-prices-will-rise') which is not specified or individuated propositionally, then no indexical cross-reference within the set of a person's beliefs can be exhibited, which would preclude testing for consistency of beliefs,[63] and would render languages unlearnable.[64] Consequently, we will be content to treat belief herein as a propositional attitude.

Further, belief may be characterized as a pro-attitude of the intellect

---

61   Griffiths, "On Belief," pp. 134-135; Peale, "Theory of Belief," pp. 117-121.

62   E.g., Willard Van Orman Quine, *Word and Object* (Cambridge, Massachusetts: Institute of Technology Press, 1964 [1960]), p. 216.

63   Ackermann, *Belief and Knowledge*, pp. 15-20.

64   Donald Davidson, "Theories of Meaning and Learnable Languages," *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, ed. Y. Bar-Hillel (New York: Humanities Press, 1965), pp. 383-394.

toward a proposition; it is a cognitive, in distinction from conative or affectional, attitude that is favorable to some proposition. At this point we can profitably reintroduce the earlier observation that belief is in some ways a voluntary matter and in some ways involuntary; the latter is stressed when we think of it as a positive intellectual attitude toward a proposition. To believe p is to take p for true or to regard it as true,[65] which explains why we customarily think that to believe p is to be favorably disposed to asserting it.[66] Our beliefs are in a sense constrained by the way we view the evidence relevant to them; they are not under our voluntary control as though we could regard just any proposition as true by an act of the will here and now.[67] Try as we may we cannot simply believe contrary to the facts as we understand them; by an act of will I may pretend to believe otherwise, but I cannot genuinely believe right now that I am not composing on a piece of paper. Some features of our ordinary speech manifest this involuntary or automatic response to the perceived evidence which is involved in belief: e.g., we speak of "being constrained to believe p," "forced to conclude that p," "being unable to resist the weight of evidence," etc.[68] These mechanical metaphors of constraint suggest that, in some sense, we can no more resolve to believe something that we do not actually see as true (or evidenced) than we can create in ourselves a taste (e.g., for peanut butter) or desire (e.g., to weed the garden) by a simple effort of the will. Our stock of beliefs changes slightly every waking moment of our lives, and we continually gain beliefs without choosing to do so or even being aware that it is happening.[69] Belief is like seeing as[70]— regarding something as having evidence ("evident-ness"), seeing something as being the case, regarding it as reliable or worthy of confidence, apprehending it-.as true in. light of the evidence, recognizing the pattern and significance of the evidence (of whatever kind, from perception to testimony). Hence we can often think of beliefs as pictures that have taken hold of and govern our lives[71]; it is impossible to believe contrary to the way in which the evidence is taken and controls our outlook. In this vein we speak of some things being

---

65   Cf. Helm, *Varieties of Belief*, pp. 81, 101; Carey, "Knowledge, Belief, and Evidence," p. 232. David Pugmire says, "What we take to i be true is simply what we believe," in "A Doubt About the Normative Theory of Belief," *Mind* 81, no. 324 (October, 1972):586.

66   E.g., Mayo, "Belief and Constraint," p. 161.

67   Price, *Belief*, p. 222.

68   Cf. Mayo, "Belief and Constraint," p. 152.

69   Cf. Peale, "Theory of Belief," p. 134; Quine and Ullian; *Web of Belief*, pp. 3-4; Price, *Belief*, pp. 230-231, 238.

70   Cf. Helm, *Varieties of Belief*, p. 152.

71   Cf. ibid., p. 137.

"unbelievable" or "beyond belief."[72] As long as the evidence is received or seen in such and such a pattern the corresponding belief is compelled in us, and our reliance upon the believed proposition will usually be seen in our counting upon its truth in our theoretical and practical inferences (e.g., seen in the "spreading" of belief that p to belief that q, etc.).[73] Even beliefs which are normally deemed unreasonable (e.g., dogmatic, superstitious beliefs held contrary to widely recognized evidence, or sheer commitments of faith) Will turn out to rest on something which is regarded 12x the believer as a kind of evidence or warrant calling for that belief and making it appropriate. It is important to stress here, of course, that the evidence which is said here to "in-form" one's belief or compel its determinate character is not "the total available evidence" (a concept whose nature is somewhat unclear)[74] but rather that which is-regarded` by the subject as evidence. Against this he cannot believe.

2.2.4 Specific Characterization: Inferential Reliance

Up to this point the discussion has been restricted to a general characterization of belief as a propositional attitude of a positive, cognitive type. To believe that p is to take p as true, to regard it as reliable, to see it as supported by the evidence. So much seems clear. But the details of the matter are not as clear and--as briefly sketched above--are quite difficult to provide. It seems precarious to go beyond the cautious and minimal characterization laid out above, all the more since this study is specifically devoted to another notion. And yet, to make use of the words of Wilfrid Sellars in facing the same problem, "It would be foolhardy--indeed downright mistaken--to claim that this formula captures 'the' meaning of 'believes,' . . . for, as with most, if not all, of the words in which philosophers are interested, we are confronted with a cluster of senses which resemble each other in the family way."[75] A successful analysis of self-deception in terms of belief will require a more precise characterization of belief than has been ventured so far. Otherwise the subsequent study might be thought to range back and forth over differing conceptions of belief, thereby being marred by equivocation. A conception

---

72   Cf. Peale, "Theory of Belief," pp. 36-37.

73   Price, *Belief*, p. 293.

74   Ackermann, *Belief and Knowledge*, pp. 34-35; cf. Price, Belief, p. 94.

75   Wilfrid Sellars, "Belief and the Expression of Belief," *Language, Belief, and Metaphysics*, ed. Howard E. Kiefer and Milton K. Munitz (Albany: State University of New York Press, 1970), p. 146.

of belief should be advanced which distinguishes it from related mental attitudes or states, and which can be consistently applied throughout the remainder of this work. This will very likely involve an artificial narrowing of the broader, ordinary usage of the term 'belief,' and it will most certainly not solve, within the framework of the present study, all of the problems which can be raised in philosophical discussions of belief. However what is a drawback or difficulty for all theories in the field is not a particular difficulty for one of them. I can hope that whatever problems may remain can be worked out with a modicum of judicious philosophical industry, consistent with the' lead suggested here.

In order to supply a precise conception of belief which can be consistently applied in the subsequent study of self-deception I will stipulate one of the manifestations of belief which (like other candidates) conforms to the general characterization set forth already. This conception does not pretend to be the only way to speak of belief or necessarily the only conception that can successfully render an analysis of self-deception; I do not assume that the treatment to be suggested now is complete, balanced, or solely adequate. For this precise conception of belief I will rest almost completely upon the work of F. P. Ramsey, H. H. Price, and D. M. Armstrong. The thesis stemming from their writings will not be argued for in this work since that would deserve and require the extensive task of another philosophical treatise. Nevertheless the answer proposed by such writers as these commands respect. It-is initially plausible, and it is not at obvious variance with-Our ordinary understanding of the notion of belief. Using their account we can easily point to clear instances of belief and recognize how they differ from related phenomena (although they surely are not instances beyond the logical possibility of controversy). But no attempt will be made, anyway, to defend this characterization of belief from conceivable criticisms which might be made of it; for that one will have to turn to the original works of these men themselves. My aim will simply be to employ their conception of belief so as to demonstrate a credible way of resolving the apparent paradox of self-deception.

Therefore the present study does not do anything quite so grand as analyzing, arguing, and settling every philosophical issue related to an analysis of self-deception. In fact, as far as the overall argumentation of this thesis is concerned, the stipulation of a precise conception of belief renders the final analysis of self-deception conditional. What I will maintain is that if the following account of belief is utilized, the apparent paradox of self-deception can be resolved in the manner to be suggested later. Yet it may be that this conception of belief is itself problematic, in which case the paradox of self-deception will have been traded in for a different form of philosophical

problem--perhaps reducing it to some other paradox about belief or to a puzzle over the correct way to portray belief. Such issues as those must, of practical necessity, be decided elsewhere. At any rate, the specific analysis of self-deception to be proposed herein would still be worthwhile in itself and could be commended over alternatives being currently advanced. Although there will be, of course, much more to be said about wider problems, I hope that the brief remarks made here about belief (bolstered by our common use and understanding of the notion) will suffice for the present discussion and allow a consistent portrayal of self-deception. So, at any rate, I will assume. As remarked earlier, it will probably turn out that belief and self-deception are a "package deal" of concepts, so that an understanding of each will be accepted or rejected together with an understanding of the other.

Many philosophical efforts to give an account of the notion of belief, even those of some very fine scholars, have resulted only in suggesting manifestations of belief and/or paraphrases for 'belief, 'believes,' etc. Others have tended to offer metaphors for belief. For instance, at one point in the above section it was said that to believe was to be "in-formed by the evidence" in a particular way; this metaphor presents a person's belief that p as a matter of his mind being imprinted or stamped in a certain way. In the *Theaetetus* (191c-e) Plato used the illustration of an imprint made by a seal on a block of wax, an imprint enduring for a greater or lesser length of time. Another portrayal of belief mentioned above is that of a "Picture that governs one's life"; such a model was occasionally used by Wittgenstein.[76] In the *Tractatus* he compared sentences to pictures and in other places he continued to hold that "in general, there is nothing which explains the meanings of words as well as a picture."[77] This probably inspired the mathematical philosopher F. P. Ramsey to; compare belief to a map. In one of his posthumously published papers Ramsey said that a particular belief is "a map of neighbouring space by which we steer."[78] This portrayal makes two characteristics stand out: belief is a map, and belief is something by which we steer. From this seminal metaphor D. M. Armstrong develops an elaborate account of belief.[79]

---

76   L. Wittgenstein, *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief*, compiled from notes taken by Yorick Smythies, Rush Rhees, and James Taylor, ed. Cyril Barrett (Berkeley: University of California Press, 1972), pp. 53-56.

77   Ibid., p. 63.

78   Frank Plumpton Ramsey, "General Propositions and Causality, *The Foundations of Mathematics, and Other Logical Essays*, ed. R. B. Braithwaite (London: Routledge & Kegan Paul, [1931] 1954), p. 238.

79   D. M. Armstrong, Belief, Truth, and Knowledge (Cambridge; j University Press, 1973), pp. 3-110.

If we think of beliefs as maps, then we can think of the totality of a man's beliefs at a particular time as a single great map of which the individual beliefs are sub-maps. The great map will embrace all space and all time, past, present, and future, together with anything else the believer takes to exist, but it will have as its central reference point the believer's present self. . . The great belief-map will be much like the maps of old, containing innumerable errors, fantasies and vast blank spaces. It may even involve contradictory representations of portions of the world. This great map, which is continuously being added to and continually being taken away from as long as the believer lives, is a map within his mind. . . . The belief-map will include a map of the believer's own mind, and even, as a sub-part of this sub-part, a map of the believer's belief-map (that is, his beliefs that he holds certain beliefs). . . . The situation is no worse than those actual pictures which contain, as part of the scene pictured, little pictures of themselves . . . .Beliefs, are maps by which we steer. Unlike entertained propositions, beliefs are action-guiding. Entertained propositions are like fanciful maps, idly scrawled out. But beliefs are maps of the world in the light of which we are prepared to act.

The task of the remaining chapters of this Part of the book will be to spell out and articulate in detail Ramsey's suggestion. The suggestion is bold and simple. But, as might be expected, its working out is laborious and complex.[80]

This task leads Armstrong into some complicated and sometimes controversial discussions of a wide range of difficult philosophical issues: consciousness, thought and language, dispositions and states, propositions, concepts, ideas, general beliefs, existential beliefs, conjunctive beliefs, etc. The details thereof will not concern us here--especially since some philosophers would likely reconstruct much of Armstrong's discussion in terms of a different metaphysic--even though such considerations would weigh significantly in any determination of the final virtue of his proposal concerning belief. For present purposes we will focus merely on the basic conception or characterization of belief advanced by Armstrong and further explained by Price (who also takes his lead from Ramsey).

Many of the treatments of belief offered by philosophers portray it as

---

80   Ibid., pp. 3-5, cf. p. 99.

an underlying positive attitude or <u>state of mind</u> (in contrast to some kind of continuous action or conscious occurrence) which can be, but might not be, detected by outside observers or even by oneself--similar, say, to a taste for oysters as a state of the palate. Armstrong does so as well, and he feels that his way of thinking about belief is implicit in Western philosophical thought. "According to this view, A's believing that p is a matter of A's being in a certain <u>continuing state</u>, a state which endures for the whole time that A holds the belief. . . . If beliefs are states, then they will be accidental and changeable features of <u>minds</u> (or, if this is objected to, of <u>persons</u>)."[81] These continuing mental states are non-relational accidental properties of the believing person, properties which are not necessarily a process taking place in him. The manifestations or expressions of the mental state of belief are notoriously multiform,[82] and for Armstrong, "The distinction: between the belief which is not being manifested, and the belief which is, then becomes the distinction between a <u>causally quiescent</u> state, and the same state causally active."[83] This is likened to the "information" in a computer's memory-banks which can be either inactive or playing a causal role in the computing process.

Armstrong distinguishes the mental state of belief from a disposition in three ways, although he does argue that dispositions are a species of state.[84] Dispositional states are stimulus-dependent, whereas belief states are not. Dispositional states are manifested, if at all, in only one way, but belief states have indefinitely many ways to manifest themselves. And while dispositional states need not have any particular structure, a belief state will have an inner complexity somehow corresponding to the proposition that is believed.[85] Armstrong later explicates this last observation in this fashion: "The elements-in-relation to be found in a particular belief-state determine <u>what it is that is believed</u>. . . [and I] will call <u>both</u> the representing elements and the representing relations involved in belief-states 'Ideas.'"[86] Armstrong says further that belief states have a self-directedness. "They have the unique, irreducible characteristic of <u>intentionality</u>. . . . Of their own nature, they point in a certain direction."[87] Belief states, that is, share the self transcending character of consciousness whereby it refers to or intends an object; this

---

81 Ibid., pp. 9, 10.
82 Ibid., p. 24.
83   Ibid., p. 18.
84   Ibid., pp. 10-14.
85   Ibid., pp. 16-21.
86   Ibid., pp. 50, 51.
87   Ibid., pp. 54, 60.

confers upon a conscious being a sort of selective capacity toward things, holds Armstrong.

> It has been argued that A has the concept of X if it is the case that instances of X, acting upon A's mind, bring it about, in suitable circumstances, that a certain sort of state comes to be in A's mind. This state, which is identified with a belief on the part of A that there is something of the sort X in some environmental relation to A, enables A to act towards the X in a discriminatory way, distinguishing it from things that are not X. It may be thought of as a 'map' of the A-X situation: a 'map' in A's mind. But-it seems to be a self-directed map.[88]

To recapitulate, then, belief is here portrayed as a continuing, intentional, mental state with a causal capacity which is stimulus-.independent, which can be quiescent or active, and which takes a variety of manifestations; the ideas which are involved in the belief state give it its determinate character as corresponding to the character of the proposition believed.

What has been rehearsed above corresponds to the map element of Ramsey's view of a belief as "a map . . . by which we steer." The steering element also needs some explication. Belief has been categorized as a mental state, and belief states are represented as differing from each other in terms of the proposition believed--in terms of the ideas which correspond to the structure of the proposition and which are intended by the belief state. These mental states or maps are further individuated by their "different causal powers."[89]

As indicated already above, belief states are either causally quiescent or causally active. This causal feature gives belief its essential character as action-guiding, and it sets belief off from mere thought (or the entertaining of a proposition).

> The difference between a belief and a mere thought is that the former is, while the latter is not, something 'by which we steer.' Given suitable dominant desires (which are also to be looked upon as causal factors), then the belief-state will co-operate with the desire so that they are jointly responsible for the subject's acting in a certain way.

---

88   Ibid. p. 65.
89   Ibid., p. 107.

. . We believe this proposition ['There is a red thing over there') if, and only if, we are prepared to act "towards the thing, or supposed thing, in a 'red-selective' way, if we should desire so to act. The mere <u>thought</u> that there is a red thing over there would be something in our mind which had the same structure (involved the same Ideas), but which was not, even in conjunction with suitable dominant desires, a cause of such selective behaviour.

It is an essential mark of beliefs, then, as opposed to mere thoughts, that, if suitable dominant desires are also present, the believer is moved to action: action having as its objective the satisfaction of the desire.[90]

Armstrong immediately indicates that this is not a description of the complete phenomenon or complex concept of belief, but simply one aspect of it. Belief is not simply a causal factor in a person which initiates and sustains certain courses of action, any more than it is A mere disposition of a person. Moreover, since a belief may be relevant to current desires or purposes on a particular occasion and yet be causally quiescent, the full detailing of a belief's operation in a particular case can be quite tedious; one must take account of countervailing conditions.[91] Thus this basic outline of the notion, of belief is simple, but it is not simplistic in its outworking. According to Armstrong, then, "A belief is a map-like state in the believer's mind, having a complex structure. . . [and] the belief-state is, and the thought is not, a potential cause or inhibitor of action."[92] "It is of the essence of belief that it can move to action in the service of our purposes and desires."[93] This "causal role which beliefs play in behaviour" is a necessary feature of belief for Armstrong, for he maintains that there is a conceptual or logical connection between belief and action.[94] The mental state; and causal view of belief which has been set forth here is Armstrong's explanation of Ramsey's statement that a belief is a map by which we steer. When S believes that p, he is in a particular mental state which, under suitable circumstances, causes him to act in certain ways.

This conception of belief is briefly recounted by Armstrong in a slightly different fashion at two points. He offers this additional way of thinking about belief:

---

90   Ibid., p. 71.
91   Ibid., pp. 71-72, 74.
92   Ibid., p. 72.
93   Ibid.
94   Ibid., p. 74.

As we may put the point, beliefs are, and thoughts are not, potential premises in our 'practical syllogisms'. . . . Such a link [between belief and action] is present in the case of all beliefs, and absent in the case of 'mere thoughts.' I suggest as a general characterization of the link: *beliefs are, mere thoughts are not, premises in our practical reasoning.*[95]

Such an account stems, as did the former one, from the suggestions of F. P. Ramsey, who said, "It belongs to the essence of any belief that we deduce from it and act on it in a certain way."[96] H. H. Price has taken the lead of Ramsey here, contending that "when we come to believe a proposition p, we add it to our stock of premises. So long as our belief continues, this proposition is available to us as a premise in our inferences."[97] Price writes that an important manifestation of belief is inferring or drawing inferences from the proposition believed; accordingly, when we believe a proposition our belief tends to extend or spread itself to at least some of the consequences of the proposition.[98] "Indeed, believing a proposition seems to consist at least partly in a tendency to draw inferences from the proposition believed."[99] Price explains the view that "a proposition which we believe is one of those propositions which we are disposed to take account of in our practical deliberations" in the following manner:

When we say that a person believes a proposition p we mean (1) that p is a member of his stock of premises (2) he is disposed to use it as a premise in his practical reasoning or practical inferences, inferences whose conclusions, if put into words, would be of the form 'let me therefore do x' as opposed to 'therefore q is true.' It may well be that he cannot ever draw a practical conclusion from any one of these premises alone. He may always have to use two of them, or more, in combination, and one of them may only have been added to his stock of premises a moment ago (e.g., 'there is ice on the road'). But-still any one of the propositions in his stock of premises has the status of a potential premise in some possible piece of practical reasoning . . . .[100]

---

95   Ibid., pp. 72, 74.
96   Ramsey, "General Propositions and Causality," p. 251.
97   Price, *Belief*, p. 254.
98   Ibid., pp. 290-291, 293.
99   Ibid., p. 98.
100 Ibid., pp. 254-255.

Price offers a couple of illustrations of this conception of belief.

> If someone claimed to believe that to-day is early closing day and yet set out on a shopping expedition this afternoon, we should doubt whether he did really believe what he claims to believe. For if he really does believe that it is early closing day he must surely be capable of drawing the very simple inference 'The shops are shut this afternoon.'[101]

This afternoon you intend to go to a village ten miles away, and you believe that the bus will be crowded. Then you will do your best to arrive early at the bus station. Or you will make the journey on a bicycle. Or you will get a lift from a passing motorist. Or you will start your journey very much earlier and walk all the way. In doing any of these actions you are using the proposition 'the bus will be crowded' as a premise in a piece of practical reasoning or inference-guided conduct. You are acting 'in the light of' this proposition or acting 'upon'. it. You are not just making bodily movements of one sort or another. You are acting thoughtfully or intelligently.[102]

Price does not mean to restrict the inference-drawing characteristic of belief to practical affairs; it also encompasses theoretical reasoning on the basis of believed premises.

> We need beliefs because we need guidance not only in our actions but in our thoughts also. We are interested in the question, 'What am I to think?' as well as in the question, 'What am I to do?' To put it another way, when we believe a proposition p we are interested in the question, 'p, so what?' And 'so what?' does not only mean 'so what am I to do?' but also 'so what am I entitled (or obliged) to believe?'[103]

As I have said already, we need beliefs for the guidance of our actions and our practical decisions. This is another way of saying that we draw practical inferences from the propositions we believe, or use them (when relevant) as premises in our practical reasoning. But we draw theoretical inferences

101 Ibid., p. 98.
102 Ibid., p. 255.
103 Ibid., p. 291.

from them too. If one likes to put it so, we use them for the guidance of our thoughts as well as our actions. . . . In other words, when we believe a proposition p, we do use that proposition as evidence to support other propositions. The inference we draw from the proposition p takes the form 'p, so probably also q. Indeed, this is one of the most important uses we have for our beliefs, once we have got them.[104]

The account of belief offered by Armstrong and the account offered by Price here, both of which stem from suggestions in the writing of Ramsey, can now be synthesized into one general characterization. Armstrong has said that belief is a map-like mental state that is a potential cause of particular action under suitable circumstances. Such a state can exercise its causal influence in our actions--whether they be mental, verbal, or-bodily--both when we are currently reflecting on our state of mind and its consequences, and when we are not. Either way, my thoughts and actions rely upon the mental state corresponding to the proposition believed. If I unquestioningly follow the advice of a friend as a matter of habit, the mental state of believing that he is truthful causes me to react as I do. If I were to think through each step in my reaction, or if someone else were to give a reconstruction or explanation of my reaction to the friend's advice, it would be said that the proposition, "This friend is truthful,' (or something like it) is used as a premise in my mental or practical inferences. In both cases, unreflective and reflective, the proposition which is believed has guided my actions and thoughts, or to put it another way, I have relied upon it in thinking and acting as I now do. As Price has said; to believe that 'This friend is truthful' is to have the proposition available in my stock of premises for use in theoretical and practical inferences. "When we believe a proposition we rely upon it. . . . A proposition is relied upon when it is available to us as a premiss for inferences, whether theoretical or practical.[105] A belief is an action-guiding state of mind. When we rely upon that state of mind unreflectively it can still operate as a cause in our thinking and behavior. When we rely upon it reflectively its direction in our thinking and behavior is viewed as part of our reasoning or implicit reasoning. Either way, to believe p is to be in a state of mind which determines a particular kind of mental, verbal, or bodily behavior under suitable conditions; it is to rely upon the proposition p and to have the corresponding mental state causally active in one's actions (including speech-actions) and reasoning.

The overall characterization of belief which has been progressively advanced in the preceding discussion can now be summarized. Belief is a

---

104 Ibid., p. 98.
105 Ibid., p. 293.

propositional attitude of a positive, cognitive type, constituted by an action-guiding mental state on which one relies in his theoretical and practical inferences. Throughout the remainder of this study of self-deception the conception of belief that will be consistently used is as follows: S believes that p if, and only if, S relies upon p (sometimes, intermittently, or continuously) in his theoretical inferences and/or practical actions and plans. The general explanation of this approach to the notion of belief will be found in the discussions of Ramsey, Armstrong, and Price as rehearsed above. Whenever 'S believes p' appears in the subsequent study, it may be translated as 'S relies upon p in his theoretical inferences and/or practical actions and plans."

To give precision to the conception of belief offered here it would be helpful to indicate briefly how belief is distinguished from such related notions as thought, judgment, assent, hope, certainty, conviction, opinion, supposition, suspicion, etc. However to do so with descriptive accuracy it must first be recognized that the words corresponding to these things are used in a variety of ways and with different horizons of distinction. For instance, with reference to thought, we can ask how a mere thought is distinguished from a belief, or we can ask how 'he thinks that p' differs from 'he believes that p'; the syndrome of 'thought'-words evidences a latitude of uses. Likewise, certain words can be used in one context in a way which defines them over against other words, and yet be used in another context where they are subordinate to the same words as part of a scale including those same words; for instance, 'I suspect that he committed the crime' refers to a kind of low-level belief, while 'I do not believe that he committed the crime, I only suspect it' evidences a linguistic context where suspicion is set in contrast to any kind of belief. It is obviously not possible here to give a thorough rehearsal of the many language-games in which 'belief' and its near relatives function in the English language. In what follows I will simply aim to set forth a few salient distinctions pertaining to the-philosophically interesting interfacings of certain words with 'belief.'

We can begin by separating those words which denote some kind of degree of belief (e.g., 'conviction,' 'opinion') from those which do not fall on such a graduated scale. The latter will be discussed first. Armstrong distinguishes mere thought from belief in this fashion. A mere thought is a datable mental occurrence or a mental state which involves exactly the same ideas organized in the same fashion as a belief state. However, unlike beliefs, mere thoughts may be acts performed by S at will, and putting them into words is their only natural manifestation. Further, beliefs cooperate with desires and purposes to cause actions, whereas mere thoughts do not; the notion of a mere thought does not have a logical connection with action, and

mere thoughts do not function as premises in one's practical reasoning.[106] "It is being argued that beliefs are thoughts <u>plus</u>. The plus is the causal role which beliefs play in behaviour."[107]

With respect to judgment and belief, the former is a mental act of asserting or affirming a proposition, whereas the latter is a mental state of a particular kind described above. Assent--like asserting and affirming--is basically a linguistic notion, the private or internal asserting of a proposition. Judging and assenting are, in virtue of their assimilation to linguistic asserting, activities of which a person is reflectively aware, but the mental state of belief is something which a person need not even be conscious of possessing.[108] Assent and entertainment will be further discussed below.

Hope and belief also need to be compared and contrasted. In one sense, although belief has an essential connection with action, hope <u>need</u> not have such a connection; a woman can hope, for instance, over a period of years that she will get married, and yet not have that wistful longing affect her behavior in the slightest (although, of course, it could have an influence on her behavior). However, in discussing another sense of hope, H. H. Price says that it involves or presupposes some degree of belief. One cannot genuinely hope that X will happen if he does not believe that it is logically and causally possible for X to happen; further, he must have at least a slight degree of confidence that X actually will happen. And yet this confidence must fall short of utter certainty that X will happen, for a degree of incertitude is characteristic of hope; thus we read that "hope that is seen is not hope" (Romans 8:24). "The person who hopes must not be absolutely sure that the event will not happen; but neither must he be absolutely sure that it will."[109] Unlike simple belief, hope involves something <u>further</u> than some intermediate degree of confidence; in a sense hope is actually a complex of belief<u>s</u> (plural). "There is another belief-factor in hope: the valuational belief that it will be a good thing if x happens (whether good in itself, or good as a means). One does not hope that x will happen if one believes it would be a bad thing for x to happen, or even if one believes it would be neither good nor bad."[110] To believe that the Dodgers will win the pennant is to have some degree of confidence that this will take place, whether you live in Los Angeles or San Francisco; but presumably a Dodger fan would hope that they will win the

---

106 Armstrong, *Belief, Truth, and Knowledge*, pp. 36, 71-74.

107 Ibid. p. 72.

108 Cf. Ibid., p. 72.

109 Price, *Belief*, p. 269.

110 Ibid.

pennant Chile a Giant fan would not.

Sometimes we hesitate to use the term 'believe' when we are aware that the normative or expected dependence of our beliefs upon evidence is askew. A couple of examples are presented by Price. ' Imagine that S sees walking ahead of himself on the sidewalk a man who resembles his friend R in build, clothing, and gait; S quickens his pace, goes straight up to this person, and in a friendly gesture throws his arm around him--only to be embarrassed by the discovery that it is a complete stranger. Since S accepted the proposition in an unquestioning manner that this person walking ahead of him was his friend R, since he took it for granted without weighing and choosing among alternatives, and since he was not at all prepared for the shock of being mistaken, some writers would avoid saying that S believed that this person ahead of him was R; instead they would prefer to say that S "was under the impression" that this person was R (e.g., Cook Wilson) or that S "was thinking without question" that this person was R (e.g., H. A. Prichard).[111] These scruples are not entirely necessary or completely commonplace, but they do represent a distinction observed by some English-speakers. On the other hand Price offers an example that does seem to match customary practice: namely, stating 'my impression (estimate) is that p' rather than 'I believe that p,' when one cannot readily marshal the evidence that supports this opinion.[112] This is not clearly a different state of mind from reasonable belief, and yet the reasons for the belief cannot be adduced. For instance, a person may say 'my impression is that he was driving a '54 Buick,' and not be able to specify the grounds on which he bases that conclusion. "There is a sense in which such impressions and estimates are quite often based on abundant evidence, although the speaker cannot state the evidence, either to others or to himself, or at best can only state a very little of it. One may have evidence, though one cannot give it."[113] In such cases it is common to use 'impression' rather than 'belief.'

In addition to thoughts, judgments, assents, hopes, and impressions we find belief to have a host of near relatives which can be organized roughly on a scale of degrees: e.g., certainty, conviction, thinking, opinion, supposing, suspicion. Price speaks of degrees of assent, while Armstrong speaks of degrees of belief. However the degrees to which they refer are perhaps better taken as degrees of some adverbial qualification of belief; belief is a positive propositional attitude constituted by a mental state which is expressed in a

---

111 Ibid., pp. 209-211.
112 Ibid., pp. 217-219.
113 Ibid., p. 218.

large variety of 'symptoms, some of which are subject to degrees of strength. Degrees of vehemence do not demonstrate that there are degrees of argument. Likewise it would not be valid to conclude from the fact that there are degrees of confidence with which beliefs are held that there are degrees of belief itself. Locutions such as "I am convinced that p,' 'I think that p,' 'I suspect that p,' etc., make reference to beliefs which are held with differing degrees of confidence or to differing degrees of tenacity with which beliefs are held.[114] These degrees are determined by the conflict between confidence and doubt within one's mental state of belief:[115] the degree of confidence fluctuates in inverse proportion to the degree of doubt so that "as we may put it, their sum remains 1."[116] "A proposition about which we have some doubt does nevertheless give us some guidance both in thought and in action; . . . we do nevertheless rely on it in some degree," says Price.[117] The degrees of tenacity which separate various forms of belief can be specified by asking how long, or under what conditions, the believer will stick to his answer of 'yes' when asked 'p.'[118] Another convenient way of distinguishing forms of belief like conviction and opinion seeks the settled feeling with which we hold a belief; we do this by specifying the degree of surprise we should feel if p were to be falsified.[119] The lowest degree of confidence is traditionally associated with suspecting; this would be indicated by saying 'I should be very little surprised if not-p.' Middle degrees of confidence in the case of thinking that p or having an opinion that p would be indicated by saying 'I would be a good deal surprised if not-p.' The highest degree of confidence goes with a conviction that p (or being absolutely sure that p); it would be expressed in saying 'I would be overwhelmed with surprise if not-p.'

The adverbial degrees which distinguish different expressions of belief have been indicated in terms of confidence (or the ratio of confidence to reservation felt), tenacity (or sticking with an affirmative answer), and surprise felt with falsification of the believed proposition. In light of his causal theory of belief Armstrong supplies this added characterization:

> It seems reasonable, therefore, to place the difference between

---

114 Cf. ibid., pp. 204, 205.

115 Ibid., pp. 287-288.

116 Armstrong, *Belief, Truth, and Knowledge*, p. 108.

117 Price, *Belief*, p. 153.

118 Cf. O. R. Jones, "Knowing and Guessing--By Examples," Analysis 32, no. 145 n.s. (October 1971):19-23; Colin Radford, "On Sticking to What I Don't Believe to be the Case," *Analysis* 32, no. 149 n.s. (April 1972):170-173.

119 Price, *Belief*, pp. 131, 276.

different degrees of belief in different causal relations to action. Full belief is a map by which we steer, a mere thought is a map by which we do not steer, partial belief should therefore be a map by which we steer . . . in some possible circumstances but not in others. (Are our circumstances such that we can afford to be wrong?). . . . The 'rational betting situation,' therefore, which Ramsey envisages, in which a man puts his money where his mouth is, and does it completely rationally, may therefore. be considered as simply being an, or the, ideal situation in which degree of belief should show up most clearly, unaffected by the inconstancy of other factors which continually obtain in ordinary life. . . . The belief is a central state, and degree of certainty of this belief is a degree of causal efficacy of this state in relation to conduct. Conduct as a hypothetical 'rational betting situation' is a mere effect of this degree of causal efficacy. [120]

Conviction, opinion, thinking, suspecting, etc., are all mental belief states, but they differ in terms of their varying causal efficacy in guiding one's theoretical inferences and/or practical actions and plans. The different forms of belief are characterized, that is, by differing degrees of reliance upon the believed proposition in reasoning and conduct.

With this background we can easily distinguish such things as conviction, opinion, suspicion, etc., from belief; they are specific forms of the more general mental state of belief. When it comes to believing things, people do not have to choose between an inert agnosticism and an unreserved self-commitment; between these extremes they can find the option of believing a proposition with a lesser or greater degree of confidence. They thereby can conduct their intellectual and practical activities in the light of the proposition, though not without some degree of doubt or mental reservation. Surely this is what we find ourselves and our neighbors doing continually.[121] It is in order to facilitate a discriminatory way of speaking of this phenomenon that we use a multiplicity of words like 'conviction,' 'opinion,' 'suspicion,' etc. These words denote a graduated order or scale of confidence with which beliefs are held. "The corresponding mental-state verbs come in a fair variety. The most typical group once more displays the familiar gamut of 'firmness' going from. . .the weak <u>suspect</u>, <u>surmise</u>, <u>imagine</u>, <u>assume</u>, and <u>suppose</u>, through the stronger <u>think</u>, <u>believe</u>, and <u>hold</u>."[122] According to Price the scale runs

---

120 Armstrong, *Belief, Truth, and Knowledge*, p. 109.
121 Cf. Price, *Belief*, pp. 155-156.
122 Zeno Vendler, *Res Cogitans, An Essay in Rational Psychology* (Ithaca: Cornell University

from suspecting (or surmising) at the bottom--traditionally the lowest degree of belief which is such:-an unstable state that one easily slips out of it into suspended judgment[123]--through opinion and thinking in the middle, to complete conviction or absolute assurance at the top.[124]

A final note can be added here by way of clarification. Sometimes the words associated with the scale mentioned above are used in a way which departs from the basic or simple picture given, and yet the departure still-reflects the underlying notion of a scale of degrees of confidence with which beliefs are held. For instance, as evidenced in the Vendler quotation preceding, the word 'believes' (or 'belief') can be used both in a general and particular way; it can denote the general mental state of relying on a proposition, or it can denote within-the scale a particular level of confidence with which one relies upon that proposition. Accordingly we must distinguish between the exact senses of 'belief' in statements like 'his belief is a matter of conviction' and 'he does not simply suspect it, he believes it.' Furthermore, there is a natural tendency to contrast belief (understood either generally as encompassing the overall scale or particularly as a definite level on the scale of confidence) with either or both of the extreme ends of the scale. Thus some writers make absolute assurance or complete conviction a matter of knowledge and certainty, not "merely" belief. At the other end of the scale the word 'suspect' can come to be used to cover situations where there is not even a low level commitment to or reliance upon a proposition, but simply an interest in the possibility that certain evidence might support the truth of that proposition. Out of the infinite range of logical possibilities a particular proposition or group of propositions receives special attention because of hints found in the available evidence. Thus one suspects that Jones is guilty of the crime (in that the evidence brings Jones to attention without as yet exonerating him), even though one is not in the mental state of believing that Jones is guilty (even with-the lowest degree of confidence). This use of 'suspect' is tied to the fact that within the scale mentioned earlier a suspicion is a belief held with the lowest degree of confidence due to slight or no evidence. It is easy to move from here into saying that one suspects that p when, given the evidence, p is as yet only a relevant possibility. When 'suspect' is not being used to denote a form of belief but rather used to stand

---

Press, 1972), p. 29.

123 Price, *Belief*, pp. 98, 132, 151, 268, 276, 286-287.

124 Ibid., pp. 204, 287, 302; on opinion and thinking, pp. 151, 209, 211, 287; on conviction, pp. 151, 285-286, 287 (cf. Armstrong, *Belief, Truth, and Knowledge* on certainty, pp. 108, 139).

in contrast to anything so firm as belief, it may be distinguished from belief in this manner. If one suspects that p, he treats it as a hypothesis which can be tested through repeated efforts to reason or draw inferences in terms of it along with one's definite beliefs; should a coherent pattern eventually emerge (e.g., the detective's use of the hypothetical premise in various reflections on the evidence turns up the criminal in the end), then one's suspicion may become a belief. But in belief, one does not merely treat p as a hypothesis or as a premise in various imaginary lines of inference; one -actually relies upon p in his theoretical and practical inferences--which is to say that he sticks by the conclusions based on p and ventures to act in terms of p.

In summary of the discussion in the present section of this study, the precise conception of belief--supplied primarily by Armstrong and Price--which will be applied consistently throughout the investigation of self-deception is as follows. Belief is at base an action-guiding state of mind; it is a map-like mental state that is a potential cause of particular action (mental, verbal, bodily). Specifically, belief is a continuing, intentional, mental state (made up of ideas which give a determinate character to the state corresponding to the proposition believed) with a stimulus-independent causal capacity to affect or guide one's theoretical and practical behavior, under suitable circumstances, in a wide variety of manifestations. In what follows 'S believes that p' will be understood as true if, and only if, S relies upon p in his theoretical inferences and/or practical actions and plans. Unlike belief, thought fails to be action-guiding, and judgment is a mental act instead of a mental state. Hope goes beyond simple belief or some intermediate degree of confidence and includes the additional factor of a valuational belief pertaining to a proposition. We say that we are under an impression, rather than holding a belief, when the evidence for the proposition involved cannot be readily adduced. We sometimes say that we suspect that p, rather than believing that p, when p is deemed a relevant possibility and treated in a hypothetical fashion in our inferences--rather than being relied upon in our reasoning and conduct. Otherwise suspicion, supposition, surmise, opinion, thinking, conviction, etc. represent degrees of confidence with which generic belief is held (except when 'belief' itself is used to denote a particular level on such a scale); they are distinguished by their varying causal efficacy in guiding one's theoretical and practical inferences. Whether or not this precise conception of belief is finally defensible, it does have an initial plausibility and will be employed hereafter in the study of self-deception; the resolution of the apparent paradox is conditioned upon it.

It will turn out that the grounds for saying that-S is self-deceived will coincide with or include those for saying that S believes that p. If S did not take

p as true or evidenced--that is, if S did not have a positive attitude or mental state such that p was relied upon in his theoretical and practical inferences--then we could not distinguish self-deception from mere ignorance of, or mere dislike for, p. It is just because S unavoidably looks upon the evidence as supporting p and is thereby in the mental state of relying upon p in his inferences (practical and/or theoretical) that his desire to avoid or manipulate that evidence in self-deception is meaningful. S does not wish to have his mind "in-formed,' by the evidence in this fashion any longer; he does not want to believe what he does. He would rather forget or hide the unpleasant truth that has gripped him--that is, to make covert that he relies upon p in his theoretical inferences and/or practical actions and plans. S's negative emotional response to p (which is meaningful because in part it is caused by his belief that p) leads him to try to escape his involuntary or uncontrived way of seeing things (viz., seeing p as true, being in a mental state of relying on p); in short, this negative emotional response to believing that p makes him attempt to escape his belief that p.

2.2.5 The Bases of Belief-Ascription

Something should now be said about the bases on which beliefs are ascribed, either by oneself or by others, according to the present characterization of belief. It should be recognized that such a discussion pertains to discerning one's beliefs and attributing them to him, not to the state of belief itself. A discussion of belief-ascription, then, is beneficial for the detection of self-deception, but it does not directly facilitate an analysis of self-deception itself. To say 'S believes that p' is to tell us something about the person himself, for only facts about the believer are relevant to ascribing beliefs to him. On the present characterization, when S believes that p there is an onset of a certain state within S--a positive attitude toward, or reliance upon, the proposition (perhaps consciously formulated)--such that he is caused to act and/or to reason in accord with p.[125] And so we have said

that S's belief is expressed in a large variety of symptoms, some of which are subject to degrees of strength. The basis for ascribing beliefs to oneself or others--the symptoms of belief--are what the occurrent and dispositional analyses of belief basically provide us. To say that 'S believes p' can be rephrased as 'S relies upon p in his theoretical inferences and/or practical actions and plans.' Consequently, in every genuine case of belief

---

125 Cf. Peale, "Theory of Belief," pp. 121, 134.

there is some way in which the. believer gives expression to his reliance upon a-proposition (i.e., his taking of p as true). The mental state of belief causes discernible effects in S's reasoning and general behavior. The expression of belief may be an explicit assertion of p, either outwardly or inwardly, or the belief that p may be expressed by the way in which S behaves, reasons, gestures, feels, etc. Earlier we noted that the occurrent analysis of belief takes its tack from a consideration of one's own beliefs, whereas the dispositional analysis begins from the perspective of others' beliefs. In terms of ascribing beliefs to a believer, both approaches have something to offer and should both be taken into account. As the dispositional analysis maintains, a belief is ascribed to someone as an explanation of his observed behavior (including his verbal behavior); we attribute a belief to S when this explains his assertions and inferences, or explains his actions and habits, or explains the strength and tone of his sense of conviction.[126] When we speak of "strong or firm belief" and "degrees of belief (or assent)" we are actually referring to some adverbial qualification on the belief, speaking of the degrees of confidence or conviction with which we express the belief.[127] However, external or observable behavioral symptoms of belief are not the only indicators of such. Inner and private belief-symptoms count as well, thereby guarding the uniqueness of the way in which a believer may come to know his own beliefs.[128] Ascribing beliefs to oneself need not in every case amount to treating oneself like another person, for one's knowledge of his own attitudes, emotions, convictions, preferences, etc., is normally much more immediate, non-inferential, and accurate than one's knowledge of another person's feelings, etc. For this reason a presumptive authority accompanies S's own avowals of his beliefs; apart from our evaluation of his behavior, his own verbal behavior (e.g., claiming to believe that p) is strong prima facie grounds for attributing a belief to him. Nevertheless, such grounds for belief-ascription can be defeated by a cautious and relatively thorough observation of his other behavioral indicators; over a period of time his actions can still speak louder than his words. But none of this denies that there are a wide variety of symptoms of belief, some public and some private, all important in their own right: e.g., decisions, bodily behavior, inaction, emotion states (like

---

126 Rorty, "Belief and Self-Deception," p. 391; cf. Peale, "Theory of Belief," pp. 139-140.

127 Griffiths, Review of *Belief*, p. 65; White, Review of *Belief*, p. 22.

128 Price, *Belief*, p. 295; cf. Griffiths, "On Belief," p. 137; Coburn, "Believing Things," pp. 95-103; Peale, "Theory of Belief," p. 140.

confidence', surprise, doubt), drawing further inferences, assenting.[129]

There are definite criteria for the ascribing of beliefs to people, then, and they are supplied by both the mentalist and behaviorist accounts of belief. Although no one of these criteria, or even a set of them, is in itself logically sufficient to indicate a belief in someone, these criteria provide fairly reliable inductive, correlations between beliefs and their symptoms.[130] In our use of them we must simply acknowledge that neither a person's actions nor his utterances, are infallible signs of his beliefs.[131] The various kinds of expressions of belief should be used to supplement each other, whether we are considering our own beliefs or those of another. In this connection we should note that a person can have a belief without assenting to it (without consciously formulating a proposition and asserting it to himself or others), and that what one assents to as his belief is fallible (subject to correction).

2.2.6 Assent and Entertainment

Assenting to a proposition is an important symptom of belief. To assent to p is to perform an illocutionary act of the expositive type; it is to spell out (either inwardly or externally) how one stands in respect to a proposition, considered perhaps as a question (e.g., 'the prices will be raised?'). The silent assertion of p can be likened to the rehearsal of outward verbal behavior. Armstrong writes that "'Assent' or 'affirmation' are linguistic notions and 'inner' assent or affirmation is presumably an imaging or other imagining of 'outward,' that is, genuine assent or affirmation to a proposition which is believed."[132] Assenting to p brings one's belief to the conscious level of experience where he can be said to be aware of p or be entertaining it in mind (and perhaps engaging in introspection). However, such introspectible or explicit assent is not necessary to belief. That there is no special logical or conceptual connection between beliefs and their linguistic expression is argued at some length by Armstrong.[133] In the first place belief is not logically dependent upon the competence of someone to express that belief verbally; we readily attribute beliefs to beings who lack the capacity to speak, and we view them as informed and deceived (which is to come into belief states) by their perceptions. Secondly, even if it should be established that there are

---

129 Price, *Belief*, pp. 267-298.

130 Mayo, "Belief and Constraint," p. 158.

131 Price, *Belief*, pp. 253-266.

132 Armstrong, *Belief, Truth, and Knowledge*, p. 22.

133 Ibid., pp. 24-35.

beliefs which can be given only a linguistic expression, it would not follow from this fact that beings without linguistic competence could not hold such beliefs. Thirdly, there is no logical link between possessing very abstract beliefs and having the corresponding linguistic competence to utter them. Therefore, assent is not necessary to the mental state of belief.

The cognitive and affective aspects of belief can sometimes be separated in a person and even be at odds-with each other (e.g., hoping for what you know cannot be, fearing what you know cannot hurt, failing to feel conviction in the face of strong proof).[134] Accordingly we can easily imagine a situation where most of the affective manifestations of a belief that p occur in S, and yet he does not assent to it, even when the proposition is attended to in mind: He does not notice that his actions, emotions, assumptions, inferences, etc. are such as would be the expected symptoms of someone who accepts p. Yet when we ask him whether he believes p, he answers that he does not; he gives the same answer when silently discoursing within himself. Even though he does not show the assent-symptoms of belief, his condition is quite obviously belief-like; most if not all, of the other symptoms of belief are present. In most respects he is just like a normal believer, and his behavior can hardly be understood without postulating in him a belief that p. It would be an artificial imposition at this point to erect some terminological rule, prohibiting us to say correctly that S "believes" p under such circumstances. Such would only screen off the complexity of human nature and behavior from us. For we can certainly imagine, if we have not actually encountered, people who would protest that they do not hold beliefs about the inferior human dignity of people of other races, and yet who evidence such an attitude in their social behavior nonetheless. As Price observes, "One may hold beliefs, permanent or temporary, without admitting even to oneself that one holds them."[135] Therefore, the criterion of assenting to a sentence when asked can fail us as a way of making belief-ascriptions in the case of unapprehended belief.[136]

The fact that belief can be divorced from explicit assent, then, shows us that there can be beliefs held by a person of which he is not aware--not consciously entertaining in mind or introspecting. Speaking of belief as a state of the mind, Armstrong adds:

There is no reason why this state should be something which the

134 Griffiths, Review of *Belief*, p. 64.
135 Price, *Belief*, p. 310, cf. pp. 300-301; cf. Peale, "Theory of Belief," pp. 66-67, 122-123, 125.
136 Quine and Ullian, *Web of Belief*, p. 5.

believer is conscious of being in. . . . Currently causal activity is compatible with the belief not being a content of consciousness. Many of the beliefs which guide our actions never enter consciousness while the action is being performed, yet the belief must be causally active at that time. Sometimes a confidently held belief turns out to be false, and as a result the action based on it is unsuccessful, yet only with failure do we become conscious that we had bean all along assuming the truth of that belief.[137]

For instance, a person can rely upon a proposition (e.g., 'There is sufficient gas in the car's tank') in his theoretical inferences and/or practical actions and plans, and yet he need not be entertaining that proposition in mind--as when he goes to start the car without explicitly thinking about it having gas in the tank; the proposition might come to mind, though, if the car would not start or the believer gets stranded down the road. Price offers another example when he says:

This spreading of belief from one proposition to another may be experienced or lived through by the believing person. . . . But sometimes we just find ourselves feeling confident of the conclusion, or feeling surprised when it is falsified, though we did not actually experience any process of inferring. The conclusion 'drew itself,' as it were. We did not consciously draw it. For instance, if I believe that one of my colleagues went to New York the day before yesterday, I feel surprised when I meet him in Oxford this evening, though I did not consciously infer that he was unlikely to be back so soon.[138]

The fact that the set of our beliefs is expanded and diminished throughout our waking moments due to the automatic adjustments of sense-experience and casual reflection already indicates that beliefs, however trivial or fleeting, can be adopted without concentrating on the adoption procedure or being aware of its results. Moreover, not everything that a person believes can be simultaneously attended to by him in thought. Therefore, it is reasonable to say that believing does not require an awareness of the belief on the part of the believer. Introspection does not invariably accompany each and every mental state or action of a person; otherwise an infinite regress would be

---

137 Armstrong, *Belief, Truth, and Knowledge*, pp. 9, 21.

138 Price, *Belief*, p. 293.

generated since consciousness is itself a mental state (e.g., one would be aware that he believed, and aware that he was aware of it, etc.).[139] Normally we are unaware of most of our beliefs, unless some occasion or reason arises which calls them to mind (e.g., a question, a challenge). Thus without being cautious or giving it much thought, people are able to make erroneous judgments about their beliefs (as well as desires, etc.). It is a false picture we entertain of an intelligent being if we think of him as incessantly talking to himself internally and always making explicit (or reporting on) his mental states or acts. Any number of philosophers have likewise noted that someone can be aware of or believe that p without being aware of their being aware of (or believing that) p.[140] Indeed, as they observe, at times an agent can have such a deep-seated and feared belief that he cannot describe what he is aware of (believes) without defeating his own purpose of not apprehending the belief to himself without extended self-examination and counsel. Therefore, with respect to assent and awareness in belief, we can concur with Ginsberg that ordinarily when S believes that p, he is able to judge or assert that p, "unless there are some factors which prevent that belief from being manifested ,"[141] whether they be distraction, carelessness, or personal motivation.

2.2.7 The Corrigibility of Avowals and Disavowals

In addition to the fact that belief need not be accompanied with explicit or personal awareness of one's belief, we must also observe that self-ascriptions of belief by way of assent and disavowals of belief are not incorrigible (i.e., there can be overriding reasons to think them false) and therefore not infallible (i.e., such reports can be mistaken). A person can be held to believe something from which he dissents, and can be found not to believe some things to which he assents; to some appreciable extent we can be mistaken about our beliefs. Avowing and disavowing, then, would not be infallible guarantees of belief and unbelief. There are limits on our self-knowledge, even though our own reports about our beliefs (as well as about pains, perceptions, etc.) have a presumptive authority and are granted a high

---

139 A similar regress argument against the view that self-knowledge accompanies all instances of belief (in which case the believer would have to possess an endless hierarchy of beliefs in order to know that he believes even one proposition) is developed by Collins, "Unconscious Belief," p. 677.

140 E.g., de Sousa, "How to Give a Piece of Your Mind," p. 76; Jaakko Hintikka, *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Ithaca: Cornell University Press, 1962), pp. 28, 53-55, 81-82, 118; Ackermann, Belief and Knowledge, pp. 7-8.

141 Ginsberg, Mind and Belief, p. 69.

degree of accuracy. In discussing this subject, Audi writes:

> We have seen that <u>normally</u> first-person, present-tense, occurrent mental state beliefs are direct, far more reliable than the counterpart beliefs about others, excellent evidence for the presence of the states they "report," . . . but they are like our beliefs about others in being fallible, dubitable, corrigible, and testable.[142]

The extent of the corrigibility of judgments about ourselves needs to be appreciated.[143] One can believe falsely about his motives, desires, and personality traits; he can be mistaken about the location and cause of his sensations. Indeed, some philosophers have argued that even the apparently incorrigible statements of someone about his immediate experiences might later be withdrawn for the sake of preserving consistency among his beliefs. With such fallibility about ourselves being noted, we should not hesitate to grant that one can believe falsely about his own cognitive states like perceiving, remembering, or even believing. The possibility of being mistaken about one's beliefs is not mitigated by the degree of confidence or conviction he feels and expresses in relation to his avowals and disavowals of belief. Such feelings have a contingent relation to beliefs; they may not always attend even those beliefs consciously entertained, and conversely their presence in high degree does not render a person's statements about his belief or lack thereof infallible. "We do not, of course, believe everything we say; nor do we wholly believe everything we declare with conviction."[144]

Some writers have been tempted to think, nevertheless, that belief is "self-intimating," so that whenever one believes something he knows that he does, and whenever one does not believe something he knows that he does not. Because our own relation to our words and deeds differs importantly

---

142 Robert Audi, "The Limits of Self-Knowledge," *Canadian Journal of Philosophy* 4, no. 2 (December 1974):266.

143 lnsofar as incorrigibility theories are made irrefutable by some philosophers, they tend to become trivial: see Robert C. Solomon, "Minimal Incorrigibility," *Australasian Journal of Philosophy* 53 (December 1975):254-256. The possibility of misclassification of one's experience is cogently urged against incorrigibility theories by: Brian Ellis, "Avowals are More Corrigible than You Think," *Australasian Journal of Philosophy* 54 (August 1976):116, 122; J. H. Chandler, "Incorrigibility and Classification," *Australasian Journal of Philosophy* 48 (May 1970):101-106; Kathryn Parsons, "Mistaking Sensations," *Philosophical Review* 79 (April 1970): 201-213. Further against incorrigibility see: Earl Winkler, "Incorrigibility: The Standard Contemporary Doctrine," *Personalist* 50 (Spring 1969):179-193; Douglas Greenlee, "Unrestricted Fallibilism" *Transactions of the Charles S. Peirce Society* 7 (Spring 1971):75-92.

144 Rorty, "Belief and Self-Deception," p. 387.

from the relations of others to them,[145] some philosophers have gone so far as to say that it does not make sense to ask whether self-ascription of belief could be mistaken; being in a specially authoritative position, only the person making the avowal (or disavowal) can say with finality what he believes.[146] Griffiths says that in cases where a person actually has consciously before him what he believes, "It is not possible to speak of a man as mistaken about what he believes:"[147] And Casey concludes:

> I have claimed that one is, in an important way, an authority on what one believes just because one so believes. One cannot be mistaken about what one believes. One may know that one has a particular belief without observing one's own behavior, without behaving at all, and without deciding how one would act or what one would assert in a particular situation. One may simply consider the matter in question and determine by introspection whether one assents to it.[148]

Arguments for the incorrigibility of avowals of one's beliefs are not as abundant as those for other forms of first-person reports (e.g., about sensations), but a few recent ones can be considered here. A common tactic is to maintain that no clear examples have been presented where S sincerely but mistakenly avows (disavows) a belief that he has consciously in mind here and now; those which are adduced are explained away or recategorized (e.g., as doubt, misreporting, mistaken belief rather than mistake in believing that one believes).[149] However, unless one is absolutely committed in advance to rejecting the possibility of such cases, examples are not difficult to find in ordinary experience (e.g., the case of racial prejudice to be mentioned below); indeed, we shall eventually see that cases of self-deception are precisely counterexamples to the thesis that one's avowal or disavowal of belief is infallible. Secondly, some defenders of the incorrigibility thesis utilize the unhelpful ploy of responding to counterexamples by ex post facto importing

---

145 Cf. Wittgenstein, *Philosophical Investigations*, Part II, section x.

146 E.g., Mayo, "Belief and Constraint," p. 150.

147 Griffiths, "On Belief," p. 129, cf. p. 132. Peale also holds to the incorrigibility of one's self-ascription of belief, in a restricted context--for him, when beliefs consciously occur to S but have no time to form dispositions or no opportunity for expression: "Theory of Belief," p. 142.

148 Casey, "Knowledge, Belief, and Evidence," pp. 232-233.

149 E.g., ibid., pp. 232-235; Griffiths, "On Belief," pp. 129, 131; John Exdell and James R. Hamilton, "The Incorrigibility of First Person Disavowals," *Personalist* 56, no. 4 (Autumn 1975):389.

new, artificial conditions that must be met. For instance, in response to Robert Audi's article, "The Epistemic Authority of the First Person,"[150] and in defense of the thesis that a person cannot mistakenly believe that he lacks a belief (that is, "sincerely" and mistakenly disavow a belief),.Exdell and Hamilton import a criterion that disqualifies examples where the mentioned belief falls short of full certainty (i.e., where there is any degree of doubt).[151] And Barker defends the incorrigibility thesis here by restricting his attention to rational beliefs; he says that if a person mistakenly believes that he lacks a belief, then he has obviously failed to employ the critical capacities associated with rationality.[152] That may be all well and good, but it-it beside the point. Whether someone deems them rational or not, people can be mistaken in their disavowals of belief.

We can find more elaborate attempts to defend the incorrigibility of avowals or disavowals of belief in Coburn[153] and U. T. Place.[154] Coburn argues that S's (non-observational) assertion that he believes p is a non-inductive and criteriological evidence that what he asserts is true. He contends that the relation between such self-ascriptions of belief and the actual presence of belief must be non-contingent, for if not it would be logically possible for a person always to be mistaken in his avowals of belief--which is "hard to credit" in light of the way in which we use the word 'believe.' Coburn offers two illustrations: if someone correctly used the word 'believe' in connection with other people and yet was always mistaken when ascribing beliefs to himself (as his behavior indicates) then we would not grant that he had mastered the concept of believing; and if we needed to translate some language where an unknown verb was employed in contexts similar to belief-sentences in our language and yet the self-ascription of that verb was always accompanied by what we deem belief-defeating behavior, then we would never translate the verb as 'believe.'

This is a weak argument, dependent as it is on drawing philosophical consequences from what we would allegedly do in counterfactual circumstances;

---

150 Robert Audi, "The Epistemic Authority of the First Person," *Personalist* 56, no. 1 (Winter 1975):5-15.

151 Exdell and Hamilton, "Incorrigibility of First Person Disavowals," pp. 389, 390, 391.

152 John A. Barker, "Audi on Epistemic Disavowals," *The Personalist* 57, no. 4 (Autumn 1976):376; Cf. "Socratic Ignorance Vindicated," Philosophical Studies 28, no. 1 (July 1975):73,74.

153 Coburn, "Believing Things," pp. 99-102.

154 U. T. Place, "The Infallibility of Our Knowledge of Our Own Beliefs," *Analysis* 31, no. 6 (June 1971):197.

I am not sure how Coburn could establish, in such extraordinary situations which run so counter to ordinary experience, just what we would do and say. He actually is stipulating a policy to follow in such cases, but that cannot ground his argument. Moreover, at best what his examples show us is that when the common correlation between self-ascriptions of belief and the presence of belief does not hold (i.e., where there is no correlation at all), we would not (allegedly) use the word 'believe'; we would indeed be at something of a loss. But that merely indicates that we expect a common correlation, not a necessary correlation (which is what Coburn attributes to self-ascriptions and beliefs). Coburn has taken ordinary evidence that S believes something (viz. S's avowal of the belief) and escalated it into indisputable evidence that S believes something, which is no more legitimate than is concluding from one's fever (a normal symptom of infection) that necessarily the person is infected. The fact that avowals of belief are often and expectedly accompanied with the mentioned belief does not demonstrate that such avowals are incorrigible. People can and do sometimes come to realize, on the evidence of their behavior, that their avowal of a belief was mistaken. While the avowal has a presumptive authority, its evidential status can be impugned in the long run by observational findings, and thus the avowal does not guarantee its truth.

U. T. Place's argument is as follows. A person cannot believe a proposition prior to consciously considering it. Moreover, when S considers a proposition, if he asserts it, then he will be unable to resist accepting it as a basis for action (i.e., unable to resist believing it). This latter claim is advanced on the basis of certain empirical, psychological findings regarding suggestibility; when a person hears something asserted, he is naturally inclined to accept it (believe it) unless one of these three considerations prevents him from doing so: the speaker is deemed persistently unreliable in giving information, the speaker is thought to have motives for mendacity, or the asserted proposition is seen as conflicting with the hearer's other beliefs. On the basis of this empirical (inductive) platform Place would attempt to demonstrate the infallibility of one's knowledge of his own beliefs. He argues that in the very act of asserting p one must hear (or be otherwise conscious of) his assertion, and thus he will believe it unless he deems himself unreliable, mendacious, or in conflict with other of his beliefs. /n the latter two cases (mendacity and conflict with previous beliefs) the speaker could not be able to believe the asserted proposition, thus making his assertion of it a lie. In the first case, says Place, to treat yourself as persistently unreliable is not rational; it is the way of madness. Therefore, if someone asserts "I believe that p,' he cannot be mistaken; if the assertion is false, then the speaker must be lying. But I find this line of reasoning less than compelling. There is every reason to be skeptical about

the finality of the empirical claims about suggestibility (e.g., the tendency to believe what one hears except under three narrowly circumscribed conditions obviously has widespread counter-examples) and to wonder about the easy transference of findings about one's response to hearing others to the case of responding to the hearing of oneself. And the thesis-protecting claim that one cannot have a belief prior to consciously considering it (for if one could, then he might unwittingly and mistakenly assert something contrary to it, thereby evidencing the fallibility of his knowledge of his beliefs) is in conflict with observations we have already made in this study. However, we need not dwell on these points. Even granting them there are at least three major difficulties with Place's argument.

First, it counter-intuitively treats assertion as an avenue for coming to belief (thereby guaranteeing that one does in fact believe what he asserts) rather than the expression of a belief; apart from lying (which Place sees as an exception to his thesis of the incorrigibility of belief avowals) why would S assert p unless p were already his belief? Second, in common with others mentioned earlier, Place dismisses one obvious counter-instance to his thesis on the irrelevant ground that it represents what he deems as less than rational behavior (viz., the case where, upon-heating his assertion, S rejects it because he sees himself as quite unreliable, thereby being mistaken is his belief-avowing assertion). As much as we may disapprove of it, this <u>is</u> a case of a corrigible avowal of a belief. Third, even granting Place's theory that assertion is an avenue to belief, we must not overlook the fact that the proposition which is asserted in the context of this discussion is 'I believe that p.' On Place's theory, when S asserts this, he comes to believe it--thus indicating that S <u>believes that he believes</u> that p. But this result is not sufficient to prove Place's original point: that in avowing a belief, S actually believes the proposition mentioned in the assertion of belief. That is, Place must demonstrate that if S asserts that he believes p, then S actually does believe p (for self-ascription of beliefs is infallible). However, granting Place's argument, he has simply demonstrated that S believes that he believes p. It is still quite possible that, while S believes that he believes, he in fact does not.

Before he ends his article, Place takes up a case of phobic behavior which appears to be a clear counter-example to the infallibility of our knowledge of our beliefs. He mentions a man who <u>says</u> that he does not believe the statement 'Cats are dangerous' and yet <u>behaves</u> quite clearly as though he believed them to be dangerous. Here the man's disavowal of a belief seems obviously mistaken. Place's attempt to rescue his thesis will likewise appear to be mistaken. He counters by saying that this man's disposition to act in a phobic manner is an independently existing impulse which does not proceed

from an acceptance of the statement 'Cats are dangerous.' He claims that this is not an example of someone who is mistaken in the assessment of his beliefs, but rather is an example of a disturbance in the rational relationship between one's assertions and behavior, a relationship implied in the concept of belief. His first remark is so question-begging and counter-intuitive that it will require much more defense than he has given the thesis presently under consideration (e.g., can there be such a thing as an emotional impulse which is completely independent of any cognitive assessment of the facts, that is, a belief?). The-second remark shares the drawback of those similar to it which have been criticized already in this study. Place's third and final remark about this example is that, if S's behavior indicates that he does in fact believe that cats are dangerous, then S's disavowing of the statement is either a lie or a mistake. The former is ruled out <u>ex hypothesi</u> because S is considered sincere in his assertion, and the latter is challenged on the basis that S does not deny for a moment that he has the behavioral disposition to act as if cats were dangerous. This last remark is a new element in Place's scenario, added after the fact. However, it can be easily countered by telling our own story with that element left out; such a story would be, I believe, closer to what we really find happening when people act fearful but deny the object of their fear. Nevertheless, even leaving this aspect of the story in the script, the fact is that a person who recognizes his fearful behavior toward cats and still denies believing that cats are dangerous is <u>all the more</u> to be deemed mistaken in his disavowals, for when he denies that the phobic behavior proceeds from a belief that cats are dangerous (and yet, as Place's story has it, he offers no other belief to account for his behavior) we have one more instance of a mistaken self-ascription of belief. Actions will, after all, speak louder than his words.

Therefore, we have found no compelling argument against the claim that one may be mistaken in his avowals or disavowals regarding his own beliefs; what one says about his beliefs or lack thereof is fallible, dubitable, corrigible, and testable. S may have a belief, earnestly avow that he does not have such a belief, and be mistaken in his avowal."[155] "It is not unnatural to say of a man that he believes that he believes although he does not in fact do so."[156] even though such a description is usually given by someone else of him (say, in the third person)--for a confession of this entire situation by the man himself "smacks of self-deception."[157] Scattered throughout Price's major study of belief are observations which support what we are maintaining here,

155 Ginsberg, *Mind and Belief*, p. 4.
156 Hintikka, *Knowledge and Belief*, p. 124.
157 Ibid., p. 125.

and interestingly he also tends to associate these observations with cases of self-deception.

> There is also the rather curious situation in which a man thinks that he believes something, but in a moment of unusual clarity and honesty is obliged to admit that he does not, and even that he ceased to believe it some time ago. . . . I may believe a proposition p without realizing that I believe it. It is possible to be mistaken, and in some sense sincerely mistaken, as to what one's own beliefs are, and also as to the degree of confidence with which one holds them. It-is even possible to believe a proposition p when one thinks that one disbelieves it or that one has an 'open mind' about it. . . . First-person present-tense belief utterances may therefore fail to convey to others what the speaker's beliefs actually are, because he himself does not always know what they are. . . . One may speak as if p were true when one does <u>not</u> believe it. . . . The difficulty is not so much about liars as about self-deceivers: people who say to themselves, as well as to others, 'I believe that p' when their actions show that they do not believe it. Probably there is some degree of self-deceit in most of us. . . . There is such a thing as unconscious self-deceit, pretending, even to oneself, that one's beliefs are different from what they actually are, without being aware that one <u>is</u> pretending. When this happens, other people may be able to discover, by observing one's conduct that one is thus deceiving oneself.[158]

Likewise Ginsberg speaks of S ascribing a belief to himself when we nonetheless, due to overwhelmingly strong evidence, think that he does not have it and will not ascribe it to him; again this is deemed self-deception.[159] Illustrations of this phenomenon are not hard to find. Place has already offered us one in his man with the phobia for cats. A common example in the literature on this subject is that of racial prejudice.[160] It seems obvious that when a person has a prejudice which he does not acknowledge and from which' he would want to dis-associate himself, and yet his social behavior (e.g., undertone in various utterances, facial expression, association patterns, decisions, inferences, hesitations) provides ample counter-evidence, we can

---

158 Price, *Belief*, pp. 33-34, 37, 256-257, 258.
159 Ginsberg, *Mind and Belief*, pp. 50, 68-69, 71-72, 75, 76.
160 E.g., ibid., p. 68; Audi, "Epistemic Authority of First Person," p. 11.

with good reason attribute to him a belief which he firmly disavows; his behavior would otherwise appear quite inexplicable. Indeed, given two factors (at least) about such a person--namely, that he does not wish to view himself as harboring attitudes of racial superiority, and that he nevertheless treats members of other races as less than on a par with him--the explanation that he believes what he disavows seems to be the only appropriate one. It must at least be granted that our natural use of 'belief' expressions countenances such an expression. We do well to be reminded here also that freedom to believe p and freedom to assent to p (inwardly or outwardly) are different matters. When S is aware of p as true (sees the evidence as indicating, say, the slightly subordinate standard of other racial members--e.g., you just cannot deny that they act, dress, and behave somewhat differently and strangely. . .) he may not be free to disbelieve it and avoid instinctively acting upon it. However, S may believe that p and still be quite free to withhold, avoid, or suppress internal and external assent to p.[161] There is nothing which says he <u>must</u> engage in (at least) an inner soliloquy in order to believe it; most people have neither the time nor the complex agility to work through soliloquizing just everything that they see as true (i.e., believe). Accordingly, there is nothing odd or absurd about the possibility that someone may come to infer for himself that he actually has beliefs of which he has remained unaware.[162] For that reason, people may have the best word on what they believe, but they do not logically have the last word.

2.2.8 The Voluntariness of Belief

Up to this point in our characterization of belief we have attended to the sense in which belief is involuntary, a positive attitude toward a proposition which is constrained by the way in which the evidence is seen. We have gone on to note that beliefs are ascribed on the basis of various kinds of behavior (inner and outer), saying that when S believes that p he is in a mental state that is action-guiding. However, the expression taken by that mental state need not be in the form of consciously attending to p, or in the form of assenting to it privately or publicly. Moreover, while avowals and disavowals are good evidence of one's beliefs (or lack thereof), they are not incorrigible

---

161 An argument against avowal preparedness as being essential to thinking something (and *mutatis mutandis* to belief) is given in J. M. Russell, "Self-Deception," pp. 67-75, where he considers the use of thought-attributive expressions to pre-linguistic children; we could add the example of deaf and dumb adults.

162 Contrary to F. A. Siegler, "Unconscious Intentions," *Inquiry* 10 (Fall 1967):256.

evidence of them.

At this point we can turn to a consideration of the senses in which belief maybe looked upon as voluntary and under one's control. The first is evidenced in the first-person present-tense use of 'belief' sentences which is somewhat performatory in function. Sometimes an utterance like II believe that p' is not calculated to give autobiographical information about one's state of mind, but rather is used to express an attitude of self-commitment, to take a stand against inner resistance, to give reassurance to another person, to adopt a program or outlook. Instead of telling others (or himself) what he believes, one can use such an utterance thereby to commit himself to holding certain views or living in terms of them.[163] As an act of self-commitment, the first-person present tense 'belief' utterance is a voluntary determination of what one shall: believe; in appropriate circumstances the expression makes it so. In order to alter how things are about himself one may adopt a program of behavior and cognitive attention which will encourage and reinforce a belief in himself; this program can be initiated by the statement, 'I (hereby) believe that p,' whether it is a secret or public utterance.[164] Having noted that we do not always believe everything we assert with confidence, Rorty goes on to write that declarative 'belief' sentences may express an attempt at self-transformation rather than an assertion of existing belief. She also notes-that-this can be one of the fruitful conditions for self-deception.[165]

A second and perhaps more important sense in which believing can be seen as voluntary or under one's control ties in with the way he comes (or. can come) to be in the mental state of belief. Earlier it was noted that the way in which a person sees the evidence, the seeing of it <u>as</u> this or that, or the taking of it in a particular way, constrains that person's beliefs. Since I see myself as right-handed, I cannot voluntarily and on the spot, genuinely believe that I am left-handed (although I can say that I am and pretend that it is true).

Believing what we desire or will has the air of paradox about it. In the first place, holding a belief on subjective grounds conflicts with

---

163 Price, *Belief*, pp. 30-34; cf. J. M. Russell, -"Self-Deception," pp. 82-90. Russell makes the insightful suggestion that the performatory quality of such utterances explains why it is that, when S mistakenly avows (or disavows) a belief, we still tend to credit him with sincerity.

164 An argument that performatives can be thought to take truth values will be found in Herbert Fingarette, "Performatives," *American Philosophical Quarterly* 4 (1967):1-10.

165 Rorty "Belief and Self-Deception," pp. 390-392.

the cognitive status of belief. Believing something, after all, entails thinking that it is true. As much as we may desire or will things otherwise, we need to wait upon experience to see what the truth of the matter is. Thus the idea that our beliefs can be self-made, rather than the effect of the world working its way on us, strikes a conceptually odd note.[166]

We have all had the kind of experience that led Hume to speak of belief as a feeling or automatic response to-circumstance; few if any people can look out a window and see dark clouds and pouring rain, only then to force themselves to believe that it is sunny and bright. Nobody can believe contrary to the way in which he sees the evidence. However, one can exercise some control over the way in which he sees the evidence; one can direct his attention, give prominence to some matters over others, suppress what he wishes to encounter, re-evaluate the significance of last consideration, etc. If belief is like "seeing-as," then we must also recognize that seeing-as is somewhat subject to one's will.[167] In that sense a person is free to ignore the grounds for a belief, in which case a belief is not compelled after all. While it may be true that S cannot voluntarily choose to believe whatever he wishes, just like that, he may nevertheless freely doubt propositions, suspend judgment about them, voluntarily inhibit the extension of his inferences, etc. Direction of our thoughts is a kind of doing, and by the direction of our attention we can encourage or thwart our propensity to believe things.[168] Thus a person is free to fortify or undermine beliefs he may have by voluntarily concentrating his attention on certain evidence, ignoring other evidence, misconstruing evidence (seeing it as something it is not), etc. In such ways as these he can deliberately cultivate a belief (whether about some objective matter or about himself and his beliefs) which turns out to be contrary to the facts.

Philosophers like Augustine, Newman, Schiller, and William James have discussed the phenomenon, and we are all familiar with it from personal experience. "Everyone knows that we can do these things, however difficult it is to talk sense about them."[169] "That belief is effected by our subjective selves is a bit too common to make these problems seem truly paradoxical."[170] We all know the experience described by Descartes, that of

---

166 Shapiro, "Self-Deception," p. 234.
167 Helm, *Varieties of Belief*, p. 152, cf. pp. 142-143 (where Helm points to this as a voluntary act of personal interpretation).
168 Helm, *Varieties of Belief*, pp. 149, 150, 153; Price, *Belief*, pp. 25, 294.
169 Price, *Belief*, p. 230.
170 Shapiro, "Self-Deception," p. 235.

weighing or deliberating about the options and then "taking the plunge" of will and assenting to one over the other. We ordinarily assume responsibility, and are held responsible, for our beliefs; they are assessed as though we had some control over them (e.g., being evaluated as more or less reasonable, justifiable, or logical, and sometimes classified as moral or immoral).[171] From another perspective the voluntariness of belief is seen when we hear others or ourselves saying "I cannot believe that" (e.g., that one's friend has been disloyal, that one's country is guilty of aggression, that a close relative has been convicted of a heinous crime). And of course the "cannot" here should actually be read "will not"--because one does not want it to be-true Cannot emotionally afford to admit it, thinks it his duty not to, or lacks the intellectual inertia to rise to the occasion.[172] Thus in many ways we recognize the voluntary aspect of believing. This can only be deemed inconsistent with the compulsory character of belief if we strictly separate desire and will into a different category from belief, making the one completely voluntary and the other fully passive. The tendency to do this may arise from recognizing that we can only believe what we think is true, whereas we can desire what we know is not so; the tendency is also bolstered by and metaphorically expressed in the model of faculty psychology in which the mind is divided into intellectual, volitional, and emotional functions such that belief is the effect of experience's influence on the mind, and desire and will are sources of our attempts to influence the world.[173] Thus we come easily to hold that belief is completely involuntary and automatic, when other aspects of our experience, reflection, and speech equally indicate that belief is under voluntary control. Price has been successful in trying to do justice to both sides of this issue. He points out that, while one cannot believe just any proposition by an act of the will here and now, nevertheless one can cultivate a belief by voluntarily directing his attention to those elements of the evidence which will support a desired belief.[174] Bringing both perspectives together, he says, "You cannot help preferring the proposition which your evidence favours, the

---

171 Cf. Mayo "Belief and Constraint," p. 152; Price, *Belief*, p. 26.

172 H. H. Price, "Belief and Will," ed. Stuart Hampshire (New York: Harper and Row, Publishers, 1966), pp. 92-100.

173 Shapiro, "Self-Deception," pp. 236, 238.

174 Price, *Belief*, pp. 222, 223, 230, 240; "Belief and Will," pp. 104-112. Similar treatments of the problem are given by Peale, "Theory of Belief," pp. 36ff., 134-135, who says that it is trying and wanting to believe that are under voluntary control, and by Shapiro, "Self-Deception," Chapter 6,-who does much to compare and integrate the standard views of Descartes and Hume on the subject (as does Price, *Belief*, series I, lecture 10).

evidence mu are at the moment attending to... ."[175] S cannot believe p unless he sees it supported by the evidence, but what he takes to be the evidence will be dependent on how he exercises his attention in matters (e.g., seeking out all the facts, or averting attention from evidence that is adverse to a desired belief, placing a construction on the perceived facts). One is compelled, then, to believe what his conception of the evidence demands. By controlling his viewing of the evidence one can indirectly control his beliefs.

2.2.9 Summary of the Concept of Belief

Therefore, in coming to the end of this characterization of belief, I trust we have done justice to the ordinary concept and its many facets--especially the sense in which belief is compulsory and the sense in which it is voluntary.

In some cases belief does seem to be the natural outcome of circumstance,, and talk of a distinct act of assent is at best highly artificial. Similarly, in some cases, notably those involving deliberation, it does seem as if self-conscious choice is the basis of belief.[176]

Although it is not a formal definition or analysis, the following characterization of belief can facilitate an account of self-deception. Belief is a propositional attitude (not excluding false propositions) of a positive, cognitive, type constituted by a continuing, intentional, action-guiding mental state (made up of ideas which give it a determinate character corresponding to the proposition believed) with a stimulus-independent causal capacity to affect one's theoretical and/or practical behavior (such that one relies upon the propositional attitude in his reasoning and conduct), under suitable circumstances, in a wide variety of manifestations (some of which are subject to degrees of strength). It has been pointed out that belief may be, but is not necessarily, achieved consciously and rationally. It has been important to note in the course of elucidating belief in this fashion that in a sense belief is constrained by the evidence, and ascribed to oneself or others on the basis of various behavioral indicators (not excluding private assent). It has been observed that 'one can believe a proposition without giving conscious attention to it or asserting it (inwardly or outwardly), and avowals or disavowals

---

175 Price, "Belief and Will," p. 106.
176 Shapiro, "Self-Deception," p. 240.

of belief are not incorrigible or infallible. Finally, we have discussed the ways in which a belief is under one's voluntary control--either by self-commitment (seen in a performatory avowal) or indirectly through the way in which one's attention is focused on the evidence.

In the subsequent study, therefore, S will be deemed to have a belief when it is a property of that person that he has a propositional attitude or mental state that operates as a contributing cause in his mental, verbal, or bodily behavior. Such a mental state will be individuated by the proposition it intends and by its causal influence in the person's reasoning and conduct. Belief is a mental state, not a mental act like judgment; unlike mere thought, belief is action-guiding. And unlike hope, a simple belief need not be accompanied by an additional valuational belief regarding the relevant proposition. Belief can be referred to as conviction, opinion, surmise, suspicion, etc. when varying degrees of confidence with which the belief is held are implicitly contrasted. In some contexts suspecting that p and believing that p are distinguished by treating p as a hypothesis (in suspicion) and relying on p in one's reasoning and behavior (in belief); in the one case S gives consideration to the inferences based on p, while in the other S sticks to those inferences and ventures to act in terms of them. When it is said in the following discussion that S believes that p, this will not entail any answer to the question of whether S rationally deliberated over p, gives p conscious attention, or assents to and asserts p. Indeed, because belief states are manifested and ascribed in terms of a variety of behavioral indicators, what S says about his beliefs and disbeliefs will be taken as corrigible. Finally, even though a person's beliefs may be constrained by the way he sees the evidence, his control over his attention to the evidence will render his beliefs indirectly voluntary. With these qualifications in mind, a precise conception of belief can be consistently applied in the analysis of self-deception. In this study 'S believes that p' will be deemed true if, and only if, S relies upon p in his theoretical inferences and/or practical actions and plans.

These points will be crucial in understanding the phenomenon of self-deception. In terms of the present chapter, the voluntariness of belief is particularly relevant. It has been maintained earlier that self-deceived people believe false propositions. We can add now that these false beliefs have been personally induced in the self-deceiver. "It is clear that desire or will plays a role in self-deceit. Without some subjective source, self-deception is no more than a mysterious case of unwarranted belief."[177]

In characterizing belief in the above way I have tried to offer

---

177 Ibid., p. 234.

commonplace observations, which are agreed to by many writers, and which are true to the complexity of the ordinary notion of belief. Where a particularly relevant philosophical challenge can be made to some element of the characterization I have ventured, a defense of it has not always been made. Thus the resolution of the apparent paradox of self-deception that will be set forth below will be conditioned on the acceptability of this conception of belief. With this proffered characterization of belief, an analysis of self-deception can hopefully be set forth which passes the tests of adequacy laid down previously. Although in what remains to be said I will not always offer specific argumentation for every aspect of the analysis being developed, the entire account of self-deception can be judged as to its success in resolving the apparent paradox. If it turns out to be an adequate analysis, the success of the final product will confer its benefit back on those elements of the analysis which were proposed with little supporting considerations of their own.

A brief illustration of how self-deception works, incorporating observations from the characterization of belief given here, can be offered by way of anticipating the final product. Imagine that among S's beliefs there are those which indicate an uncharitable attitude toward his neighbor, like racial prejudice. S may have a strong motive for concealing from himself that he holds such beliefs. If asked, he might sincerely deny having them, even while they are operative in his affections. All of us know what it is to have beliefs about ourselves which we are reluctant to admit; we do not want to be taken as the type of person who has such beliefs or attributes. Some may be so painful or shameful to acknowledge that we steadfastly refuse to become aware of them (to consciously formulate them in mind), and we may twist the necessary evidence so as to be able to assent contrary to them or (mistakenly) disavow them.

2.3 The Deceived Belief Must Be Genuine

We have maintained that deceived people believe false propositions, and we have elaborated a basic characterization of belief. It will turn out on the analysis being developed here that self-deception actually involves <u>two</u> beliefs which are in conflict. This will be defended in chapter 4. What can be observed here, however, is that the conflict that exists within the self-deceiver can be adequately described as a conflict between two beliefs, and need not be portrayed as a conflict between knowledge and belief. That is, rather than saying that the self-deceiver <u>knows</u> one thing and believes contrary to it, it

will be sufficient simply to say that the self-deceiver <u>believes</u> something and yet believes something contrary to it. The <u>contrary</u> belief in either case will be false. However, there is no need to maintain that the other belief to which it is contrary is true and held on good evidence; that is, there is no need to say that it is knowledge (a true belief held on good evidence) to which the false belief is contrary in self-deception. What the self-deceiver <u>takes</u> to be true (i.e., believes) need not actually <u>be</u> true. What is at issue is not whether the self-deceiver holds a false belief in conflict with a true one. It is equally appropriate in self-deception that the conflict be between a false belief and another false belief, for it is the conflict-state that constitutes the condition for self-deception. As long as the self-deceiver actually believes a proposition to be true, it can be objectively false and still serve to set up or generate a conflicting (and similarly false) belief. Our analysis of self-deception need not become complicated, then, with a mixture of knowledge and belief. A person can deceive himself about a belief which he holds whether or not that belief actually has good supporting reasons and turns out to be true or not. Those are extraneous matters here. The important thing is that the self-deceiver believe some proposition and then (falsely) believe something which is incompatible with it. Accordingly the present analysis of self-deception will be made simply in terms of belief.[178]

As noted above, the analysis of self-deception offered here will portray it as a conflict between two beliefs. We can now add that these will be a first-order and a second-order belief. Initially S gains a belief, recognizing that (as he sees it) the evidence supports some proposition (p); he takes p as true and "expresses agreement" (in some way or ways) with it. Aspects of his experience or behavior are properly explained on the basis that he believes p. This is the first-order belief. Nevertheless, S wants to ignore or escape the unpleasant fact that p represents; he wishes to evade his belief that p. By means of ignoring, misconstruing, or somehow adjusting the perceived evidence--that is, by exercising control over his attention--S generates <u>another belief</u>, namely that he does not believe that p. This is the second-order belief. In it the first-order belief is iterated; it is a belief about-his

---

178 At one point Audi maintains that a plausible account of self-deception can be given "only if" we do so in terms of unconscious knowledge ("Epistemic Authority of First Person," p. 12). However, he needs no response since he says on the very next page (p. 13): "We might wish to allow that a self-deceiver need only have a true unconscious belief rather than unconscious knowledge." But he reverses himself again within the same paragraph, going on to say that it is "perhaps necessary" to use the notion of unconscious knowledge in accounting for self-deception. Whatever his actual position turns out to be, he offers no argument for the requirement that knowledge be incorporated into an analysis of self-deception.

beliefs. This second-order belief is the deceived belief, and as said already, it is-a false belief (whether or not the first-order belief is likewise false). As such it cannot amount to knowledge. Note well that S is said to <u>add</u> a belief (a deceived one), not to alter or drop the previous one. He believes that p, yet avoids it by believing that he does not believe that p. He recognizes p as true but will not acknowledge that recognition. We can observe here that, in the service of not believing that he believes p, S can go--and often does--so far as to believe not-p; however, this is not necessary. In self-deception S is not so much interested in creating or inducing another first-order belief to counteract the other one (although he may do so out of the desire to appear rational); his primary motivation is to forget or evade the dreaded belief, to be convinced in the short or long run that he does not believe that first-order belief. Thus we will contend that to his original belief S adds an unreasonable belief about his beliefs. By this deceived belief S is mislead from the truth about himself.

This account is necessarily brief and sketchy, awaiting elaboration in chapter 4. The present description has been offered simply to facilitate clarity in discussing matters which pertain to belief and its function within self-deception, which is the subject matter of this chapter. However, before proceeding further in the discussion, it might be well to take note of the fact that the outline of self-deception offered above is supported by the insights of other authors who comment on the phenomenon in passing. Hamlyn characterizes self-deception as a case where someone believes p but conceals from himself that he does so.[179] Ackermann says that the man who deludes himself has a belief, yet believes that he does not have that belief.[180] Likewise Price comments that the self-deceiver unwittingly pretends to himself and others that his beliefs are different from what they actually are.[181] We can move on, then, with the assurance that the sketchy and formal account of self-deception offered in this section of-the chapter is not so artificial as to lack initial plausibility.

Given these preliminaries, the point which must now be made is that the belief which functions in self-deception is genuine belief. Because the self-deceiver is concerned for the truth and is not simply play-acting, he really believes that p, and really believes that he does not believe p. He will make efforts to sustain that second-order belief as rational (even though he does so with techniques that cannot but impress others as pseudo-rationality), and

---

179 Hamlyn, *Theory of Knowledge*, pp. 84-85.
180 Ackermann, *Belief and Knowledge*, p. 7; cf. pp. 6, 10, 69.
181 Price, *Belief*, pp. 257-258.

there exists strong behavioral evidence that he genuinely holds the first-order belief as well. I have belabored the point that in self-deception S genuinely believes something(s) because it is required by the notion of self-deception, but also because certain efforts can and have been made to undermine it. For instance, with reference to the self deceiver s attitude toward the evidence which is adverse to his desires (and thus to what I will deem S's first-order belief) Shea has contended that S can intentionally distort the evidence for p without ever being driven to believe p (i.e., without ever being aware of its weight and import).[182] In such a case we would conclude that self-deception does not involve a genuine believing of something--which is contrary to the present thesis. However, Shea's claim is difficult to accept. If S is not aware of the weight and import of the evidence, how can his distorting of the evidence seem to be intentional? On the other hand, if S is aware of the evidence's weight and import, how can he not be said to believe it, given the characterization of belief offered earlier? The fact is that S attempts to distort the evidence just because he believes the (unpleasant) truth which it supports. The first-order belief is fully a belief.

So also is the deceived belief (which I have portrayed as a second-order belief), even though various efforts have been made to down-grade the nature of the mistaken belief found in self-deception.

It might be thought that S is merely <u>pretending</u> either to believe what he says, or to be ignorant of the truth as he really perceives it.[183] This would be done, we are told, in order for S to conceal his real belief ("knowledge" according to some accounts) from others, and thereby cause them to refrain from trying to get him to see the truth (or better, the truth as he perceives it). S's actions and claims may give us the appearance that he actually believes what he says (e.g., that he does not believe p), but there would be serious drawbacks, according to the objection, in saying that he does really believe what he avows here. In the first place, nobody could be so irrational as to believe what he must recognize as false. Secondly, S does not act normally with respect to the available evidence; he strains, twists, and distorts it as though trying to convince himself of something. Thus it should be thought that in self-deception S is merely putting on a performance; he must be lying or pretending about his belief. But this is unacceptable. It may be true that the self-deceiver does not altogether act like a normal believer, but on the other hand he does not quite act like a pretender either. He behaves in ways which depend on the truth of what he says he believes; he relies on it, takes

---

182 Shea, "Self-Deception," pp. 40-42.

183 This hypothetical position is broached by Lerner, "Emotions of Self-Deception," pp. 131, 134-136.

it seriously, reasons on the basis of it, etc. To dismiss the self-deceiver as a liar or fake is to leave some of his actions inexplicable; for instance, it would give us no apparent reason or motive for his pretending. In his avowals of the deceived belief and in his behavior relevant to it the self-deceiver gives evidence of seeking the truth and not merely of responding to his avowed belief "as-if-true." He seems to us to believe what he says, and not simply to "make-believe." And thus the category of pretending does not seem applicable.[184] If the self-deceiver is said merely to feign ignorance of the truth in order to conceal his knowledge from others,[185] then we have a case of hypocrisy and not really self-deception. Audi has claimed that the belief manifested in self-deception is an "as if" belief; however, a page later he acknowledges that it could be real belief.[186] However, in a subsequent article he instate that in self-deception it is as if S believed the mistaken and avowed proposition; and since it is not strictly correct to say that S believes the proposition, when he avows it he is actually lying.[187] But then S is deceiver only, and not deceived; so it is not strictly correct to deem the situation described by Audi as one of self-deception.

Others, like Lerner,[188] have reasoned that the belief involved in self-deception is neither real belief nor a pretense of belief, especially in light of confessions of self-deception which run "I believed p, but not really believed it." Contrary to Canfield and McNally,[189] who assume reductionistically that S either really believes or does not believe at all, other alternatives are open. Belief in self-deception is somehow defective, flimsy, manufactured, irrational; simply to call it 'belief' neglects its peculiarity and ambivalency, but to deny that S really believes conflates self-deception with mere lying and pretense. . . .So it is suggested that in self-deception we find not full-fledged, wholehearted belief, but rather half-belief. Because be thinks that certain features of the concept of belief are absent, Siegler also suggests that self-deception involves only half-belief.[190] S only sort of believes what he says; he has "near belief."[191] This suggestion must also be deemed unacceptable. Without full, genuine belief on the part of S we do not have self-deception, but only vacillation, change of mind, lack of confidence, or insincerity. It is

---

184 Cf. Shapiro, "Self-Deception," pp. 33-34.

185 As suggested by Bruce, "Investigation of Self-Deception," pp. 71-72.

186 Audi, "Epistemic Authority of First Person," pp. 12, 13.

187 Audi, "Epistemic Disavowals and Self-Deception," p. 383.

188 Lerner, "Emotions of Self-Deception," pp. 136-141, 158.

189 John Canfield and Patrick McNally, "Paradoxes of Self Deception," *Analysis* 21, no. 6 (June 1961):143.

190 Siegler, "Analysis of Self-Deception," pp. 162-164.

191 Cf. Price, *Belief*, p. 23; cf. p. 312.

not as though S sometimes gives weak assent and sometimes weak dissent; he fully and confidently avows his mistaken belief without hesitation. There is very good evidence that he really believes what he says, just as there is strong evidence that he believes something incompatible with it. We have no lack of the features of belief; it is just that we have too many beliefs fully evident. Siegler claim that features of the concept of belief are absent is unconvincing since the alleged feature which he adduces is that of consistency--which has been faulted previously in this study.

Finally, the genuine character of the belief(s) which plays a part in self-deceiving might be undermined if one looked upon the self-deceiver's declarative statement of belief as no more than a performative utterance attempting to effect a self-transformation by creating a commitment to some project. As Rorty suggests, instead of seeing S's declaration as asserting a belief, we should see it as an attempt to bring about some change in an aspect of himself.[192] Yet because of the indeterminate character of such declarations, S can focus on their general appropriateness and keep the details necessary for their enactment out of focus--thereby permitting S to believe (descriptively, not performatively) contrary to his avowed belief. In self-deception one's self-conception (on the basis of the performatory avowal) does not match up with the facts. This is a helpful insight, but two problems attend the suggestion that the deceived belief is performatory in self-deception. First, it would not apply generally to all recognized cases of self-deception but only to a subsection of the instances; thus we would still need a general analysis of self-deception, and it would just be wrong to say that the deceived belief in such cases is generally performatory in character. Second, as observed by Szabados, if Rorty's suggestion precludes that the self-deceiver actually believes (in the ordinary, descriptive sense) the mistaken proposition about which one is deceived, then it would mistakenly reduce all cases of self-deception to wishful thinking or hoping contrary to fact.[193] The person who deceives himself is not totally indifferent to questions of truth and evidence, as one whose utterances were simply performative in nature; the self-deceiver actually makes himself believe the (false) proposition which is mentioned in his avowal of belief, or else he is not genuinely deceived after all.

Therefore, we find no reason to look upon the belief that is operative or avowed in self-deception as somewhat less than full, ordinary, genuine belief. The self-deceiver really believes what we attribute to him on the basis of his

---

192 Rorty, "Belief and Self-Deception," pp. 387-388, 391-392, 395-396.
193 Szabados, "Rorty on Belief and Self-Deception," pp. 465-467.

behavior and avowals. Rorty has rightly observed in an article on the subject of belief and self-deception that self-deception is not a distinctive or problematic process, for the same mechanisms are at work in it as in ordinary cases of belief (e.g., selective focusing on evidence, indeterminacy in assertion). "We are not, then, strangers to any of the raw materials of self-deception: they are operative in our normal beliefs and avowals."[194] We conclude, therefore, Chat self-deceived people believe (in the normal sense) false propositions, and that self-deception can be adequately analyzed according to a person's beliefs (where belief is understood according to the characterization of it offered in this chapter). However, one last obstacle needs to be overcome before moving directly into a resolution of the apparent paradox of self-deception in terms of belief.

2.4 Fingarette's Belief-Free Model for Self-Deception

As mentioned in the Introduction, there is one suggestion about resolving the paradox self-deception which neither accepts nor rejects the other-deception model of self-deception. This new model says that we can account for the phenomenon much better without incorporating the mental elements of belief or knowledge. Indeed, these latter are seen as the cause of the perplexity or paradox surrounding self-deception.

Herbert Fingarette's fascinating study of self-deception is the only book exclusively devoted to this subject that is presently available.[195] In it we find a challenge to the announced project of construing self-deception in terms of belief. Fingarette does not think that our everyday language offers us anything but "hints and imputations by way of paradox" (5) with which to discuss self-deception. In such everyday language "self-deception is readily and commonly described by use of varied combinations of terms from the 'cognition-perception' family" (34), among which would be the term 'belief.' For this reason ordinary linguistic habits will be philosophically unhelpful in analyzing self-deception. "Paradoxes arise in connection with self-deception when we characterize it primarily in terms of belief and knowledge" (34). Therefore, although he does not propose the total elimination of the 'cognition-perception' family of terms, Fingarette does call for "a fundamental change of emphasis," so that self-deception will now be seen in terms of purposive action; that is, Fingarette proposes to give us

---

194 Rorty, "Belief and Self-Deception," p. 393.
195 Fingarette, *Self-Deception*. Page numbers from this book will be given in parenthesis in the course of the present discussion.

a "fundamentally new, 'volition-action' account of self-deception" (35). Re turns away from the analytical tradition, due to its misguided preoccupation with the question of whether the self-deceiver knows the truth and with its pervasive use of mental language. Re challenges the unquestioned assumption of earlier philosophers that belief plays a role in self-deception (12), trying to show that this is unnecessary by giving a more satisfactory account of self-deception in terms of one's failure to "spell out" his engagements in the world. Fingarette's view that belief is not central to an adequate analysis of self-deception is not argued directly; rather he bases the acceptability of this claim on the success of his own alternative account of self-deception. That is, the thesis that self-deception can, be understood apart from belief is offered indirect support.[196] Accordingly we must ask whether Fingarette's understanding of self-deception accomplishes its goal of avoiding paradox, whether it in fact obviates the need to speak of self-deception in terms of belief, and whether it is true to the phenomenon itself. If not, then there is no reason to assume that we may not proceed to develop an account of self-deception which makes belief a basic component of it.

Fingarette does not think that the problem of holding incompatible beliefs is. really at the heart of the paradox of self-deception: "The deep paradox of self-deception lies not in this at most mildly odd condition, but in the element of knowing, intentional ignorance" (29). The way in which he proposes to explain this willful ignorance of the truth is in terms of a new model of consciousness, "one in which we are doers, active rather than passive" (38). The traditional model of-consciousness has been passive, construed in visual terms (35). In its place Fingarette proposes to substitute the model of an active power to say something. On his active, volitional model of consciousness one becomes explicitly conscious of something through the intentional act of "spelling out" his "engagements in the world." It involves the operating skill of giving a subjective description of how one sees himself and his surroundings and thus is a way of paying attention to things. This skill at spelling out one's engagements in the world--that is, spelling out "what one does or what he undergoes as a human subject" (40)--is best explained in Fingarette's own words:

> To become explicitly conscious of something is to be exercising a
> certain skill. . . . The specific skill I have particularly in mind as a
> model for becoming explicitly conscious of something is the skill
> of saying what we are doing or experiencing. I propose, then, that

---

196 Szabados, "Self-Deception," p. 56; "Rorty on Belief and Self-Deception," p. 465.

> we do not characterize consciousness as a kind of mental mirror,
> but as the exercise of the (learned) skill of 'spelling-out' some
> feature of the world as we are engaged in it. . . .

I have purposefully chosen as the name of this skill a phrase, 'spelling-out,' which could not in the context be taken literally but must be taken in something like its colloquial use. . . .

Colloquially, to spell something out is to make it explicit, to say it in a clearly and fully elaborated way, to make it perfectly apparent. Typical uses which I have in mind are: 'He is so stupid you have to spell everything out for him'; 'He let me know without actually spelling it out': and 'you know perfectly well what I mean--do I have to spell it out for you?'

Applied to 'becoming conscious of something,' the phrase 'spelling-out' may refer, but need not, to the actual and elaborate saying out loud, or writing down, of that which one is conscious of. However, the phrase 'spell-out' is _ intended to suggest strongly an activity which has a close relation and analogy to linguistic activity. Sometimes--but by no means always--the 'model' activity (literally making something explicit in language) is also an instance of the skill (becoming conscious) for which it serves as model. However it is clear that one often becomes explicitly conscious of something, or, to use the phrase which I now propose to use synonymously, one often spells-out something, without any evident utterance, even to oneself, or with only allusive or cryptic ones.

What the exact connection is between spelling-out and perfectly straightforward examples of linguistic activity, I do not know. I think there it always a close relation. The point of my speaking of a 'model' here is that I wish to avoid even attempting a definitive account (38-40).

One's engagement in the world need not and usually is not spelled out by him; for that spelling out to take place there must be some special reason--we must be prodded into it (40-42).

However, there are situations where there is an overriding reason not to spell out one's engagement but to systematically avoid doing so (43). Even when normally appropriate, S will persistently avoid spelling out his engagement, for to do so in such cases would be to acknowledge his personal identity in a way which is destructive of his unity as a person or his self-conception. And here we move into the region of self-deception.

> In general, the person in self-deception is a person of whom it
> is a patent characteristic that even when normally appropriate he
> persistently avoids spelling out some features of his engagement

in the world. . . . The self deceiver is one who is in some way engaged in the world but who disavows the engagement, who will not acknowledge it even to himself as his. That is, self-deception turns upon the personal identity one accepts rather than the belief one has (43, 66-67).

When a person has assessed a situation and committed himself to avoiding the spelling out of his engagement, he must of course avoid becoming explicitly conscious that he is avoiding the spelling of it out. He must avoid his <u>avoiding</u> through reduplication of his efforts, thus calling for a self-covering policy of not spelling things out.

> The original reasons for refusing to spell-out the truth will also serve as reasons against spelling out the prior assessment and commitment not to spell-out the truth. For to spell-out the assessment and the policy adopted would, of course, require spelling-out the engagement at issue, the very engagement the self-deceiver has committed himself <u>not</u> to spell out (48).

So then, self-deception requires the adoption of an avoidance policy wherein one has chosen to stay ignorant of certain engagements in the world; accordingly self-deception is a purposeful act. Fingarette tries to explain why people deceive themselves, pointing out that sometimes our engagements in the world conflict with our conscious image of ourselves. In refusing to avow those engagements the self-deceiver can avoid acknowledging his personal identity in a way destructive of his unity as a person. Self-deception, then, results from an expedient policy of refusing to spell out one's engagements in order to preserve one's particular, achieved identity; it is the individual's way of protecting his self from internal criticisms generated by his own unacceptable engagements (chapter 3). Interestingly Fingarette does not view self-deception as stemming from a lack of personal integrity--self-deception actually tries to guard that integrity and unity of personality--but from a lack of courage to confront the reality of one's situation (140). We cannot hope, moreover, to avoid an inveterate tendency to self-deception unless we work at developing the skills required to articulate the character of ' our individual and social engagements or forms of life.[197]

---

197 Beneficial discussions of Fingarette may be found in: Bruce, "Investigation of Self-Deception," chapter 4; Shapiro, "Self- Deception," pp. 117-118, 224-229; Drengson, "Self-Deception," chapter 5; Oser, "Invitations to Self-Deception," pp. 116-129; Szabados, "Self-Deception," pp. 56-59; Alan R. Drengson, Critique of *Self-Deception*, by Herbert Fingarette, in

This is briefly a summary of Fingarette's basic approach to self-deception. For all of its beneficial aspects (e.g., the notion of a self-covering policy), does it indirectly show that self-deception should be analyzed apart from the notion of belief? It would not appear so; in the final analysis Fingarette's account proves inadequate. In the first place, Fingarette's account does not banish paradox but merely restates it in new terms. A self-deceived mother who persistently expresses her conviction that her son is a good boy (when he is in fact a scoundrel) must be credited with some measure of sincerity if she is not to be made out a mere hypocrite; thus she spells out her engagement on this matter. However, because she is deceiving herself, she does not really believe in the boy's virtue and is apparently resisting the spelling out of her engagement here. Thus the old paradox has reappeared (in Fingarette's new terms): the mother does, but does not, spell out her engagement. The paradox can be seen another way. Fingarette portrays the self-deceiver as able to spell out his engagement, but unwilling to do so. On the other hand, the self-deceiver is said to have made himself unable to spell out his engagement (by means of the imposed policy of avoidance). In that case the paradox is that the self-deceiver is able, but unable, to spell out his engagements in the world. A third version of paradox in Fingarette's treatment of self-deception pertains to consciousness. He says that the self-deceiver maneuvers in order to disavow something about his engagement in the world; that would certainly seem to mean that the person is somehow conscious of that engagement. Yet because this person refuses to acknowledge that engagement and avows contrary to it, Fingarette is committed to the idea that the person is not conscious of that engagement. However, it then seems that this account of self-deception is afflicted with the paradox that a person is both conscious and not conscious of something at the same time. The self-deceiver is both conscious of the truth which he decides to avoid, and unconscious of the truth as avoided. Fingarette says that the self-deceiver has overriding reasons for not becoming conscious of the 1 truth; he has assessed his situation and determined that he has such reasons, thereby being conscious of the truth. Yet he adopts a policy on the basis of these reasons which precisely keeps him from becoming conscious of the truth. Again it appears that Fingarette has made the self-deceiver conscious of something of which he is not conscious. We must conclude, then, that Fingarette's new approach to self-deception has done nothing to eradicate the apparent paradox which must be

The Canadian Journal of Philosophy 3, no. 3 (March 1974):475; Drengson, Review of *Self-Deception*, by Herbert Fingarette, in *Dialogue*, no. 1 (1973):142; Fell, Review of *Self-Deception*; King-Farlow, Review of *Self-Deception*; de Sousa, "Review Discussion: Self-Deception."

confronted.[198] If he rejected the doxastic analysis of self-deception because it led to paradoxes regarding the person's beliefs, then Fingarette has no good reason to prefer his own account of the matter over the doxastic one.

It can be added in passing that Fingarette's critics have also faulted him for not being true to the ordinary phenomenon of self-deception. His account does not cover all the various kinds of cases of self-deception. In the first place he seems to rely upon extreme illustrations that might not even be deemed matters of self-deception (as opposed to delusion, mental disorder, etc.).[199] Second, some cases of self-deception can take just the opposite form from that suggested by Fingarette; instead of refusing to spell out their engagements, some self-deceivers engage in an artificial and misleading overdoing of them (e.g., escaping responsibility through their constant parading of admissions of what they are up to).[200] Here the self-deceiver pursues the very reverse of a failure to spell out his engagement in the world; instead he demonstrates a super-honesty and overly explicit description of his engagement, along with an inappropriate "rationality" and emotional detachment about everything he does. Third, Fingarette's approach to self-deception does not cover all cases because it applies only to those calculated to preserve one's personal identity or integrity, whereas there exist other (less ego-centric) motivations for self-deception.[201] Fourth, it has been argued that Fingarette cannot successfully differentiate self- deception from cases of wishful thinking.[202] What he says is unique to self-deception (intentional cultivation of a belief due to a passionate desire for p to be so) and to wishful thinking (no purposeful inducement of a belief which is known to be false) could be equally said of the contrasting notion in each case. Fifth, we can add the remark that Fingarette's analysis of self-deception would appear to "explain" the troublesome concept at the price of appealing to an even greater obscurity--namely, the view of the self as a community of originally independent forms of engagement (learnt patterns of behavior). Given Fingarette's view of the person as made up of a community of primitive selves (or patterns of engagement), there would be no genuine self-deception, but only the deceiving of one aspect of the

---

198 The preceding paradoxes in Fingarette's proposal are observed by Szabados, "Self-Deception," pp. 58-59; Drengson, "Critical Notice on *Self-Deception*," p. 479; Saunders, "Paradox of Self-Deception," pp. 569, 570; Bruce, "Investigation of Self-Deception," pp. 111-119.

199 Drengson, "Critical Notice on Self-Deception," p. 484.

200 Hamlyn, "Self-Deception," pp. 50-51; Drengson, "Critical Notice on *Self-Deception*," p. 478.

201 Drengson, "Critical Notice on *Self-Deception*," p. 482.

202 Béla Szabados, Wishful Thinking and Self-Deception, *Analysis*33, no. 6 (June l973):202-203.

person by another aspect of the person. Finally, not a few commentators on Fingarette's proposal have indicated that the spelling-out metaphor is itself in need of much greater spelling out.[203] Just how linguistically explicit must it be? Is Fingarette formulating the concept as he goes along or using a familiar explanatory device? Fingarette's account is especially unclear as he discusses "explicit consciousness," for at times he suggests--contrary to the force of the modifying adjective--that all consciousness is explicit consciousness. Consequently, if Fingarette's account of self-deception is not true to the ordinary phenomenon, and if his account is itself afflicted with inner paradoxes, then he has not offered indirect proof of the need to avoid using the notion of belief in the analysis of self-deception, his claim that it should be set aside rests squarely upon the positive virtue of his alternative account. But those virtues have now been found wanting.

The most conclusive rebuttal of Fingarette's claim that self-deception should be analyzed in volition-action terms and not in terms of belief is that the former demands the latter, and thus in the long run we have no choice but to utilize the notion of belief in our analysis. The person in self-deception, says Fingarette, avoids spelling out his engagement, and the term 'engagement' characterizes those things which someone does or undergoes as a human subject. But believing is surely something that we either do or undergo as human subjects, and therefore believing something must count as an engagement in the world within Fingarette's system of thought. Thus it turns out that his account of self-deception does not exclude belief after all--in which case his account can hardly stand in the way of utilizing the notion of belief in an alternative analysis of self-deception. Fingarette wanted to shift discussion away from 'belief' and 'knowledge' in the analysis of self-deception and dwell rather on consciousness and personal identity, but it could hardly be expected that he would be able to give an adequate analysis of these latter two things independently of cognition terms (e.g., beliefs are obviously involved in one's self-identity). Furthermore, Fingarette needs to have some way of determining that a person's disavowal of a belief is a manifestation of self-deception and not simply mendacity. He attempts to do this in terms of one's sincerity in avowals or disavowals of an engagement in the world. Yet it is difficult to see how he could plausibly and clearly talk about sincerity (or insincerity) in a manner which is completely independent of belief.[204] Therefore, we can concur with John Turk Saunders when he

---

203 King-Farlow, "Review of Self-Deception," p. 78; de Sousa, "Review Discussion: Self-Deception," p. 313; Drengson, "Critical Notice on *Self-Deception*," p. 478.

204 The need and inescapability for belief to be incorporated into Fingarette's account of self-deception is indicated in: Szabados, "Self-Deception," p. 58; Drengson, "Critical Notice

says:

> The "volition-action family" of terms is so laden with notions having
> to do with belief, perception, consciousness, and knowledge that a
> nonparadoxical account of self-deception cannot be advanced in
> terms of the latter family without also being advanced in terms of
> the former family. The two families crucially overlap.[205]

We are led to conclude by the previous considerations that Fingarette has
not resolved the paradox of self-deception that has troubled philosophers;
he simply ignores it. And even if his action-language analysis proved helpful
in seeing through some moral paradox in self-deception, we would still need
a mental language analysis to resolve the original epistemological paradox.[206]
He has neither avoided belief nor paradox in the long run. Fingarette's study
of self-deception, therefore, presents us with no good reason to attempt an
analysis of that phenomenon independently of the notion of belief.

---

on *Self-Deception*,' pp. 478, 481.

205 Saunders, "Paradox of Self-Deception,' p. 570. The same point is made by King-
Farlow, "Review of *Self-Deception*," pp. 77-78.

206 Bruce, "Investigation of Self-Deception," p. 166; cf. Rorty, "Self-Deception," p.
391.

Chapter Three
Self-Deception and Other-Deception

3.1 An Answer to Arguments Against the Other-Deception Model

The preceding chapter elucidated and defended the claim that self-deception minimally involves the believing of a false proposition. The next chapter will maintain that incompatible--and hence two--beliefs are operative in self-deception, and that claim will be motivated by the desire to model self-deception on the pattern of other-deception (or reflexive deception on interpersonal deception). Before getting to that, however, we must stop to answer the question of whether self-deception <u>should</u> be modeled on other deception. Some philosophers have been willing to do so (e.g., Demos, Ring-Farlow, Penelhum, Hamlyn), whereas others have not (e.g., Canfield, Paluch, Mounce, Daniels, Bruce). Those in the latter group usually promote non-standard uses of epistemic vocabulary in analyzing self-deception, which accounts for similarities between them and some advocates of modeling self-deception on other-deception.

The view which seems most natural when one first begins to reflect on self-deception is that deceiving oneself is exactly like deceiving someone else, except that in self-deception the deceiver and deceived are one and the same person. We tend to begin thinking in this way, I suppose, because we speak of "deception" in each case. And there do seem to be common elements: e.g., a false belief is inflicted on someone, the one who inflicts it attempts to conceal the truth. Yet some have argued that our compulsion to model self-deception on other-deception is to be resisted, for it leads into error and misconception which prevent resolving the paradox of self-deception.

One argument of this sort maintains that the reflexive use of normally interpersonal verbs changes their ordinary sense and thus creates paradox when the reflexive use of the verb is taken in its normal sense. For instance, a comic can amuse others by telling them his jokes, but he cannot amuse

himself by telling-himself his jokes; therefore, amusing yourself should not be thought of as the same sort of thing as amusing others. Likewise, deceiving yourself is not the same thing as deceiving another. The reflexive use of certain verbs is not typical of their interpersonal use (e.g., 'teach', 'remind', 'hide from,' 'invite,' 'surprise,' 'surpass'); literally doing these things to oneself is impossible. So the use of them in reflexive phrases (e.g., 'S hid it from himself') must have a special logic all of its own. Indeed, without there being two persons involved, the use of certain inherently other-regarding verbs has no clear meaning at all. In self-deception S is not to be regarded as deceiver and deceived, for in actual fact he is neither (just as in playing squash with himself, S is neither winner nor loser).[1] If we liken self-deception to making oneself do something (i.e., self-command) it will be clear that the interpersonal and the reflexive use of the verb are not parallel.[2] However, this kind of objection is not telling against modeling self-deception on other-deception. It should be noted that in the case of counterexamples (e.g., amusing others, amusing oneself) we are not made to think that one cannot do X reflexively; we are not prone to eliminate the reflexive phrase completely, reducing the action to some other description altogether. Moreover, it is possible with some imagination to think of literal reflexive counterparts to these verbs—as long as it is kept in mind that S does not do them in complete solitude, but rather in special circumstances which enable his doing X to himself. For instance, while S cannot hide a present from himself in -a barren room and on the spur of the moment, he can end up hiding a present from his wife at home so thoroughly that he actually hides it from himself as well. Furthermore, in answer to this objection we need to observe that the reflexive use of a verb does not always, change the normal sense of it (e.g., 'speak of,' 'love,' 'kill,' etc.). Therefore, the question remains open whether 'deceive' is more like those verbs whose reflexive use is extraordinary or more like those verbs whose reflexive use is rather ordinary.

The second kind of argument against modeling self-deception on other-deception maintains that doing so unavoidably leads to paradox (Sartre, Canfield, Bruce, Shea). If S tries to induce in himself a belief which he knows is false (similar to what he would do if deceiving R), he can never be successful; nobody can believe or be conscious of the truth and falsity of p at the same time, for such incompatible beliefs are impossible.[3] In self-deception- -unlike other-deception--the deceiver cannot form a conscious intention to induce a false belief at the very moment that the deception succeeds (or-is

---

1   Champlin, "Self-Deception: Reflexive Dilemma," pp. 281-286, 294-297.

2   Canfield and Gustavson, "Self-Deception," pp. 32-36.

3   Ibid.

maintained). Therefore, the procedure followed in self-deception must be different than that followed in other-deception.[4] The inconsistency involved in thinking of self-deception along the lines of other-deception is eradicable; it cannot be eliminated until one of the inconsistent elements is eliminated.[5] This line of argument against modeling self-deception on other-deception is not convincing. In the first place, there are many kinds or varieties of other-deception; it is not a uniform procedure, as the argument seems to assume. Some forms of other-deception could well be parallel to some forms of self-deception, while other forms of both are not alike. Then again, even when a particular form of other-deception seems to create inconsistency when applied to self-deception, the resolution need not come by elimination; consistency could also be gained by further qualification of the elements involved or of the procedure followed. Moreover, others have contended that there very well could be cases using consciously chosen lies to deceive oneself,[6] although the illustration offered will later be challenged in this study. Contrary to the present objection, in some cases of self-deception the potentially undeceiving information is readily available to the self-deceiver (e.g., Sartre's case of "bad faith," wherein a young lady allows her hand to remain in that of her date since she does not want to stifle his advances, even though she does not admit her own desires to herself). Finally, it has been claimed that an analysis of self-deception which utilizes the component of incompatible beliefs (just as they are found in cases of other-deception) cannot avoid the paradox of S being simultaneously conscious of the truth and falsity of some proposition; this will be countered in chapter 6 of this study.

The final kind of argument against modeling self-deception on other-deception which is found in the literature maintains that there are elements of difference between the two. For instance, the difference in beliefs between deceiver and deceived which appears in a case of other-deception has a different cause than in the difference manifest in a case of self-deception. In other-deception the difference in belief traces to the difference for S and R in the warrant-giving power of S's say-so, whereas in self-deception the difference in belief traces to an essential role played by emotion, so that evidence is not properly treated.[7] However, arguments such as these, even when they are descriptively accurate, only show at best that self-deception and other-deception are not completely alike in every respect. There are

---

4   Shea, "Self-Deceptions" pp. 23-24.

5   5Bruce, "Investigation of Self-Deception," pp. 18-20.

6   King-Farlow, "Self-Deceivers and Sartrian Seducers," pp. 131-133.

7   Daniels, "Self-Deception and Interpersonal Deception," pp. 284-249.

specific differences between self-deception and other-deception. But this can be granted without doing any damage to other-deception as a <u>model</u> for self-deception; disanalagous features do not disqualify a model, but are rather expected in it. Unless this line of argument aims to show that self-deception and other-deception have <u>no</u> important features in common, it can be overlooked. Later we will observe the common elements of self-deception and other-deception which justify using the same word 'deception' in each case.

3.2 Inadequacy of Analyses Which Deny the Parallel

An indirect way to rebut the charge that self-deception should not be modeled on other-deception would be to show that analyses which avoid the parallel are not adequate. Three sorts of such analyses can be found: (1) those which claim that there is no such thing as self-deception, (2) those which propose that epistemic vocabulary is given a special sense in self-deception, and most importantly (3) those which maintain that self-deception does not involve incompatible beliefs but only a peculiar form of a single belief on its own.

A. E. Murphy reasons that, as a form of deception, self-deception would need to be intentional; however, since it is thought that in self-deception S is ignorant of what he is doing, there must not be any genuine deception involved. Moreover, if the self-deceiver were to be aware of what he is doing, then he would be defying the good; yet an incoherent view of human action is involved in the thought that a man could knowingly do what he deems not to be good. Accordingly self-deception cannot be a form of deliberate wrongdoing. Therefore, Murphy denies that self-deception exists, for it would involve intentional ignorance, which is self-contradictory, and deliberate wrongdoing, which is contrary to human nature.[8] His first reason overlooks the possibility (to be explored later) that S intentionally makes himself unaware of the truth, thereby progressing from awareness to unawareness, but still ending up in a conflict state of incompatible beliefs. His second reason rests on a questionable view of human nature, as illustrations of the "forbidden fruit syndrome" indicate (e.g., if one wishes to keep children from throwing stones at his plate-glass window, the worst thing be can do is to post a large sign prohibiting it).

Paluch argues that no strict cases of self-deception exist because there is

---

8   Murphy, *Theory of Practical Reason*; cf. Jordan, "On Comprehending Free Will," pp. 186-193.

no analogue between self-deception and other-deception. According to him, there is an asymmetry between first and third person forms of the expression 'knows p, but believes not-p,' and this asymmetry forces us to introduce non-standard uses of epistemic words like 'know' into the analysis of self-deception. For instance, he says, such a non-standard use is found in the weak models of self-deception by Demos and Freud; for them the analogy between self-deception and other-deception reduces to S not actualizing his capacity to undeceive his victim. But that analogy breaks down: S is aware of this capacity of his in other-deception, but unaware of it in self-deception. Furthermore, the notion of unconscious knowledge which is used in these weak models is opaque. Paluch goes on to say that even weaker models of self-deception, wherein S is held responsible because he ought to have known better than to believe as he does, are not parallel to cases of other-deception either. The responsibility for the false belief is not the same in both cases, for S is not aware of his motives and has no power of self-analysis in self-deception, but he does so in other-deception. Thus Paluch concludes that there are no cases where S strictly "deceives" himself.[9] However, his line of reasoning rests on a number of erroneous assumptions: e.g., that if S cannot declare his incompatible beliefs then he cannot be described as having them; that confessed awareness of incompatible beliefs is incompatible with the conflict between them; that all deception involves mendacity; that belief must be exclusive in nature automatically ruling out contradictory beliefs). These premises in his argumentation are discussed and soundly defeated by David Pugmire.[10] Awareness of a conflict between what one knows and what one believes would not rule out self-deception unless it ipso facto eliminated the conflict; however, if someone were to declare such a conflict in the present tense, it would be absurd to think that it marked the revocation of the conflict-engendering belief. Moreover, we can grant that a self-deceiver does have a disposition to confess this conflict within him, but without saying that this disposition need coincide with the period during which the description of it is true; that is, confessions of self-deception typically come in the past tense. The fact is that it makes perfect sense to speak of deception by accident or inadventure (e.g., "the bright sunshine outside deceived me into thinking that the winter temperature had risen"; "your vigorous gestures misled me into thinking that you were angry with the desk clerk"). The fact that people can hold contradictory beliefs has not only been discussed previously in this study, but it is an obvious fact of human experience (which explains the existence

---

9   Paluch, "Self-Deception," pp. 263-278.
10   Pugmire, "'Strong' Self-Deception," pp. 339-343.

of logic courses, psychological counselors, etc.). It is not an unexceptionable necessary truth that a proposition can be believed only at the expense of all its rivals.

That there are analogies between self-deception and other-deception--contrary to the claims of the above critics--will be shown at the end of this chapter. Therefore, no successful case has been made that self-deception does not strictly occur. Indeed, as mentioned in the Introduction, any analysis which concludes that the well-known phenomenon of self-deception is only an appearance is much more likely–to be suspected of error and confusion than is the occurrence of self-deception itself. Denying the genuineness of self-deception dismisses the problem in an unsatisfactory way; no adequate substitute notion has been offered to cover the vent or state now described as 'self-deception.' Moreover, nothing is accomplished by claiming that words like 'knowledge,' 'belief,' or 'ignorance' are not used in their full sense when referred to in analyses of self-deception. Some philosophers have claimed that the knowledge and ignorance of self-deception are like their counterparts in other-deception but distorted. Self-deception gives a false image of both. Thus other-deception and self-deception are not so distinct that in self-deception S fails to be a deceiver, but nor are they so identical that self-deception becomes nonsensical. The epistemic vocabulary, we are told, is obviously used in a different sense when applied to self-deception.[11] Such an approach accomplishes nothing because it simply re-asserts the perplexing nature of self-deception; nothing is resolved or explained in this manner. Moreover, at best these claims show us that while there is a similarity between reflexive and interpersonal deception, we must also pay attention to their differences. This is precisely what I shall note at the end of this chapter.

A third approach to the problem of self-deception which does not aim to model it on other-deception maintains that we should "look and see" how self-deception expressions function or under what conditions they are used; in this way it will be possible to describe the phenomenon independently of the paradoxical features which perplex us. When this is done, certain philosophers conclude that self-deception does not entail the holding of incompatible beliefs, but rather the holding of a single belief under peculiar circumstances.[12] Thus when we say that S is "self-deceived" we are saying that it is <u>as though</u> he had been deceived by someone else but is himself at fault. Siegler says that we should notice how self-deception expressions function. Since self-deception is not parallel to other-deception,

---

11   Mounce, "Self-Deception," pp. 62-69.
12   Cf. Lerner, "Emotions of Self-Deception," p. 149.

either there exists no self-deception, or self-deception is not akin to other-deception (i.e., the common epistemic-words are used differently), or self-deception expressions function differently than other-deception expressions (yet without the meaning of the constituent words being changed between them). Siegler prefers the last alternative, saying that the use of self-deception expressions is to attribute responsibility to S or oneself for unwarranted beliefs; the first-person use of such expressions functions as a self-reprimand or resolve. To say that S is self-deceived is to say that he ought to have known better, and he would have known better if he had properly received the obvious evidence.[13] Canfield and Gustavson advance a non-paradoxical analysis of self-deception along the same lines, claiming that it is parallel to making oneself do something (i.e., self-command) and not to deceiving another person. They claim that the other-deceiver "knows" the truth in a different sense than that of the self-deceiver since the latter lacks conscious misrepresentation. They conclude that self-deception occurs when S believes p under circumstances adverse to it; the evidential conditions are adverse to p, yet S believes it anyway.[14] The analysis offered by Shea is in the same vein. Self-deception does not involve incompatible beliefs, but holding a belief simply because one wants to--because he has a desire to do so, regardless of adverse evidence.[15] Likewise Exdell and Hamilton portray self-deception as a condition of ignorance which is not due to any of the ordinary causes; in self-deception one's normal ability to reach rational conclusions has been impaired or psychologically incapacitated by passion or emotion.[16]

Each of these suggestions has value, but none of them is a sufficient account of self-deception. While it may be necessary to self-deception that S desires to believe p even though evidential conditions are adverse to it (and thus S ought not so to believe), these same conditions characterize related actions and states. Ignorance, obstinance, and wishful thinking are among them. Siegler is difficult to criticize because he later admitted that, while self-deception expressions have anon-descriptive function, this fact does not preclude that such expressions could also be claiming that something is the case. And when Siegler laid out the truth conditions for self-deception expressions he pointed out that the similarities with other-deception are greater than the differences, and he rendered a conflict-state description of

---

13   Siegler, "Self-Deception," pp. 29-42; cf. "Demos on Lying to Oneself," pp. 473-475.

14   Canfield and Gustavson, "Self-Deception," pp. 32-36.

15   Shea, "Self-Deception," passim.

16   Exdell and Hamilton, "Incorrigibility of First Person Disavowals," pp. 392-394.

self-deception which is similar to the one finally to be proposed in this paper.[17] However, these things represent what appears to be a later turn of thought in Siegler's treatment. His earlier treatment of self-deception expressions would have to be faulted for overlooking the fact that some uses of self-deception expressions have no bearing on elucidating the cardinal features of the normal concept of self-deception. Such idiomatic expressions are peripheral to an analysis of the concept because they are figurative, hyperbolic, and secondary to the full paradigmatic sense (similar to "insanity" expressions: e.g., "You would have to be insane to believe that"). Nis analysis would not fit all and only cases of self-deception, for it would not draw a sharp distinction between self-deception and unwarranted belief or ignorance.[18] Against the analysis offered by Canfield and Gustavson we can point out that treating self-deception as a form of self-command is prima facie odd. Moreover, what they say about making another person do something is not fully applicable even to that kind of action (e.g., intention, request, and compliance are not involved in "making" someone drop his bottle of gin). Further, it is not at all clear what 'belief-adverse circumstances' means in their treatment. On various possible interpretations self-deception will end up merged with simple trust, wishing to believe, mere error, ignorance, refusal to believe, disinclination to believe, etc. Furthermore, when it is suggested that only extraordinary means (e.g., drink, drugs, hypnosis) could bring the self-deceiver to admit the truth, it appears that his contrary belief is either beyond his control or not devised with some sort of awareness of the truth which he wishes to avoid. But in that case we no longer have genuine self-deception, but rather delusion, insanity, mental disorder, etc.[19] In a similar fashion Shea's suggestion that self-deception is merely believing p because of a desire to do so should be faulted as too broad an account, not distinguishing it from other forms of unwarranted belief. In the same way Exdell and Hamilton's view of self-deception as reason being blinded by emotion is so broad a characterization that it encompasses wishful thinking as well; their example of self-deception, moreover, indicates that the form of emotion-blinded reasoning they refer to is actually a matter of pig-leaded refusal to believe some obvious truth, and

---

17   Siegler, "Analysis of Self-Deception," pp. 147-164.

18   Cf. Gardiner, "Error, Faith, and Self-Deception," p. 231; Bruce, "Investigation of Self-Deception," pp. 99-101; Shea, "Self-Deception," pp. 160-161.

19   Cf. Gardiner, "Error, Faith, and Self-Deception," pp. 227-231; Frederick A. Siegler, "Self-Deception and Other Deception," *The Journal of Philosophy* 60, no. 24 (November 21, 1963):759; Szabados, "Self-Deception," p. 55; Shea, "Self-Deception," pp. 162-164; Valberg, "Rationality and Self-Deception," pp. 102-104, 118; Drengson, "Self-Deception," chapter 3; Shapiro, "Self-Deception," chapter 2.

not strictly a matter of self-deception in any distinct sense.[20]

Finally we can mention Bruce's attempt to analyze self-deception without its entailing a conflict state or incompatible beliefs, thereby denying the parallel between it and other-deception. He maintains that the self-deceiver knows p, but tries to conceal this knowledge by feigning ignorance; S pretends not to know p in order to avoid acknowledging it himself.[21] On the one hand this approach fails to distinguish self-deception from related states like pretending, and on the other it leaves us wondering how S could avoid acknowledging his knowledge of p without believing that he does not know p. In the first case we do not have self-deception but lying; in the second case we have self-deception, but not without admitting incompatible beliefs (or at least two beliefs) as inherent to it.

Therefore, not only have we not found a telling argument against modeling self-deception on other-deception, but we also find that attempted analyses which deny the parallel are inadequate for specifically describing self-deception. Thus we may safely proceed on the natural assumption that self-deception is in some important respect(s) analogous to other-deception.

3.3 Self-Deception Is Not Literally Other-Deception

In response to the denial that self-deception` should be modeled on other-deception we need not go so far as to make reflexive deception lust exactly .like cases of inter-personal deception. This is done when the self-deceiver is actually treated as two separate personalities: an individual with an inner duality, or an individual at two different times. Some analyses of self-deception insist that we not construe the person's mind in a unitary way. For instance, Rorty says that there is no paradox in self-deception because there is no united self to be deceiving itself; individuals are a multitude of self-conceptions.[22] Likewise, King-Farlow says that we should not insist on construing a person's mind in a unitary way; rather the mind should be metaphorically viewed as a large, loose sort of committee with an irregularly rotating chairmanship.[23] But the classic expression of this type of response to self-deception is found of course in Freud, who posited a deceiver within the deceived person; by his doctrine of the censor or ego Freud effectively

---

20   Audi, "Epistemic Disavowals and Self-Deception," pp. 381-382.

21   Bruce, "Investigation of Self-Deception," pp. 146, 151-152.

22   Rorty, "Belief and Self-Deception," pp. 404-406.

23   King-Farlow, "Self-Deceivers and Sartrian Seducers," pp. 134136. Other hints at a divided personality explanation of self-deception can be found in: Price, *Belief*, pp. 33-34; Audi, "Epistemic Authority of First Person," p. 12; Hintikka, *Knowledge and Belief*, p. 86.

held that two different personalities were at work within the mind of one individual Such analyses as these allow for self-deception to be a literal case of other-deception because more than one personality is at work.

Freud's account has special problems. His notion of the unconscious, of course, has proven to be philosophically puzzling. His application of it to cases of self-deception is equally confusing. According to him the self-deceiver unconsciously represses dreaded information; thus the self-deception is independent of S's purposes and is somehow imposed on him by forces beyond his control. But we stop to wonder just how such a defensive activity could be completely unconscious. And if it is somehow wholly unintentional, that would suggest quite strangely that there is no agent of the repression-- unless it is an "inner" agent completely beyond the person's control, which is a confusing if not primitive view of personality. And yet, on the other hand, Freud sometimes suggests that self-deception <u>is</u> purposeful repression by the person himself, and not by some personified part of his mind. In whatever way this inconsistency may be resolved, a further problem is generated by Freud's teaching that S represses information so as not to become conscious of it; the implication is that S knows the truth, but knows it unconsciously. But how could such unconscious knowledge ever be detected? The traditional answer is that through psychoanalysis S comes to realize and admit his latent knowledge. However, how can we clearly distinguish between S coming to realize repressed information about himself and S coming to acquire new beliefs about himself under the direction of the psychiatrist? Finally, Freud does not seem to have escaped paradox altogether in his account. Freud says that the ego or censor knows the dreaded truth, while the person does not; yet since the ego is in the person's mind (or a facet of its operations), it certainly seems appropriate to say that the person does know the truth, even though simultaneously he does not- know the truth. In the long run Freud only <u>restates</u> the problem of self-deception at a different level (i.e., it re-emerges at the level of the censor standing between the conscious and unconscious mind) and does not resolve it at all.[24]

As for other suggestions of multiple personality in S which may not share Freud's special problems, they nevertheless depend heavily on metaphors and thereby do little to resolve the paradox. Obviously it is nonsensical to speak of the soul as divided into portions, some of which can literally act as agents, and thus having souls within souls; hence the metaphors. However, such metaphors treat self-deception as other-deception and seem to escape

---

24   Cf. Bruce, "Investigation of Self-Deception," pp. 62-67, 157-160; Gardiner, "Error, Faith, and Self-Deception," pp. 226-227.

the paradoxes only by giving a figurative account of an answer. The answer or resolution itself is still wanting. Moreover, in self-deception whatever it is that knows and deceives is also that which is ignorant and deceived. But this is impossible on accounts which divide the person into parts with deceiver and deceived separately assigned. Since two different personalities are literally involved (i.e., deceiver and deceived are not the same <u>simpliciter</u>), this cannot be a genuine case of <u>self</u>-deception.[25]

Another kind of analysis which treats self-deception as too literal a parallel to other-deception is one which postulates, not a duality within the person, but a duality of <u>times</u> for the person.[26] Self-deception is identical with other-deception, minus the simultaneity of incompatible beliefs in the latter. In self-deception S is both deceiver and deceived, but with a temporal distinction. S resolves to perpetrate a deception upon himself, taking steps to avoid counter-evidence to a preferred belief (or if he encounters it, insuring that he will not take it as such or that he will be able to dwell on contrary interpretations of the evidence as being more plausible); hereby S succeeds in inducing a belief in himself which he formerly would have taken as false. By intentional selectivity regarding the evidence S makes himself forget p and comes to believe not-p. He is successively deceiver and then deceived. This is a literal parallel to other-deception because the deceiver and deceived are not really the same person, but rather S at t1 and S at t2 respectively. S is a different person as deceived than he was as deceiver. There are at least three significant reasons why this account is unacceptable. First, and ironically, is the fact that this attempt t parallel self-deception with other-deception overlooks an important difference between its analysis of self-deception and its analysis of other-deception. In the latter, the belief (of S) and disbelief (of R) <u>need not</u> be at different times, whereas in the former they must <u>necessarily</u> be at different times. Second, at best this analysis covers only a subsection of the varieties of self-deception and is not thus an adequate general account of the phenomenon; there are genuine cases of self-deception where S does not progress from one belief to its opposite, but never entertained the incompatible belief to begin with. Finally, and most importantly, based on this analysis the deceived (S at t) does not realize that his newly acquired belief is erroneous; S would only know that earlier as deceiver (S at t). Therefore, this is not a genuine case of <u>self</u>-deception, but merely a matter of a <u>change of belief</u> (replacing one belief with an incompatible one).[27]

---

25 Cf. Daniels, "Self-Deception and Interpersonal Deception," p. 246.

26 E.g., Szabados, "Rorty on Belief and Self-Deception," p. 471; "Self-Deception," pp. 63-64.

27 Cf. Shea, "Self-Deception," pp. 40-42, 159-160; this view is also quickly dismissed

So then, while we cannot agree with those who insist that self-deception cannot be modeled on other-deception in some fashion, neither can we go to the other extreme of too closely identifying self-deception with other-deception. The former is not literally a case of the latter.[28]

3.4 The Common Ground between Self-Deception and Other-Deception

Although there is no telling argument against modeling self-deception on other-deception, the most convincing reason that we should see a parallel between the two is because it is possible to locate common ground to them. In locating it, we keep in mind that self-deception and other-deception are analogous, not identical. If self-deception were literally a case of other-deception then we would have two persons in some sense, and no genuine self-deception. Thus self-deception cannot be expected to have every trait in common with other-deception, any more than acting and reacting, warning and forewarning, heterosexual and homosexual, teaching others and teaching oneself, need to have each and every characteristic in common in order for each pair to belong to a common genus. To recognize a special act of self-manipulation in self-deception and thereby portray it as exactly like other-deception is as misleading as to ignore a special self-manipulative activity on the part of S in self-deception and claim that it is wholly different from other-deception.[29] The possibilities need not be reduced to rigid adherence to other-deception in every particular or complete abandonment of it as a model for self-deception.[30] There is a sense in which the deception--the perpetrating of a false belief--is analogous between self-deception and other-deception; the two activities or states are linked by a family resemblance.[31] If we can discover the basic character of deception, then we will be in a much better position to decide how a man can, or cannot, really deceive himself.

> How can a man deceive himself? How can a man deceive another? It is easy to understand how a man can deceive another and difficult to understand how a man can deceive himself. But if we can understand the activity of deceiving then we should have no difficulty in understanding how the activity can be performed in

by Demos, Siegler (1968), Paluch, and Mounce on similar grounds.

28   Cf. J. M. Russell, "Self-Deception," pp. 25ff.
29   Shea, "Self-Deception," pp. 2-3, 76.
30   Gardiner, "Error, Faith, and Self-Deception," pp. 242-243.
31   Wilshire, "Self, Body, and Self-Deception," p. 422.

various circumstances. And if there are circumstances which do not permit the performance then we should be able to find out why. We know how a man can bite himself. When a man tries to bite himself while he is in a strait-jacket, we can explain why he cannot succeed. When a man tries to bite the back of his neck, we can explain why he cannot succeed. When a man tries to bite his teeth we can explain why he cannot succeed. How is it with deceiving?[32]

It is helpful at the outset to note that there is more than one concept or pattern for deception. There is every reason to expect this plurality to characterize self-deception just as it does otherdeception.[33] There are many senses in which we might say S "deceived" R. By something S says or does, R is misled into believing contrary to the truth or to S's own beliefs (about the truth); this need not be wittingly done on S's part, and he can mislead in this fashion both by commission and omission of certain (verbal and bodily) actions. For instance, R might be overheard saying, "By going to the window you deceived me into thinking a guest was soon to arrive." Here R indicates that S's action misled R as to the truth (or at least as to S's own beliefs) even though S did not intend for it to do so. Another sense in which we can speak of S "deceiving" R pertains to an <u>attempt</u> on S's part to practice deceit on R -- that is, intending to mislead R (from the truth, or S's beliefs about the truth). For instance, S may speak and act as though a borrowed sports car were really his own, aiming to get R to believe that the car belongs to S. However, even when R is not taken in by him, R can say "S, you are deceiving me by talking in that way." S was acting in a deceitful manner, purposely trying to induce a false belief in R. However, not all forms of deception require the attempt to bring another person to a new, false belief. Imagine that Johnny breaks his mother's prized-vase and then hides all the pieces; when she returns home he distracts her from the vase's previous location, keeps her mind on other subjects, plays as though nothing is different in the house whatsoever. He is practicing deceit on his mother without making her believe some new proposition which happens to be false; she is not thinking anything regarding the vase just now, but simply has (dispositional) belief that the vase is perfectly all right -- a previously held belief which Johnny continues to nourish. So again we see a broad variety in senses or patterns of

---

32 Siegler, "Self-Deception," p. 29.

33 Paluch, "Self-Deception," p. 272; Hamlyn, "Self-Deception," pp. 48, 51; Champlin, "Self-Deception: Reflexive Dilemma," pp. 288289, 296; Drengson, "Self-Deception," p. 91; Shapiro, "Self-Deception," pp. 21-22, 28.

"deception." Finally, we can mention deception in the sense of <u>successfully</u> practicing deceit on another. Here S believes that p, but he intends to get R to believe (or to continue to believe) the false proposition, not-p. For instance, believing that there will be a surprise party for R tomorrow, S pretends that everyone including himself has forgotten R's birthday, and R is led to believe that his friends did not remember his birthday. Other common and numerous examples of these various forms of deception can be readily imagined.

But what do, they all have in common? 'Net the keeping of R from believing the truth; this is too broad a characterization, for according to it S is deceiving R when S fails to call R and tell him that S has broken his shoelaces. Not purposiveness or intent, for S's deception of R can be accidental and something of which S is totally unaware. Not the inducing of a false belief in R, for S can deceive R by acting {and speaking) in a way which keeps R from changing his (now false) belief about something. Not the successful misleading of R, for S can be accused of practicing deceit even when his attempt fails to reach its goal. I would suggest that what is common to all of these forms of deception is this: S does or says something which induces R, if he believes anything at all about p, to believe falsely (i.e. in a way which is out of accord with the actual state of affairs). To put it briefly, S deceives R when he induces him to believe a falsehood. We will find the same basic meaning for 'deception' in cases of self-deception, except here S and R are the same person. S does something which has the effect of making himself believe falsely; he is responsible for his own believing something which is not actually true. He has brought about false believing in himself, and thus is his own deceiver.

It is evident from the above discussion that S could in some sense be his own deceiver even when the deceit was not purposely practiced or intended. And I think that there are some loose uses of the term 'self-deception' which in fact do not involve all the elements of the strongest forms of deception (e.g., intended, and successful, misleading from the truth). It is to such uses that Hamlyn would be referring when he argues that, since other-deception does not always require that R is ignorant of the truth when S attempts to practice deceit on him, self-deception can be exactly like other-deception. His point is that self-deception need not be intentional misleading as to the truth.[34] Such a variety of self-deception has been termed "weak" self-deception. Without exploring all the traits and implications of, or criticisms against, such a weak form of self-deception, this paper will hereafter confine

---

34   Hamlyn, "Self-Deception."

itself to a consideration of <u>strong</u> self-deception.[35] This is, I believe, the central concept in mind when people ordinarily speak of self-deception, and it is the only philosophically interesting one.

We have seen that self-deception could legitimately be modeled on other-deception because there is a common sense for 'deception' in each case. A further parallel between self-deception and other-deception would be the presence of incompatible beliefs. In other-deception the deceiver believes one thing, but the deceived believes something incompatible with it. In self-deception one person (being deceiver <u>and</u> deceived) holds an incompatible pair of beliefs himself. This will be explored further in the next chapter. However, before proceeding to it, a further important parallel between self-deception and other-deception can be observed. In other-deception the deceived's erroneous belief rests on his implicit view of the deceiver's beliefs. This can be explained by illustration. S deceives R into believing that the sports car driven by S is not borrowed but-owned by S himself. In this case S believes p (that the sports car is borrowed), and be influences R to believe not-p (that the sports car is not borrowed). Now even though R may not explicitly think about S's own beliefs, S must necessarily have-given the impression (by his gestures, words, behavior) that he himself believes not-p (that the car is not borrowed). R must misconstrue, even if implicitly or dispositionally, what S's own beliefs are. If R believed that S believes p, then all of S's performance and words would be for nothing they could not convince R of not-p. S's actions, verbal or otherwise, could not be evidence of not-p for R as long as R believed that S believes p. S will have to portray himself as not believing p if he expects R to be influenced to believe not-p by his behavior. This is all to say that in other-deception S (the deceiver) must believe p, and it is necessary that- R (the deceived) believe that S does not believe p. The same is true in self-deception, I will argue. S holds two conflicting beliefs. On the one hand, S believes that p (as the deceiver). Yet on the other hand, S will also believe that S does not believe that p (as the deceived). The self-deceiver believes something, but does not believe that he believes. This latter second-order belief is false, and S has been influenced to come to hold that false belief by his own activity. Be is responsible for getting himself to believe (falsely) that he does not believe something. And in this basic characteristic of self-deception we find a precise replica of what happens in cases of other-deception.

There is a further sense in which other-deception and self-deception

---

35   Cf. Paluch, "Self-Deception," and Pugmire, "'Strong' Self-Deception" for use of these designations.

are alike: namely, in the defensive maneuvers that the deceivers practice. By observing such tactics as keeping discrediting information from becoming the topic of sustained and explicit contemplation and giving it an unnatural interpretation we are often led to identify a deceiver -- whether he be deceiving others or deceiving himself. The reactions of the other-deceiver and the self-deceiver to adverse evidence constitute another rather obvious point of contact between the two.

> Not counting cases where admissions are easily won, it is just this sort of defensive conduct which we most often rely on in making a decision both about suspected self-deceivers and suspected other-deceivers. We look for behavior which seemingly systematically results in relevant information being either kept from or discredited in conversation. With a person whom we suspect of having such purposes, who will not admit it, we may come to be convinced we are right (both for self-deceivers and other-deceivers) upon witnessing certain characteristic sorts of defensive operations, and characteristic kinds of awkwardness which occur when the person is questioned. It may beat this point that self-deceivers and other-deceivers alike first move more clearly into the category of perpetuating falsehoods, when we 'corner them' with our questions. Of course both will avoid occasions which promise such interrogations.[36]

Such defensive tactics are at hole equally in interpersonal and in intrapersonal deception.

Therefore, we conclude that the arguments against modeling self-deception on other-deception are not telling ones, and that the proposed analyses of self-deception which deny the parallel are inadequate. We have seen that self-deception is not a literal case of other-deception, but is simply analogous to it. And finally we found the following four important areas of commonality between self-deception and other-deception: (1) the deceiver's actions influence the deceived to believe a falsehood, (2) a set of incompatible beliefs is involved in deception, (3) the deceived must entertain a. false belief about the deceiver's own beliefs, and (4) when confronted with discrediting evidence the deceiver responds in a characteristically awkward fashion. In terms of such basic traits as these the other-deception model of self-deception has not been inappropriate.

---

36   J. M. Russell, "Self-Deception," pp. 30-31.

Chapter 4
Self-Deception as a Conflict State of Incompatible Beliefs

4.1 Incompatible Beliefs Need to Be Attributed to the Self-Deceiver

Over against philosophers mentioned in the last chapter who have maintained that self-deception involves S in believing only a falsehood, many others have said that the self-deceiver is involved in a mental conflict of a special kind--namely, holding to incompatible beliefs (e.g., Demos, Penelhum, Pugmire, Rorty). According to the latter opinion, self-deception is a conflict state wherein S has an unacknowledged belief (hidden) which allows for a contrary belief to persist (in open declaration) until this first belief comes to be acknowledged. The self-deceiver believes something but wants to suspend or refuse that belief;[1] accordingly he keeps himself from "believing" it in the fullest sense--by not viewing himself as believing in this way, thereby resisting admission of the belief and disavowing it.[2] He believes that p, but then brings himself sincerely (but falsely) to disavow the belief. To be reasonable he should surrender the deceived (but favorable) belief and submit to the natural tendency (or willingness) to assert the original, unwanted belief. The self-deceiver is in a state of mental conflict between his beliefs.[3] There is good reason to suppose that self-deception does involve such a conflict state of incompatible beliefs.

The fact that we say S is "deceiving" himself, along with the model of other-deception, gives initial impetus to viewing the self-deceiver as

---

1   This aspect of the situation is stressed by Shea, "Self-Deception," chapter 6; Daniels, "Self-Deception and Interpersonal Deception," p. 249.

2   This aspect of the situation is stressed by Valberg, "Rationality and Self-Deception," pp. 248ff.; Audi, "Epistemic Authority of First Person," p. 12.

3   A similar characterization (but from the perspective of longing to believe something) is given by Pugmire, " 'Strong' Self-Deception," pp. 345-346.

holding a belief (as deceiver) and something incompatible with it (as deceived). Moreover, from observing people who are customarily said to be in self-deception it <u>appears</u> that S both does and does not believe the same proposition, given the difference between his verbal and affective behavior. Indeed, those treatments which say that the self-deceiver only believes a false proposition (rather than a true one and a false one together) deny a necessary condition for <u>self</u>-deception; a man who believes a false proposition can as easily have been deceived by someone else. If S believes a falsehood, he may be said to be deceived, but that fact does not constitute a case of <u>self</u>-deception.[4] Moreover, the false belief which the self-deceiver comes to hold and maintain is not simply a comforting belief, accepted because of its attractiveness to S when there is no relevant and perceived evidence against that belief. This would simply be wishful thinking or groundless faith.[5] In self-deception it is essential that S's false belief be held against adverse evidence--evidence of which S is not ignorant but has taken as such, thereby believing the opposite of his false belief.[6] If S does not really <u>believe</u> p when he holds the <u>false</u> belief that not-p, then his is not essentially a case of self-deception.

In a case of self-deception S recognizes the evidence in favor of p. As pointed out in chapter 2, he takes p as true, seeing that it is evidenced. If he did not see this, or if he did not realize the significance and import of the evidence for p, then he would merely be ignorant, dull, or naive rather than self-deceived. Of course there are situations where S does not clearly see the import of the evidence, or equally sees the import of conflicting evidence; in such cases S may not believe, or may only half-believe. Our point is simply that when S actually does see p as evidenced, then he can be said to believe it; when he "expresses agreement" with it (in some observable way) we may ascribe the belief to him. In self-deception S recognizes the evidence as supporting p; he takes p to be true. Attributing this belief that p to S is called for to explain his actions, emotions, gestures, habits, awkward treatment of evidence, etc. But not only does the self-deceiver believe p, he will not admit p. He holds the belief that p, but with the tendency to avow that p inhibited—leading him to think that he does not have the belief when he retrospectively examines his behavior.[7] Although the notion is difficult and will not be utilized here, an "unconscious belief" is taken by some writers in

---

4   Cf. Bruce, "Investigation of Self-Deception," p. 6.

5   Lerner, "Emotions of Self-Deception," pp. 156-157.

6   Cf. Penelhum, "Pleasure and Falsity," p. 258; Bruce, "Investigation of Self-Deception," p. 92.

7   Cf. Robert Audi, "The Concept of Believing," *Personalist* 53, no. 1 (Winter, 1972):61.

virtually the same fashion: S is said to have such an unconscious belief when we are compelled to ascribe a belief to him which he would not express or attempt to defend.[8] Instead he confesses or declares something which is incompatible with p altogether, and such declarations of this incompatible belief seem most sincere. Attributing this incompatible belief to S is called for to explain his assertions and sense of conviction in them. However, this incompatible belief is false--which is why we say that S is "deceived." Nevertheless, it is a special kind of false belief. It mistakenly leads S to suspend or deny the assertion of p's truth, and it cannot be assimilated completely to certain standard categories of mistake. As Daniels points out, everybody <u>has</u> a set of beliefs (a body of evidence) which <u>warrants</u> the believing of some propositions and not believing others--whether or not S actually believes and disbelieves in a manner which conforms to what is warranted by his body of accepted evidence. Usually, when S's body of evidence warrants him to believe p, then he does so openly; S will profess what he sees as evidenced, what presents itself to him as true. However, sometimes this does not happen if S does not profess what his perception of the evidence would warrant--if he declares things incompatible with his own body of received evidence--then we naturally ask why this is so, and the answers to this question invariably fall into certain general categories (e.g., S is slow-witted, brainwashed, confused, incoherent). However, when S's false belief is a matter of self-deception, his believing contrary to the evidence cannot be explained on the basis of these standard categories.[9] Self-deception may shade back and forth into various of these categories, and it may be difficult for others to distinguish it overtly from these related conditions, but nevertheless it is a distinct thing. It can seem to us that S does not believe p since he does not admit it and because he in some ways acts like a man who is - ignorant of p's truth. However, we cannot overlook various rationalizing efforts which seem to have the object of escaping' the belief that p--contrary to his disavowals--or else his "concealing" activities would make no sense.[10] Unlike many categories of mistaken belief, the deceived belief that he does not believe p can be rectified by an act of will on S's part; by being honest or exercising thorough self-scrutiny he could come to admit the dreaded belief. It is within his power to come to believe p in a way different from his original believing--that is, he can come to believe it with an avowal of that fact.[11] So then, there appears to be a special kind of false belief involved in self-deception.

8   E.g. Collins, "Unconscious Belief," pp. 671, 673, 679.
9   Daniels, "Self-Deception and Interpersonal Deception," p. 248.
10   Cf. Bruce, "Investigation of Self-Deception," pp. 51-52, 59-60.
11   Valberg, "Rationality and Self-Deception," pp. 101-102, 161-164.

The self-deceiver's false belief is not simply a matter of pretending (as-it-were-ignorance), for then there would not be any genuine <u>belief</u> that is false; self-deception would reduce to the attempt to deceive others. Nor is it a matter of simple error, ignorance, stupidity, or naiveté; S's false belief is more than an indirect mistake, an incidental oversight, or an instance of carelessness. He not only wrongly disbelieves p, he also believes p at the same time—which makes his error stand out as self-deceived. S's false belief is not simple faith or trust, for these latter conditions do not incorporate a simultaneous and incompatible belief along with the false belief. Nor is it pure bias (which, unlike self-deception, does, not require the suppression of a belief in effect) or forgetfulness (which merely brings about a change of mind or belief). The deceived belief is not simply an instance of half-belief; there is no vacillation, no indecision about which of two acknowledged and incompatible beliefs to give up. S is not merely refusing to admit that his professed belief is false or acting as though its falsity made no difference to him. If he believes that his professed belief (which we know to be incompatible with another belief of S's): is false after all, then <u>ex</u> <u>hypothesi</u> S is under no illusion. The self-deceiver, however, does not believe that his professed and cherished belief is false, for then he would simply be acting in an obstinate fashion. Instead, the self-deceiver has a concern for rationality and a respect for the need of evidence supporting his beliefs. Finally, the false belief held in self-deception is not a matter of delusion (which is externally caused, and which need not involve a conflict state of mind) or mental disorder (in which case it would be outside of S's power to undeceive himself). The false belief held by the self-deceiver has a unique character or context; it is a belief held in conflict with another belief which undermines the truthfulness of the first.

In self-deception S believes p, but he believes something incompatible with p as well; he seems to satisfy the conditions for both belief and disbelief that p. In observing his actions (verbal, bodily, etc.) we must settle for a consistent description of inconsistent behavior. Consider an example. We know that S is quite aware of a large body of evidence pointing strongly to his wife's infidelity. However, the grounds which we have for claiming that S believes that his wife has been unfaithful also support the claim that S is deceiving himself about the matter. S consistently avoids or tries to avoid situations in which he is likely to become aware of her infidelity (e.g., coming upon her and her lover alone together). S tries to say and do only what one who believed in his wife's fidelity would do, but this performance is unrealistic or somehow flawed. He gets angry with her over every trifling matter, unlike the way in which he used to treat her. Yet at other times he is overly solicitous toward her in public; his behavior strikes us as exaggerated--an attempt to

cover up the way he really feels. S tries too hard to be convincing, seizing upon even an oblique reference to marital relationships to deliver a protracted speech on the importance of interpersonal trust in marriage; his outburst culminates with an emotional affirmation of his own wife's trustworthiness in sexual matters. S astonishes and embarrasses us when the subject of fidelity comes up in conversation. When a friend visits S at home and asks about the whereabouts of his wife, a curious reaction is evoked in S. He fidgets, looks away, slightly blushes, answers in a halting and slightly strained voice that his wife must be out shopping. When evidence of the infidelity of S's wife comes into the conversation he handles it in a stilted or twisted fashion, all the while satisfying himself with his plausible and reasonable interpretations of it in favor of his wife's fidelity. He is willing to defend her, to act decisively on the assumption of her trustworthiness, to offer a (pseudo-)rational account of all the evidence bearing on the subject. He evidences sincerity in his affirmation of his wife's fidelity, and yet he seems to misconstrue evidence in a way which no intelligent man normally would.

Here there is some reason to say that S does believe what he asserts, that his wife is faithful to him--otherwise there would be no deception toward himself, but only lying toward us. Yet there is also reason to say that S does not believe what he asserts--otherwise there would be nothing for him to deceive anybody (including himself) about. The evidence that S believes p (his wife's fidelity) is not conclusive evidence against the claim that S disbelieves p; yet the evidence that S disbelieves p is not conclusive evidence against the claim that he believes p. However, in neither case are we inclined to feel that the evidence for one or the other is somehow unclear, unnatural, or unconvincing. There is good reason to say that S believes both p and not-p. In particular, S tends to declare his disbelief in that to which—as his other behavior indicates--he sees the evidence obviously pointing. S's pattern of rationalizing the evidence, all with the air that "there is nothing to avoid," is a process indicating the presence of incompatible beliefs. Such behavior becomes incomprehensible on any other description; no other explanation makes sense of the complete pattern. S denies p and is intractable in the face of counter-evidence; his behavior is firm, his tone sincere. He behaves as would a man who is ignorant of the truth of p; he appears oblivious of the force of the evidence and gives no hint of allusion to confessing the truth of p. And yet his continual maneuvering to rationalize matters and hide his emotions makes no sense unless he does in fact believe p. In fact, when S later confesses to his self-deception, he does not uncommonly say things like "I believed p all along, but wanted so badly not to that I disbelieved it." Self-deception involves the conflict state of incompatible beliefs.

In rendering an analysis of self-deception it is important that we do not confuse it with related but distinct notions.[12] The previous example[13] can help us to see the difference between self-deception and some other activities which are commonly confused with it. The husband in the above example is not simply wavering in his belief or insincere in his declarations.[14] The evidence given in his behavior shows movement in conflicting directions; the overall effect of his behavior, assertions, inferences, etc. makes us conclude that the husband is in a conflict state. He does partially satisfy the various criteria for belief and also those for disbelief. There are signs which, in the absence of contrary signs, would add up to belief; and likewise, there are signs which, in the absence of contrary signs, spell disbelief. The contrary signs would be sufficient by themselves to show one belief or the other, but together they require us to give a coherent description of inconsistent behavior.[15] The inconsistency is present at one and the same time, as we especially note when considering the husband's rationalizing.[16] Unlike a suave other-deceiver, the self-deceiving husband is less than calm in the face of adverse evidence, and his devious treatment of it is not so subtle. His pattern of avoiding evidence, his continual maneuvering to hide the truth from himself, is successful in his own case precisely because he knows when and how to rationalize. He feels the need to rationalize (thus believing that p), and what he thereby does is precisely to rationalize (thus believing that not-p). Just in giving his lame answers to dangerous questions the husband demonstrates incompatible beliefs simultaneously. The husband does not believe that his wife is faithful simply because he wishes it were true. His is not merely a case of wishful thinking, wherein one's emotions control his reasoning and yet where there is no counter-evidence in terms of which one should know better than to believe as he does. Unlike wishful thinking, self-deception forces one to "convince himself" of a belief.[17] There is adverse evidence that the self-

---

12   E.g., it is distinguished from ignorance, error, delusion, and pretending by Demos, "Lying to Oneself," pp. 509-591; from wishful thinking, naiveté, obstinance, vacillation, blatant irrationality, and faith by Szabados, "Rorty of Belief and Self-Deception," passim; from prejudice, willful ignorance, and wishful thinking by Oser, "Invitations to Self-Deception," pp. 1-2.

13   This example was drawn from Bruce, "Investigation of Self Deception," pp. 44-51.

14   As would be suggested by Szabados, "Rorty on Belief and Self-Deception," p. 468; Exdell and Hamilton, "Incorrigibility of First Person Disavowals," p. 392.

15   Cf. Audi, "Epistemic Disavowals and Self-Deception," p. 383; Rorty, "Belief and Self-Deception," p. 394.

16   Lerner, "Emotions of Self-Deception," pp. 151-156.

17   Cf. on wishful thinking: Price, *Belief*, pp. 25, 28; Audi, "Epistemic Disavowals and Self-Deception," p. 382; Szabados, "Wishful Thinking and Self-Deception," passim; Bruce,

deceived husband has to reckon with and deflect. Finally, the husband is not simply making a simple error in self-knowledge; he is not merely mistaken about his own beliefs. Sometimes a person can detect that what he has casually assumed or said that he believes does not square with his observed behavior; the inconsistency in his behavior is readily recognized when pointed out to him (e.g., "Well, you're right, I guess I don't believe that <u>Mind</u> is the most interesting and challenging journal after all; I always read it last, and only if I have time"). This is not so with the self-deceived husband. He has a strategy of avoiding recognition of his behavior as inconsistent. Consequently we say of this man that he is not wavering in belief, insincere, engaging in wishful thinking, or simply mistaken about himself. His rationalization demonstrates a distinct condition of incompatible beliefs known as 'self-deception.'

4.2 The Nature of the Incompatibility

We have seen above that in self-deception S believes two propositions which are incompatible with each other. It should be noted, however, that S's deceived belief need not be the direct contradiction of what he (silently) believes; although the induced belief can be contradictory to his unacknowledged belief, it is not necessary that it be so. What is necessary is that the self-deceiver attempt to deny or escape his dreaded belief that p. While he may do so by arguing against p and maintaining not-p, his most immediate concern is to belie his belief that p--to counteract his taking p as true and recognizing the import of the evidence. Thus he (minimally) holds a false belief <u>about what he believes</u>; he will not acknowledge that p has been accepted as evidenced--that he has believed it. This distinction will enable our analysis of self-deception to cover a wider range of cases, for it sometimes happens that the form of deception to which a self-deceiver succumbs does not transform him into an advocate of an opposite point of view (so that he asserts and defends not-p) but into an agnostic on the question (so that he denies having believed p, without going on to believe not-p).

People not only hold beliefs (e.g., that Chicago is in Illinois, that salt dissolves in water, that Aunt Mary is coming for a visit) they hold beliefs about beliefs--that is, beliefs of a higher order (e.g., that geographical beliefs are warranted by an encyclopedia, that a particular scientific belief is held by another person, that one's own belief is not pleasant). Some of these higher or second-order beliefs are about one's own beliefs; one believes that his beliefs are such and such. S may believe that the prices will rise (i.e., a belief

---

"Investigation of Self-Deception," chapter 3.

justified on the basis of economic facts), and he may believe that he believes that the prices will rise (i.e., a belief justified on the basis of certain personal facts); these are logically different matters.[18] If belief can be likened to a map by which we steer, a belief about one's own beliefs would be likened to a map of one's map--similar to an inset portrayal of the map on the map itself.[19] To say that S believes that he believes p is to say that he relies in his theoretical inferences and/or practical actions or plans on the proposition that he relies in his theoretical inferences and/or practical actions or plans on p. We can imagine, for instance, that S applies for a job and must answer certain questions on a personal data sheet, some of which inquire as to his beliefs about things (e.g., "Do you believe that women are inferior to men?"). When he answers such questions he is reflecting what he believes about his beliefs. Imagine that S has never had such a question brought consciously to mind in the past, but now he must say whether he believes women to be inferior or not. It surely makes sense to think of S stopping to reflect on the matter, going over his past behavior in memory, weighing the various lines of evidence as to whether he had been relying on the proposition that women are inferior or not. Further, we can imagine that he makes a mistake in answering such a question about his beliefs; after considering the question and deliberating over the evidence from his past experience (internal and external), S might decide that upon reflection he does not believe that women are inferior--even though in actual fact he does rely on that proposition in his theoretical or practical inferences, as those who know how he treats his wife habitually could testify. The questionnaire brought S to have a belief about his beliefs (in this case, a mistaken one). This belief about one's own beliefs cannot be reduced to a simple belief (first-order) or expression of belief that p. In the first place, the one belief is about women's inferiority, while the other belief is about oneself; in the above case S's answer on the questionnaire cannot be the expression of a belief in women's equality because in fact he does not believe that (and yet neither is he lying). S's answer is an expression of his belief about his beliefs--that is, an expression of his belief about his state of mind (as explaining his behavior) or about himself. As James Cargile points out, to say that someone believes that he believes p is uncommon, but there is nothing wrong with being uncommon in this case the uncommon phrase has as natural a use as any, without anything turning on the speaker's

---

18 Since third-order beliefs only rarely vary in truth or practical significance from second-order beliefs (Ackermann, *Belief and Knowledge*, p. 24), they will not be considered here.

19 Armstrong, *Belief, Truth, and Knowledge*, pp. 3-4.

philosophical confusion or lack of understanding of the English language.[20] His observation is supported by the extensive discussions carried on by competent philosophers on the topic of second-order beliefs or 'believing that one believes' throughout a wide range of journals.[21]

The grounds for ascribing a second-order belief to a person are somewhat different than those for ascribing a first-order belief,[22] although they are often indistinguishable to casual observation. Ronald de Sousa discusses the fact that second-order beliefs are usually tied to the linguistically articulate self; their acquisition demands some linguistic sophistication, and they are commonly attended with assent or conscious attention (whereas first-order beliefs can more easily remain something of which the person is unaware). Thus it happens that ascribing beliefs is dependent upon observation of the full range of one's behavior, but ascribing second-order beliefs (as in the above example) is tied to that person's verbal behavior (i.e., what he avows and disavows).[23] Moreover, it may be that the mistaken or strange instance "wears the pants" in the family of second-order belief utterances; we are brought to speak of S's beliefs about his beliefs when something seems fishy, anomalous, or downright wrong about his belief avowals in the context of his other behavior.

Because believing a proposition does not automatically bring with it the belief that one believes this proposition, it is possible that someone not believe that he believes p, or even mistakenly believe that he does not believe p. There is nothing about the mental state of belief which requires that it be self-intimating. Being in the mental state of believing p does not thereby create the separate mental state of believing that you believe p. Some epistemologists have, to be sure, contended that 'S believes that p' entails 'S believes that S believes that p.' However a few observations are in order here. First, Castaneda maintains that this is obviously not so,[24] and he goes on to show the extensive qualification that is required in order to hold to such an entailment. Second, when Lehrer argues in a way which supports the idea that believing p entails believing that you believe p, he makes it very clear that he is "not attempting to analyze the ordinary meaning of the words 'know' and 'believe.'"[25] Third, Cargile offers this related consideration:

---

20   James Cargile, "On Believing You Believe," *Analysis* 27, No. 6 (June 1967):179.

21   E.g., *Synthese* 21 (1970) contains a number of articles on the subject.

22   Ginsberg, *Mind and Belief*, pp. 48, 51.

23   de Sousa, "Review Discussion: Self-Deception," pp. 317-319.

24   Hector-Neri Castaneda, "On Knowing (or Believing) that One Knows (or Believes)," *Synthese* 21 (1970):191.

25   Keith Lehrer, "Believing that One Knows," *Synthese* 21 (1970): 133, cf. 135.

Some logicians have constructed systems with a modal belief operator, which have an axiom to the effect that believing entails believing you believe. . . . The trouble with this axiom is that it needs some explaining itself. By applying it repeatedly I can derive from the fact that I believe that Johnson is President of the United States the consequence that I believe that I believe that I believe that I believe that I believe that I believe that Johnson is President of the United States. And it would be natural to want to know what this means. Perhaps such inferences can be eliminated by restricting the number of times the axiom can be applied to a given belief, but we should then ask the reason for this arbitrary restriction. . . . [And moreover] then the axiom is of no philosophical interest.[26]

Eberle responds to the principle of iterated belief, that if one believes that p then one believes that one believes that p, by appealing to the ordinary conception of belief:

The customary notions of belief and knowledge seem such that (a) if a person acts as if it were the case that p, and the outcome of his actions matters to him, that tends to confirm that he believes that p, (b) his outspoken denial that he believes that p tends to confirm that he fails to believe that he believes that p, and (c) such actions and denials are not treated as both confirming and disconfirming the very same thing.[27]

He concludes that the mentioned principle of iterated belief does not seem plausible. Even if it should be uncommon or restricted to certain kinds of contexts (say, psychological or religious), the notion that someone believes that p and yet believes that he does not believe that .p can be readily understood, used, and illustrated. "Someone may flaunt atheism and find that under pressure he reverts to prayers. And if someone wants to say 'He believed that he did not believe that there is a God, but he was wrong,' we can understand what he is getting at.[28] Cargile goes on to give a lengthy example in a more mundane area, a belief about going to the theater; the scenario is imaginative, but it shows us a natural use of the words 'S believes that p, and

---

26 Cargile, "On Believing You Believe," p. 183.

27 Rolf A. Eberle, "A Logic of Believing, Knowing, and Inferring,' *Synthese* 26 (1974):362.

28 Cargile, "On Believing You Believe," p. 179.

S believes that he does not believe that p.'

> Suppose that a research institute has, over a period of years, built a fabulous lie detector with a staggering record for accuracy. The machine has a formidable array of attachments to go around the subject's head, wrists, midsection etc., and when the subject says something into a speaker, the machine can respond 'yes', 'no', 'in a way yes and in a way no', or 'no comment', to the question of whether the speaker believes what he says. Within the limits thus imposed by its terminology, the machine's -performance is astounding. It has never been found to be wrong. It has disagreed with some criminals who refused to confess, but this is to be expected. On religious and psychoanalytic cases the machine is inclined to say 'no comment', but when it risks a more positive estimate, the best psychiatrists are awed by its shrewdness. Since the original genius who first started work on the machines, there is no one in the huge research team who has a good overall grasp of how the machine works. But they know it is in good mechanical order the day it tells Jones, a member of the team who was speaking into the machine while warming it up, that he does not believe that he went to the pictures last Tuesday.
>
> This is disturbing to everyone, because Jones insists he did go, and investigation quickly proves that he did. The ticket-seller remembers him, and Jones' wife went too. And the machine credits Mrs. Jones with believing she went. A thorough investigation is ordered. The machine is working as usual on all other subjects and all other beliefs of Jones. The research team is faced with the following possibilities: (1) the machine is reacting to some irrelevant feature of Jones which has nothing to do with the question whether Jones believes he went to the pictures last Tuesday, by a fantastic coincidence; (2) something has got into the machine, a ghost perhaps; (3) Jones does not believe he went to the pictures last Tuesday, amazing as it may seem.
>
> The researchers reject (1) because the machine has worked on such a wide range of subjects that the chances of its reacting to some trait unique to Jones are astronomically low. They reject (2) because they are sure the machine is a machine and they don't believe in ghosts. So they accept (3). They reason that belief is an imperfectly understood notion and the machine is an expert on it. They don't understand the machine, but neither does the

man in the street who is told by the mathematician that there are sets whose members can be paired off with the members of a proper subset of themselves, or by the physicist that the universe is four dimensional, or finite but unbounded, etc., fully understand his informant. The ordinary man was sure it was otherwise, but he gives up saying so and believes what he is told. His understanding may be incomplete, but he is better than a mere parrot, and is wise in deferring to authority. .The researchers see themselves as in a similar position. True, the ordinary man can aspire to go to a university and get explanations, while it seems nothing like this is open to the researchers. But they hope for an explanation someday too, even if it is not so clear where it will come from.

So the final decision is that Jones went to the pictures last Tuesday but does not believe that he did. And Jones agrees. He wants to sign the report with everyone else. And a few months later, the fantastic coincidence which had been ruled out turns up. The machine was reacting to Jones' having an extremely rare form of dandruff. It does the same with anyone suffering from this disease. May we not say, then, that Jones believed he went to the pictures while mistakenly believing that he did not believe this? What more natural use for 'Jones believes he does not believe that he went to the pictures' could we hope to find?[29]

The preceding observations have been offered in order to explicate the notion of a second-order belief. Given the characterization of belief offered in chapter 2, one can rely on a proposition in his theoretical and practical inferences, and it sometimes turns out that this proposition is about what he relies on in his theoretical and practical inferences; here we would have a mental state which intends another mental state. What this thesis maintains is that in self-deception S believes that p, and he also believes (minimally or implicitly) that he does not believe that p.[30]

This bears a noteworthy resemblance to other-deception. When R has been deceived by S (either by coming to believe a false proposition or being

---

29   Ibid., pp. 179-180.
30   The personal pronoun mentioned in the iterated belief is important here. We do not say that Smith believes that Smith believes p, but that <u>Smith</u> believes that <u>he</u> believes p. These two forms do note express the same thing technically, and so the personal pronoun is crucial in self-belief sentences. Cf. Hector-Neri Castaneda, "Indicators and Quasi-Indicators," *American Philosophical Quarterly* 4 (1967) 85-86; Cf. "On Knowing (or Believing) That One Knows (or Believes)," pp. 191-194.

kept from a true proposition), R may not have formed an explicit belief about any of S's beliefs. <u>If</u>, however, R were to consider S's beliefs, R would not think that S believes contrary to the deceived belief held by R--otherwise R could hardly be misled by S. If R has come to believe (falsely) that the sports car being driven by S belongs to S because of S's behavior, attitudes, or words, then R (at least implicitly) believes that S does not believe otherwise about the ownership of the car. If R believed that S believed that the car belonged to someone else, then R would not be led to believe that the car belonged to S--even if S directly lied about the matter to R. As long as R believes that S believes other-wise, R cannot be taken in or deceived into the false proposition. Thus we can say that a condition of other-deception (whether intentional or inadvertent) is that the deceived believes that the deceiver does <u>not believe</u> that p (when the deceived has come to believe falsely that not-p). This observation can now be applied to self-deception. Seeking first and foremost to escape his awareness of the (dreaded) truth-- his belief that p--it will be sufficient for S to believe that he does not believe that p (i.e., to have a false second-order belief about his dreaded first-order belief). Although he might, he need not go to the extreme of believing that not-p; even if he does, though, he will still implicitly believe thereby that he does not believe that p. And when the self-deceiver takes the stronger course of believing the outright contradiction of his (dreaded) belief, this will be in the service of avoiding his belief.

To conceal the fact that he <u>believes</u> that p, S will often be required to attempt to conceal the <u>truth</u> of p; otherwise his not believing that he believes that p could appear arbitrary, unreasonable, or self-serving.[31] Therefore, it is usually the case that the behavioral indicators of the false second-order belief (that he does not believe that p) will overlap with and appear indistinguishable from the indicators of a false first-order belief (that not-p). To put it another way, a man who believes that he does not believe that astrology determines one's fate will behave in ways which are virtually identical with those of a man who believes that astrology does not determine fate. Therefore, although there is an important logical difference between the two forms of disbelief (viz., not believing that p, or believing that not-p),[32] in the remainder of this study I will use the weaker formula (not believing that p) to cover both attitudes. The two tend to coalesce in practice (at least for the purpose of detecting self-deception), and in self-deception the weaker form is implicit in the stronger. If .the self-deceiver were to be asked if he believed the dreaded

---

31   Cf. Bruce, "Investigation of Self-Deception," pp. 53-54.
32   Cf. Price, *Belief*, p. 262; Ackermann, *Belief and Knowledge*, p. 26.

proposition, he would answer that he does not (whether he would go on to assert its contradiction or not). Using the example offered above of the husband who deceives himself about his wife's infidelity, the two possible forms of incompatible beliefs would be as follows:

> A-1: S believes that his wife has been unfaithful
> A-2: S believes that he does not believe that his wife has been unfaithful
>
> B-1: S believes that his wife has been unfaithful
> B-2: S believes that it is not true that his wife has been unfaithful

For convenience these two forms of incompatible beliefs may be designated "first-order incompatibility" (B) and "second-order incompatibility" (A). Second-order incompatibility is minimally required in self-deception, and it is always implicit in first-order incompatibility.

The paradoxical appearance inherent in cases of self-deception is due to the incompatibility of beliefs held by the self-deceiver. In other deception S believes one thing, but R believes something incompatible with it; in self-deception S and R are the same person, in which case one person holds two incompatible beliefs. That strikes us as strange, somehow irrational. And of course such a conflict state is odd or awkward. It amounts to believing what you believe is false. Believing p, S would ordinarily recognize the unlikelihood of not-p or of his not believing p; yet he believes that he does not believe p anyway. The nature of the incompatibility of these beliefs is not narrowly logical; it is broader than the incompatibility of contradictory propositions.[33] The incompatibility is practical or behavioral. It is practically incompatible for S both to believe and to disbelieve; such calls for inconsistent behavior on his part. Because S believes that his wife has been unfaithful (or, that he does not believe that his wife has been unfaithful), he should be lovingly solicitous toward her and back up the claim to her virtue. As a result, his attempt to do one is marred by the influence of the other; his attempt to be loving is a matter of overacting, and his attempt to be rational toward the evidence degenerates into rationalizing. It is in a practical sense that it is awkward, strained, and difficult to believe second-order incompatible beliefs.

4.3 Objections to Analyzing Self-Deception as Incompatible Beliefs

---

33   Cf. Lerner, "Emotions of Self-Deception," pp. 141-142.

Various reasons have been offered for rejecting any account of self-deception which describes it as a conflict state of incompatible beliefs. Before proceeding on in our discussion to the motives and mechanisms for self-deception, these objections should be answered. First, it has been claimed that incompatible beliefs are logically impossible or inconceivable; it has appeared to some writers that making the self-deceiver out to disbelieve what he believes creates an unacceptable paradox in the analysis. Such critical claims as these are, however, mistaken. On the present analysis the object of the self-deceiver's belief is not said to be mistaken, nor is a contradictory description of his belief offered. Rather, the self-deceiver is said to have two beliefs at one and the same time: a first-order belief and a (false) second-order belief. No contradiction appears in this description, for the mentioned beliefs are not on a par. They are detached beliefs and should not be conjoined as a single, complex object. Moreover, since 'S believes that p' does not entail 'S believes that he believes that p' there is not an implicit contradiction to be drawn out by deduction.

Even if we consider the case of a first-order incompatibility in self-deception (S believe that p, and S believes that not-p), there still is no logical contradiction. Many writers have pointed this out,[34] and some even feel that a person could recognize and assert these conflicting beliefs while retaining them both.[35] But the important fact is that the truth-functional components of the sentences analyzing self-deception are not placed in contradiction with each other. The sentences are consistent; it is the self-deceiver who is in some sense inconsistent. It is not impossible to believe things in a conflicting fashion, but it is an unstable condition. The fact is, as observed by a number of philosophers (cf. section 2.2.2 above), that not all of our beliefs are rational ones--precisely because they contradict other of our beliefs. Such mental states should not be defined out of existence. A historian might write that a certain general was active and decorated in the Second World War; he might write a number of articles throughout his life on these accomplishments and draw inferences from his belief that the general was active in the war. Yet he might else- where list the date of this general's death as prior to the Second World War, and he might write articles which assume that time of death. One mental state (viz., the belief that this general was active in the war) causes a particular form of behavior and thinking, while the other mental state (viz., the belief that the general died prior to the war) causes a different form of

---

34   E.g., Canfield and McNally, "Paradoxes of Self-Deception," pp. 140-144; Shea, "Self-Deception," pp. 9-10; Pugmire, "'Strong' Self-Deception," p. 343.

35   E.g., Collins, "Unconscious Belief," pp. 670-671.

behavior and thinking. Human nature is <u>capable</u> of such incoherence, even if we believe that it <u>ought</u> to be avoided, even if it usually appears only in children or the intellectually naive and careless. The plain fact is that people sometimes rely on one proposition in some area of their lives and rely on its contradiction in other areas of their lives. The two mental states and their consequences are separate states of affairs which do not cancel each other out. Bertrand Russell once wrote:

> If it were said that to believe both [of two contradictory propositions) is a psychological impossibility, we would rejoin that, on the contrary, it is often done, and that those who cannot yet do it need only practise the 'will to believe' until they have learnt to believe that the law of contradiction is false--a feat which is by no means as difficult as it is often supposed to be.[36]

The characterization of belief which we have offered in chapter 2 above was drawn from the account given of it by D. M. Armstrong. The same author goes on to say, "It is possible to believe contradictory propositions simultaneously."[37] This is not to say that S believes and does not believe that p; that would mean that he is, and is not, in some mental state--which is logically precluded. Rather, it is to say that S believes p, and that S also believes that not-p. Here we have two separate mental states with differing causal influences in behavior (influences which, because of the nature of the case, are in tension with each other).

> If we identify beliefs with categorical, structured, states of the believer, we can give a straightforward account of a situation of the sort Bap & Ba-p. There are simply two numerically different states, the one encoded for the proposition 'p', the other for '-p'. Manifestations, if they occur, are divided without difficulty into two classes by their <u>causes</u>: the two different states. The manifestations of the belief that p are brought about (in conjunction with other factors) by the belief-state which is the belief that p, while the manifestations of the belief that -p are caused by the belief that -p.[38]

---

36   Bertrand Russell, "Pragmatism," in *Philosophical Essays* (London: George Allen & Unwin, [1910) 1966), p. 82.

37   Armstrong, *Belief, Truth, and Knowledge*, p. 104.

38   Ibid., p. 105.

Relevant to this kind of situation, Armstrong remarks that the human mind is often a disorderly and untidy kind of thing. A person can believe that p and believe that -p simultaneously--sometimes failing to bring the two beliefs together in conscious reflection (perhaps for emotional reasons, cf. Orwell's "double-think"), but sometimes becoming conscious of them and still not giving one up (e.g., Hume's comment in the Appendix to the *Treatise of Human Nature*: ". . . there are two principles which I cannot render consistent, nor is it in my power to renounce either of them").[39] Armstrong continues:

> Recognition of one's own irrationality does not necessarily abolish it. And if distinct beliefs are distinct structured states, then it is easy to see how the belief that p, the belief that -p, and the knowledge both that the two beliefs are held and that they are incompatible, could co-exist in the one mind.[40]

To say, then, that S holds contradictory beliefs is to say that S relies on p in his theoretical inferences and/or practical actions and plans, and that S relies on -p in his theoretical inferences and/or practical actions and plans. An example of this would (sometimes) be S's saying one thing and doing another; for instance, S avows that the bank will close early today, and yet he does not prepare his deposit slip and drive his paycheck to the bank until late in the afternoon. Sometimes it would be unfair to write this off and discredit S in this fashion, "He says that he believes that p, but he really does not." A lie detector test or extensive observance of S might very well vindicate his sincerity in avowing p, even as his actions "speak" contrary to his words. It is logically and psychologically possible to believe contradictory things.

The personal expression of that conflict or instability should not be dismissed as nonsense out of hand; it is perfectly possible that a person who expressed his incompatible beliefs would have a profound insight into himself, and not be fostering an incoherent description of himself which is to be precluded as impossible.[41] It is not a logical truth that a belief must exclude all of its rivals. It is an obvious fact of experience that this does not happen for us or others. It is especially strange for a philosopher--whose work in large measure calls for detecting and correcting contradictions in one's thinking--to maintain that people cannot hold contradictory beliefs. So then, it may be contradictory of a person to believe p and believe not-p, but

---

39   Ibid., pp. 104, 105.
40   Ibid., pp. 105-106.
41   Cf. de Sousa, "How to Give a Piece of Your Mind," pp. 53, 72.

it is not contradictory to <u>say</u> that he does believe these things.

If this is true in the case of a first-order incompatibility, then it is all the more obvious in the case of a second-order incompatibility. Shapiro reflects agreement with the characterization of belief offered earlier in this study when he concludes:

> In the end the "paradox" of contrary beliefs turns out to be a false paradox. That men have contrary beliefs appears odd at first sight, but when looked into the oddity vanishes, like the oddity of being sixteen on one's fourth birthday. The source of the strangeness was ultimately found to inhere in the assumption that the mind is transparent and rational. This, taken together with the correct view that we can only believe what we think true, does make holding contrary beliefs a case of explicit contradiction and so impossible to believe. It is also clear what sort of impossibility this new paradox involves. It is a case of true paradox. The two claims, the assumption of the mind's transparency and rationality, along with the definition of belief, leads [sic] to the denial of contrary beliefs, and is [sic] a case of <u>reductio ad absurdum</u>. The solution to the new paradox requires the rejection of the transparency model of mind and the attendant claim of the mind's rationality. If we replace this view of mind with a more complex account that allows for "unpicturable" and unintegrated beliefs, then paradox is avoided and the existence of contrary beliefs is explained. Thus the "paradoxes" of contrary belief come down to a mistaken conception of mind, an inadequate theory parading as an obvious fact. The answer to "Why does self-deception as involving contrary beliefs appear odd?" is that we wrongly presume a too rational account of mind.[42]

Another line of objection to incorporating incompatible beliefs into our analysis of self-deception would argue as follows.[43] If S believes p, then p is necessarily reportable by S. Such a report would be in one of the following two veins. (1) In connection with his declaration of the false belief, S might say "I believe p, and I believe-not-p." (2) Or in order to preserve the unacknowledged nature of the belief which is incompatible with his professed (false) belief,-S's report might be given like this: "p, but I do not

---

42   Shapiro, "Self-Deception," p. 97. (The reference to being sixteen on one's fourth birthday pertains to someone born on February 29.)

43   Paluch, "Self-Deception," pp. 268-270.

believe it" (cf. Moore's paradox). The first option declares awareness of the conflict and thereby dissolves self-deception into mere vacillation (or some related condition), and the second option (according to some philosophers) cannot be accepted as making an intelligible assertion. In answer we can note the following things. The premise that belief is necessarily reportable is questionable. Regardless, the fact is that the reason why S does not report his belief that p is <u>not</u> because he cannot do so, but because he <u>will</u> not do so (for reasons to be discussed in the next chapter). Further, if the presence of incompatible beliefs entails S's ability to confess his belief that p, we should still observe that the disposition to confess this belief need not be simultaneous with the holding of the belief. Perhaps S can indeed confess his belief that p, but always in the past tense--e.g., "I guess that I really did believe p all along." This observation also undermines the appeal to Moore's alleged paradox, for the expression of S would now be case like this: "p, but I did not previously believe it." If the objector now insists that any belief <u>must</u> be presently reportable, we should dispute the claim by reference to the distinction between occurrent and dispositional beliefs, and to the fact that people do make mistakes about (at least some) of their beliefs. Finally, it should be obvious that a declaration of incompatible beliefs does not in itself <u>eliminate</u> the incompatibility which constitutes self-deception, but rather <u>affirms</u> it. People have been known to confess such an incompatibility with that incompatibility continuing to hold for them.[44] The suggestion that self-deception would require an awareness: and appreciation of one's incompatible beliefs on conflict theory proposals, and that thereby the effort to deceive oneself would be undermined, will be treated in a later chapter.

A third objection to the conflict state analysis of self-deception claims that it cannot genuinely distinguish self-deception from such related conditions as lying, pretending, vacillation, change of mind, half-belief, faith, self-conscious irrationality, etc.[45] In some cases such a claim rests upon faulty premises (e.g., that in self-deception S must be unaware that his professed belief is against the evidence, or that genuine belief cannot be present in a conflict state). For the rest, the distinction between self-deception and these related conditions has been already laid out above. The reason why S is said to believe what he professes, even though it is incompatible with the evidence which he has recognized (in his earlier belief), is that he shows behavioral indications of sincerely believing it; he applies a form of rationality in support

---

44   Cf. Pugmire, "'Strong' Self-Deception," pp. 340-342; Lerner, "Emotions of Self-Deception," p. 149; Shea, "Self-Deception," pp. 11-12.

45   E.g., Gardiner, "Error, Faith, and Self-Deception," pp. 232235; Szabados, "Self-Deception," p. 54; "Rorty on Belief and Self-Deception," p. 468.

of it, and shows no signs of conceding the opposite.

A final objection is similar to this preceding one. The argument is that the claim that S holds incompatible beliefs is in principle unverifiable. Nothing could show that such a conflict state exists, for we can never know whether we have evidence of conflicting <u>beliefs</u> or conflicting <u>evidence</u> about beliefs. That is, we could never distinguish incompatible beliefs from vacillation, change of mind, ignorance, etc. Nothing could jointly discredit all of the other options. Moreover, even if S later confessed to believing p while professing not-p, this would not prove that incompatible beliefs were actually held, for we cannot distinguish between S realizing an earlier truth about himself and S coming to a new belief about himself.[46] These objections, however, can be disputed. In the first place, they rest on the assumption that incompatible beliefs are incoherent and thus impossible; accordingly, they say, nothing could count as evidence for them. But this is imposed prejudice, nothing more. It is not logically impossible for S to hold incompatible beliefs, and the fact is that people do it all the time. Further, the claim that nothing could count as evidence of this state of affairs is mistaken. There are times when S's rationalization of counter-evidence, personal habits, etc. cannot be explained on the assumption that related conditions are present (e.g., half-belief, vacillation, change of mind). The various options can be eliminated after continued study of the evidence, and S's false belief does not fit readily into any of the standard categories for explaining mistake (cf. the example of S's deceiving himself about his wife's fidelity above).[47] The evidence, upon inspection, is not conflicting or ambiguous; certain features of S's behavior clearly indicate a belief that not-p; there is no evidence that S is changing his mind or unsure of himself. Thus the conflict is not in the evidence as such, but in what the evidence indicates about S himself. If the objectors complain, saying that such allegedly "clear" evidence cannot really be an indicator of conflicting beliefs because those who dispute this possibility need <u>never accept</u>, that this "evidence" is such, our response can be short. This is always true for any claim, for no claim is immune from revision or absolutely forced by the evidence; those who wish to escape it can find theoretical measures by which to do so. Such tactics are, however, vain. We all know from our own personal experience the difference between indecision and holding incompatible beliefs; as such we can imagine what the difference would be in evidence for the two respectively.

Therefore, we can conclude this chapter by noting that to this point in our

---

46   Shea, "Self-Deception," pp. 11-21; Siegler, "Demos on Lying to Oneself," p. 470; "Self-Deception," pp. 31-32.

47   Cf. Lerner, "Emotions of Self-Deception," pp. 148-151.

study we have seen how self-deception requires believing a false proposition, how it can be modeled on other-deception, and how incompatible beliefs need to be ascribed to the self-deceiver. This chapter has shown the necessity of the incompatible beliefs and discussed their basic nature. What remains to be indicated in this analysis of self-deception is that this conflict state comes about in a specific way or context.

Chapter Five
Self-Deception as Motivated Rationalization

5.1 Rationalizing Adverse Evidence

In what has gone before we have spoken of self-deception as a state in which S can be found--namely, the state of holding incompatible beliefs of a certain sort. And this is true to the way in which we often speak of self-deception: e.g., "Nixon is not an ignorant man; his condition must be one of self-deception." It is, however, equally true that we speak of self-deception as an activity: e.g., "He is deceiving himself to think that she really cares for him." Self-deception is as much an activity as other-deception is; both can refer to misleading someone from the truth. Audi does not look upon self-deception in this way, as the following quotation makes clear:

> I am not suggesting that self-deception is voluntary...and I do not believe it is an act. But just as one can deceive someone by a complex pattern of behavior without an <u>act</u> of deception proper, so one can get oneself into a state of self-deception without an act deserving that name.[1]

It is not quite clear how we are to take this claim, and Audi presents no argumentation or illustration of it apart from what is here cited. Of course one can unintentionally deceive another person, but usually not without doing something (or failing to do something when it is expected). And even if there were a corresponding self-deception which is unintentional, one would still presumably need to <u>do</u> something to "get oneself into a state of self-deception" (to use Audi's words). The fact that self-deception <u>can</u> be intentional, as will be argued in the next chapter, will indicate that it should

---

1   Audi, "Epistemic Disavowals and Self-Deception," p. 384.

count as a human action. Self-deception does give the impression of being intentional and active. In the idioms by which we refer to it 'deceive' is ordinarily an active verb, and deceiving oneself certainly sounds like an action (e.g., we say "he is trying to deceive himself . . .," or "he is lying to himself"). Furthermore, self-deceivers are typically held responsible for their condition and blamed for it, as though it were not something that "couldn't be helped." Finally, the control that a self-deceiver must exercise over his attention in order to induce a false belief in himself (as discussed in chapter 2) is certainly the kind of thing people can intentionally do.[2] Thus Szabados seems to be more accurate than Audi, saying:

> It seems to me that there are two essential features of self-deception that we need to be reminded of. First, belief is, quite literally, involved in self-deception. Secondly, self-deception involves actions although it does not wholly consist of actions. Philosophers in general have been so preoccupied with self-deception as a 'problematic' mental state that they have neglected what might be called the 'dynamics' of self-deception. It is this feature of self-deception that we have in mind when we describe someone as <u>deceiving</u> himself or when we tell someone to stop deceiving himself. . . . It is when our young man's cherished belief is challenged, when evidence is brought forth which is against his belief (p) and supports the contrary belief (not-p) that the dynamic aspects of self-deception are most clearly seen. By resisting the natural implication of the evidence, by forcing an unnatural inter-pretation on the evidence, he obstinately clings to his belief.[3]

Self-deception is not a condition imposed from outside the person; he does not "undergo" self-deception or become infected with it. Rather, he does something to himself or for himself; he carries out a project. In other-deception we recognize that R can be in the state of mind of being deceived, or not thinking correctly about something; correspondingly, S can be seen as engaging in the activity of deceiving R by any action which generates, or tends to generate, that state of mind. Likewise, in self-deception S can be said to be engaging in the action of deceiving himself whenever his action(s) has the effect of generating a deceived state of mind as part of a conflict state of incompatible beliefs. When some activity, that is, tends to cause S to

---

2   Cf. Lerner, "Emotions of Self-Deception," pp. 160-162.
3   Szabados, "Self-Deception," pp. 62-63.

hold incompatible beliefs of the requisite kind, and that activity is one of the deceived person's, then we say that he is actively deceiving himself.

As is obvious from the <u>kind</u> of evidence we earlier took to indicate the existence of incompatible beliefs in S, the conflict state of self-deception cannot as such have been produced in just any way. If it should so happen that S inexplicably believes p but also believes not-p, we would look upon him as slow-witted or incoherent, not self-deceived. It can happen that S believes p, but believes that he does not believe p, through thoughtlessness, lack of adequate reflection, carelessness and accident. In such a case we would not attribute self-deception to him. The kind of conflict state which counts as an instance of self-deception must come about under certain conditions and in a certain manner. It must be caused by S rationalizing counter-evidence to his professed (false) belief. In a sense that latter belief is manufactured or made possible because of the way in which S manipulates the evidence adverse to it.

One of the ways in which the reflexive form of deception is unique is that it requires the presence of strong counter-evidence in order to be genuine.[4] When S deceives R, he may very well do so even though the available evidence is not clearly or apparently against the way R ends up thinking.[5] But when the deceiver and deceived are the same person, the recognized presence of evidence which is adverse to S's professed (false) belief is essential to the story. S believes a falsehood in belief-adverse circumstances. The presence of this counter-evidence which S appreciates is important because without it self-deception would be indistinguishable from cases of -wishful thinking (hope), intellectual indecision, groundless faith, and the like. We say that S's professed belief is deceived precisely because the available evidence does not really warrant it; this adverse evidence is taken as such by S but unacknowledged, and it is the belief which this recognition represents that gives the lie to his professed and incompatible belief.

Moreover, S's false belief in the face of this counter-evidence is not simply the result of an obscurantism, obstinate irrationality, or emotionalism which is indifferent to the truth. By this professed belief S is not refusing to reflect on any evidence, trying to take others in, or concealing a new lack of certainty.[6] Rorty does not appear to me to be descriptively accurate when

---

4   This condition for self-deception is stressed by Canfield and Gustavson, "Self-Deception," pp. 32-36; Penelhum, "Pleasure and Falsity," p. 258; Siegler, "Self-Deception," p. 42; Bruce, "Investigation of Self-Deception," pp. 96-98.

5   Paluch, "Self-Deception," p. 272.

6   Contrary to Reilly, "Self-Deception: Resolving Epistemological Paradox," pp. 392-393.

she suggests that the person in self-deception may be uncritical, disinterested in the grounds of belief, and holding questions of truth in abeyance.[7] Since the self-deceiver is guarding himself from something, he cannot afford to withdraw from keeping the evidence "under control" in a seemingly rational fashion. When he considers the dreaded proposition, he cannot let the question concerning it be left open; he must rationalize away supporting evidence for that proposition.[8] S's professed belief, false as it may be, is thus for S genuinely a matter of evidence and rationality. That is why S is actually deceived about himself. There is evidence which presents an obstacle to S's easily endorsing his (deceived) belief, and S is not oblivious of that fact. By his reaction he intends to take such evidence into account and give it a credible handling--or, better, what S considers a credible handling. Self-deception involves a curious or perverted rationality. S does not directly blind himself, turn away from considerations of evidence, push through to a mindless faith; he is not simply being stupid, obstinate, or prejudiced. He is confronting the evidence; he would claim that his professed belief is amenable to the evidence. However, although S is not blatantly irrational or unreasoning in the face of the adverse evidence, his general concern for rationality is misguided, artificial, somehow distorted. It appears to others as <u>pseudo</u> rationality.[9] This explains, incidentally, why self-deception is not possible before a person has learned the correct or accepted procedure for interpreting evidence and establishing truth and falsity; they must first be mastered before S can misuse them in the practice of self-deception.[10] S makes a show of following rational standards, but in reality violates them. He does not take the best evidence available or give it a natural construction; he gives undue prominence to implausible matters so as to increase the light thrown on his professed belief (and to hide his recognition that p has support of the evidence to relegate it to shadows). For all of his apparent sincerity, S's handling of the evidence comes across as a "cover story."[11] There are signs of insecurity which hint that S is, through an effort of will, explaining away counter-evidence to reassure himself as well as us. He attempts to be rational, but ends up merely rationalizing.

Thus S comes to, and maintains, a conflict state of incompatible beliefs

---

7   Rorty, "Belief and Self-Deception," pp. 388-390; cf. Szabados, "Rorty on Belief and Self-Deception," pp. 466-467.

8   Cf. Bruce, "Investigation of Self-Deception," pp. 138-142

9   Cf. Hamlyn, "Self-Deception," p. 51; Szabados, "Self-Deception, pp. 64-67; "Rorty on Belief and Self-Deception," p. 467; Siegler, "Analysis of Self-Deception," p. 161.

10   Mounce, "Self-Deception," p. 69; Szabados, "Self-Deception," p. 65.

11   Lerner, "Emotions of Self-Deception," pp. 146-148.

by rationalization of the adverse evidence. He believes p to be true, seeing it as supported by the evidence. That is why the self-deceiver is said to recognize the evidence and its adverse character. Yet he does not wish to hold p; he brings himself to believe that he does not believe p--not by simply denying p, but by rationalizing away the evidence for it. He attempts to conceal his recognition of p's truth (his belief that p) by concealing the truth of p itself (just as the model of other-deception would lead us to expect). In his later discussion of self-deception Siegler came to see self-deception in this light, and his words provide a convenient summary.

> We might hold that A knows (or believes) that p in that he is disposed to acknowledge that p and reveal by what he does that p given certain conditions. But given other conditions he is neither prepared nor disposed to acknowledge that p nor to reveal by what he does that p. The latter conditions lead (cause) him to deny that p and to act in ways that conform to a belief that not-p. Indeed, though it might be an exaggeration, we could say that he actually believes that not-p as a result of a desire that not-p and a fear that p. In this case the full thesis would be that A is in self-deception if and only if:
>
> (1) he knows (or believes) that p.
> (2) he believes that not-p as a result of desire and fear;
> (3) he believes that not-p though he has good reason to believe that p.
> (4) he misconstrues or distorts at the level of evidence and inference.[12]

Although S's pseudo-rationality is not convincing to others (i.e., they do not come to believe something incompatible with the belief that p), it is self-guaranteeing for S because he is such a willing victim of the deception. Consequently we say that he genuinely believes the (false) belief that he professes, even though it is incompatible with another belief of his. Because S rationalizes to avoid the obvious force of the evidence, because he is artfully steering around something, we have reason to say that he believes what the evidence points to (p). But because the evidence is rationalized away in a convincing manner for S, we have reason to say that he believes the conclusion of his line of thought (not-p, or that S does not believe p). His

---

12   Siegler, "Analysis of Self-Deception," p. 161.

behavior simultaneously indicates belief that p and disbelief that p.

By improper focusing on or handling of the evidence the self-deceiver suppresses his belief that p is true (or the truth of p), thereby avoiding the discomfort of openly acknowledging the dreaded truth. He is not merely mistaken, naive, or forgetful regarding the significance of the evidence; he is not simply trying to deceive others. He is actually deceiving himself. At one point S begins to realize p, but cannot bear it. This subsequently causes him to take the evidence to mean something else. By maintaining a distorted concentration on certain aspects of the evidence-situation S can keep himself from acknowledging the uncomfortable truth by believing that he does not believe it, even though he does in fact realize its truth. He unnaturally dwells on his overt, favored belief in order to keep his dreaded belief covert. Such rationalization of the counter-evidence shows that S is not just jumping to a desired (but false) conclusion,[13] or evidencing simple bias.[14] Nor is he merely indulging in wishful thinking. In wishful thinking S's belief need not be false; it need simply be a belief held without adequate substantiation or good reason. There may not be a dreaded conclusion which S wants to avoid, nor does there have to be adverse evidence present in order for S to engage in wishful thinking. Self-deception and wishful thinking have some things in common, to be sure: e.g., a belief is held in which S has a personal stake, and which would not be held in the absence of that motivation. In wishful thinking, however, the evidence, if considered at all, is slightly in favor of S's desired belief and he jumps to the conclusion--not simultaneously believing, or having strong grounds to believe, that this conclusion is false; he is not forced to pervert the procedures of reason. Over against this, in self-deception there necessarily is evidence against S's belief, and S resists this evidence by ingenious tactics. He recognizes the good grounds for taking p as true, but explains away the evidence through pseudo-rationality.[15]

Therefore, we have seen that self-deception can be an activity--the activity of rationalizing counter-evidence of one's professed belief, thereby bringing about a conflict state. If the notion of self-deception is not understood in this way it becomes indistinguishable from other related ones. In order for the holding of incompatible beliefs to be counted as self-deception, then, this situation must have come about under certain conditions (the presence of adverse evidence to S's professed belief) and by a special route (pseudo-

---

13   Ibid., p. 151.

14   Mounce, "Self-Deception," pp. 67-69.

15   Cf. Szabados, "Wishful Thinking and Self-Deception," passim, which parallels the notion of wishful thinking in Mounce, Gardiner, and Lerner, and which contrasts with that of Bruce, "Investigation of Self-Deception," pp. 101-103.

rational concern for the evidence, or rationalization).

5.2 Self-Deception Requires a Motivational Explanation

The self-deceiver's behavior regarding the adverse evidence, his rationalization, is not a mere happening, or state, or feeling; it is a human action, the exercise of a power or ability to avoid, distort, over-emphasize, misconstrue the evidence--to follow rational procedures in an artificial manner (akin to the exercise of an actor's ability to fight wars, play athletic games, make love, etc. in an artificial manner on the movie set). Like other human actions, rationalization can be exercised in a variety of ways: automatically, deliberately, impulsively, unwittingly, reluctantly, intentionally, thoughtlessly, purposely, etc. Not all descriptions of rationalization, therefore, will qualify as cases of self-deception. A man may, for instance, rationalize away all the evidence against him in a law court without succeeding or even trying to convince himself; he may be a perjurer, or simply trying to escape punishment, etc. A scholar may impulsively rationalize the evidence put forth in his rival's latest article, only to catch himself later taking a cavalier and unworthy attitude in this matter. Such cases as these would not constitute self-deception. Therefore, the description of rationalization in self-deception must contain more than reference to its circumstances, manner, consequences, and the like; this activity must be described in particular explanatory terms--it must tell us <u>why</u> the self-deceiver is behaving this way toward the evidence.

It must offer a motivation for the deceived belief. Bruce maintains, to the contrary, that it is impossible to have a motive for believing something because believing is not something we <u>do</u>.[16] This comment is not only directly contrary to our common experience (i.e., people do in fact have motives for believing some things), but it is sufficiently answered in chapter 2 with our discussion of the sense in which believing is voluntary. Penelhum takes a less extreme stand, holding that a motive is not necessary for self-deception (even though it may be possible and customary to have one).[17] We have seen, however, in the last chapter that self-deception is not adequately accounted for by saying that it involves believing something in the face of adverse evidence (i.e., what Penelhum's proposal is left saying). This does not distinguish self-deception from faith, obstinance, stupidity, etc. Without the presence of a motive working on the self-deceiver we would be uncertain that he does in fact recognize the evidence and its import; his ambivalent behavior could

---

16   Bruce, "Investigation of Self-Deception," p. 104.
17   Penelhum, "Pleasure and Falsity," pp. 259-260.

be given other explanations. Gardiner would appear to be correct, then, in responding to Penelhum by saying that a purposive aspect is crucial to self-deception; the self-deceiver does not face the evidence realistically because he has a compelling interest in believing contrary to the perceived evidence.[18] An analysis of self-deception should say why S behaves as he does.

Of course, asking why S did X is notoriously vague or ambiguous as a question. It could be interpreted as asking after the purpose of S's doing X, what drove S to do X, what are the antecedent factors, or law-like factors (habits, dispositions), or teleological factors involved in X, or even the state of S's mind when he came to do X, etc. A human action could be explained in terms of cause, reason, purpose, motive; and given different theories, each of these latter elements could be subsumed or related in particular ways to others of them. The best advice here is undoubtedly to resist trying to give a completely general answer to such a wide-open question, especially when the various kinds of answers offered by philosophers are fraught with further obscurities, disagreements, and difficulties; it would be unrealistic to make my present ambition the resolution of all these theoretical issues. In a more or less pedestrian fashion we can indicate the kind of explanation which must be offered for S's behavior if it is to be taken as self-deception. The explanatory motif is readily identified and put to use in common discourse without creating insuperable difficulties for adequate communication. It will serve our present purposes quite sufficiently.

What we want to say is that the rationalizing behavior of the self-deceiver must be given a motivational explanation. And this can be roughly specified in both positive and negative ways. For instance, human beings do certain acts or take certain actions in a variety of ways; sometimes they are done without great efforts of will, or deliberation over them, or prior processes of intention or decision; S can be said to have genuinely done X, even if he did it unintentionally, non-voluntarily, or for no purpose whatsoever. There are, how-ever, restrictions on motivated behavior descriptions; if X was done from a motive, then this action is incompatible with mere reflex, absent-mindedness, carelessness, force of habit, etc. This observation should not be taken to imply that (in common discourse anyway) a "motive" is some antecedent psychological occurrence, a disposition, or an intention. Feeling tired, tactlessness, and intending to file your tax-form are not explanatory motives for going to bed, going uninvited to the boss's party, or going to the post office respectively. To give a motivational explanation of X is to relate it to certain facts about S's desires (or wants)--in contrast, say, to his habits

---

18   Gardiner, "Error, Faith, and Self-Deception," pp. 239-240.

or what agitates him; motive implies desire (but not vice versa necessarily). For instance, "S went home early out of a desire to play ball with his son"; this offers a motivational explanation equivalent to "What motivated S to go home early was his desire to play ball with his son." Such explanations are amenable to an 'in order to' form of expression, but need not imply that there was some aim (goal) beyond the action which S was trying to achieve: for instance, if S goes to a wedding in order to fulfill a social obligation, fulfilling the obligation is not something <u>further</u> to going to the wedding, but going is the fulfillment. The desire implied in motivational explanations may be implicit and unexpressed, as when we say, "S did it out of gratitude." When we speak of motives, moreover, we speak of them as correlative to actual or envisaged actions; they do not exist in abstraction by themselves but always in reference to deeds. To ask after these motives is appropriate especially under such conditions as these: S's action was somehow important, an 'in order to' explanation is available, but standard or customary reasons do not readily apply (e.g., "What was Nixon's motive for moving out of the White House?"). These are some of the common parameters within which motivational explanations function in everyday speech. When we give S's motive for doing X, we give the desire for the sake of satisfying which the deed was done.

With this general background in mind, we would distinguish the activity of self-deceptive rationalization from other varieties of this activity by describing it as <u>motivated</u> rationalization; S's behavior in handling the evidence is to be explained in terms of a motive S has (i.e., some desire to be satisfied). Self-deception is more than an error about one's beliefs, more than holding incompatible beliefs, more than engaging in rationalization. It involves deception by the self, of the self, about the self, and <u>for</u> the sake of the self's desire(s).[19] Motives are crucial to self-deception. The self-deceiver refuses to deal with the evidence realistically <u>because</u> he has a compelling interest in believing contrary to it. Part of the reason why the conflicting evidence presented by S's behavior (e.g., S seems to be trying to dodge something by his "reasons," and yet he takes the contrived "reasons" seriously) can become evidence of conflicting beliefs is that we discover that the proposition which S denies by his false belief is one which he has a reason to dislike. Hereby alternative explanations of S's false belief can be ruled out.

It is not irrelevant whether S's deceived belief concerns matters which are emotionally vital to him. S's deception of himself may not be ego-centric or selfish in nature (e.g., the false belief need not be about S himself); however, S

---

19   Cf. Rorty, "Belief and Self-Deception," p. 393.

will have an interest at stake in maintaining the rationality (evidential support) of his (deceived) belief nonetheless. For instance it may concern a loved one or an issue of principal importance in S's life. Self-deception calls for a motive, and that is why it is centered on subjects which hold the potential for controversy, where strong emotions can be felt. S cannot deceive himself gratuitously, about just anything at all, but only about that pertaining to which S has some desire. Where S has no feelings on the matter, in cases where S does not think that anything important is involved, we do not find self-deception; about such things S can have no <u>motivated</u> rationalization and thus no self-deception.[20] Of course, we must be careful to note that motivation is a necessary--<u>not</u> a sufficient--condition for self-deception. S can have an interest in a subject without having a vested interest; he can have a preference without being biased in the treatment of evidence, etc. What S avoids and rationalizes away (his belief that p) is something which is somehow threatening, upsetting, anxiety-producing for him; p is distressing to think about and painful to speak of--i.e., S has some negative emotion toward it. The evidence may all point to the fact that S's son will soon die, or that S has lost his job out of incompetence, or that S's supposed lover really has designs for another man, etc., and such is too hard to face. In reaction S manufactures artificial evidence for an incompatible belief; he rationalizes, distorts, misconstrues the adverse evidence out of a desire to minimize his unpleasant feelings. His motive to avoid acknowledging p <u>overrides</u> the available reasons for acknowledging where the evidence points.[21] He has a desire to believe otherwise. His rationalization: of the evidence is motivated rationalization. This explains the ease with which we all fall into self-deception--the issues involved touch us in a way arousing desire, especially the desire to avoid psychic pain.

In conclusion, self-deception is not simply being mistaken with a motive, for that could as well apply to instances of wishful thinking where the belief happens to be false. Self-deception does not reduce to a mere belief that not-p (or that one does not believe p) due to a desire that this belief be true. The self-deceiver cannot take as his conscious reason for believing that he does not believe that p (or for believing that not-p) that he does not want

---

20 Valberg, "Rationality and Self-Deception," pp. 186-206; Lerner, "Emotions of Self-Deception," p. 169.

21 Cf. Bruce, "Investigation of Self-Deception," pp. 81-83, 143149, 177-197. Bruce has a helpful discussion of the nature of being "crushed" by some truth if it were fully acknowledged. For other explorations of self-deception as generated by desire or motive, see: Shea, "Self-Deception," chapters 7, 8; Valberg, "Rationality and Self-Deception," chapter 4; Shapiro, "Self-Deception," chapters 4, 5.

to.[22] On the other hand his desire does not operate upon him as a mysterious and unconscious cause of his belief. Unlike Lerner we would refrain from saying that the deceived belief is not at all the result of reasons.[23] Rather, what happens is that S has a desire to avoid the psychic pain represented by the truth of p; this is not his reason for believing contrary to p, but it does cause him to manufacture the evidence (to distort and rationalize the evidence available) which will justify his believing as he wants to. In issues which touch upon us in a very emotional way such motivated rationalization can successfully bring us to deceived belief since we exercise a control over our attention and over the way in which the evidence will be construed.[24] The professed belief is false, held contrary to another (incompatible) belief, and generated (or sustained) through motivated rationalization. The self-deceiver has a desire to avoid the psychic pain or discomfort associated with taking p to be true; thus, without eliminating that belief, he brings himself to believe something incompatible with it--in the face of adverse evidence, through a process of rationalization. We speak of the resultant state, as well as the activity through which it is created, as "self-deception." To be true to the phenomenon and yet distinguish it from related conditions, our analysis of it must include not only false belief, not only incompatible beliefs, not only deceived belief held in the face of adverse evidence, not only rationalization, and not only motivation; an adequate analysis must incorporate all these elements. Self-deception is a conflict state of incompatible beliefs generated by motivated rationalization of adverse evidence. Our study has therefore led us to conclude that an adequate analysis of self-deception would be this: by *rationalizing the adverse evidence, S brings himself to believe falsely that he does not believe that p, because he is motivated by that belief's distressing character to deny it.*

22 Valberg, "Rationality and Self-Deception," pp. 179-180, 185-186, 219-220, 224-226, 232.

23 Lerner, "Emotions of Self-Deception," pp. 132-134.

24 Cf. Daniels, "Self-Deception and Interpersonal Deception," pp. 250-251.

Chapter Six
Awareness and Purpose in Self-Deception

6.1 The Panacea of Drawing Distinctions
Regarding Consciousness

One lingering Problem pertaining to self-deception still requires some attention. It may be the fundamental or underlying perplexity which makes most of us scratch our heads over the possibility of self-deception. We tend to model self-deception on other-deception, and in the latter case--when we have a strong case of deceiving someone, not merely inadvertent misleading--it certainly seems that the deceiver is aware that his victim's belief is incompatible with his own and is thereby (as he sees it) false. But do we want to say this about self-deception, where S is deceiver <u>and</u> deceived? Is S aware (as deceiver) that his belief (held as deceived) is false? Is he conscious of his conflict state? Are his motives recognized by himself? These are critical questions, for ordinarily if R were aware that the belief promoted by S conflicted with S's own conviction, or that S's presentation of the evidence was governed by an ulterior motive, then R would not be successfully deceived by S. Awareness of these things seems to undermine the effort of deceiving. Therefore, if the self-deceiver were aware of his incompatible beliefs, or the falsity of his professed belief, or the motive behind his rationalization, he could not genuinely deceive himself. Whereas in other-deception the deceiver could make his behavior explicit to himself, such explication would preclude self-deception.[1] If S is not aware of the incompatible beliefs and falsity of the one professed, then he is not a (strong) deceiver; if he is aware, then he could not be deceived.

Obviously some distinction must be drawn to salvage self-deception as a successful enterprise which can be coherently described. As we have

---

[1]   Cf. Paluch, "Self-Deception," p. 269; Gardiner, "Error, Faith, and Self-Deception," pp. 224-225, 230-231, 239-240; Mounce, "Self-Deception," pp. 63-64.

seen earlier, some writers attempt to skirt the problem by speaking of the knowledge or belief in self-deception as not being of the "ordinary kind." In the normal sense the self-deceiver does not have knowledge, belief, motive, intention, etc., we are told. Because such things are kept hidden by the self-deceiver, they are thought to be of a peculiar nature.[2] But such comments are unhelpful and perhaps misleading. First, they do not resolve the paradox of self-deception, but simply reaffirm that there is a perplexity to be resolved; they remind us that an ordinary, paradox-free, description of self-deception seems elusive. Secondly, the peculiarity of the self-deceiver's situation is here portrayed as due to extraordinary elements with which he works (e.g., twilight belief, as-it-were-knowledge) rather than considering that it might be due to an extraordinary combination of ordinary elements or even an extraordinary personal response to (or handling of) the ordinary elements. To be sure, self-deception is somehow peculiar. But the peculiarity might well be that S <u>should</u> know better than he does, but does not. Complicating our epistemic vocabulary can only give the vague appearance of a solution.

Problems also attend the popular attempt to resolve the present perplexity by drawing some sort of distinction with respect to the consciousness or awareness of the self-deceiver. For instance, it is urged that the self-deceiver is not ignorant of his beliefs, only unconscious of them; that is, S makes himself unconscious of his knowledge through quasi-rational techniques that enable him to avoid giving attention to something about himself. Thus the resolution is that the self-deceiver has unconscious knowledge.[3] In a similar vein we are told that the self-deceiver has a deep-level belief which exposes his surface-belief as untrue; that is, the resolution comes in the form of distinguishing between conscious and unconscious belief.[4] Others have maintained that the self-deceiver sustains incompatible beliefs by having two levels of awareness: viz., simple awareness and attentive awareness--the latter, but not the former, involving that S notices his awareness.[5] Another tack is found in those who argue that S gets himself to believe what he disbelieves by selective focusing on aspects of his experience; the result is a distinction between S's strong consciousness of/that something (i.e., that which he readily affirms and on which he focuses explicit attention in the forefront of his mind) and S's weak

---

2   E.g., Lerner, "Emotions of Self-Deception," pp. 142-143; Mounce, "Self-Deception," pp. 65-67, 69.

3   Hamlyn, "Self-Deception," pp. 53-56.

4   Valberg, "Rationality and Self-Deception," p. 164.

5   Demos, "Lying to Oneself," pp. 593-594; cf. for critical responses: Siegler, "Demos on Lying to Oneself," pp. 472-473; Paluch, "Self-Deception," pp. 270-274; Bruce, "Investigation of Self-Deception," pp. 9-13.

consciousness of/ that something.[6] Likewise it is said that if we reject the "vehicular conception of consciousness" (which categorizes it as a form of perception or linguistic expression, thereby requiring some medium for realization) we will be able to distinguish between non-explicit and explicit consciousness, thus enabling us to teach that in self-deception S is conscious of the truth but prevents himself from coming to explicit-consciousness of it, wherein he would spell it out to himself.[7] In the same manner, but perhaps a bit more precisely, Pugmire contends that in self-deception S is aware of an unwanted belief, but he renders that awareness abstract by failing to attend to its details; he gives it minimal and insufficient attention, controlling his thoughts in the face of the truth, and thereby ignoring unpleasant aspects or details of what he knows.[8] The last three proposals are clearly similar to each other, suggesting that the paradox of self-deception can be resolved by distinguishing strong/weak, explicit/non-explicit, or detailed/abstract consciousness (awareness).

The thrust of these various proposals is easy enough to explain and appreciate, and some of the discussions are beneficial in ways somewhat tangential to the current problem (e.g., especially helpful comments on theories of consciousness). Indeed, when this paper's own solution is elaborated shortly there will be some areas of overlap with discussions of this type, but hopefully without the unnecessarily awkward device of a consciousness-distinction itself. Four basic problems attend the foregoing solutions to the paradox of self-deception, all centering on their notions of weak consciousness (e.g., unconscious, unnoticed, non-explicit, abstract). First, the concept of inattentive consciousness (awareness without attention) is, if not unintelligible, at least dubious and far more problematic than the notion which it is intended to explain. Secondly, various analogies and metaphors used break down or offer little assistance in understanding these concepts. For instance, Demos likens his notion to an unnoticed pain (e.g., S goes to the movies to distract his attention from a headache). But how would we ever be able to establish that S did in fact have an unnoticed pain? S could not tell us later about it without remembering the pain, but he could hardly be expected to remember something that he did not notice. Moreover, applying such an analysis to a later confession of self-deception, we would derive the curious formula, "I did not notice my belief that p at the time when I was noticing my belief that not-p." So also, Bruce's discussion of

---

6   Saunders, "Paradox of Self-Deception," pp. 561-563.

7   Bruce, "Investigation of Self-Deception," pp. 13, 68, 75-77, 119-120, 121-155, 173; cf. Fingarette, *Self-Deception*, pp. 34-46, 62-66, 94-100, 113-116, 121, 125, 137.

8   Pugmire, "'Strong' Self-Deception," pp. 344-346.

non-explicit consciousness becomes quite obscure once he gets around to illustrating it. He speaks of a man coming home early one afternoon to find a car parked outside of his house--the car of a suspected paramour to his wife; he is non-explicitly aware of the truth and prevents himself from coming to explicit consciousness by driving out to the lake for the distraction of a beer. However, in discussing this non-explicit consciousness, Bruce eventually says nothing more than that it is like a swirl of feeling (or vague uneasiness, being temporarily disoriented, etc.). Such an analysis leaves much to be filled in and elaborated; it hardly resolves the problem which provoked this discussion. Similarly problematic are the "subterranean" metaphors used by some writers (unconscious or deep-level beliefs, knowledge) due to their opacity and figurative nature. The metaphors are also somewhat inappropriate in that: (1) for the self-deceiver the unwanted information remains somewhat in his "conscious" mind in the form of nagging doubt or desire for further efforts at rationalization, and (2) such a device seems bizarre when we think about the resolution of other apparent paradoxes of reflexivity. For instance, if we think of S as in a relatively simple, unadorned room it is quite difficult to imagine what "hiding something from oneself" could mean or how it could be successfully accomplished. Obviously, wherever S places an object (e.g., his pen) he will know its location, which seems to preclude the possibility of having hid it from himself. And yet we are not prepared to discard the common phrase "hiding it from oneself" as senseless. But surely the way to maintain the sensibility of the expression is not to suppose that S's knowledge of his pen's location has somehow gone underground. This is not how we would naturally explain the use of the phrase. Instead, we would reflect on the circumstances in which the use of the expression is most appropriate; we would "look and see" (e.g., S hides his pen from himself, not just arbitrarily or in any situation, but by throwing it into a crate of identical pens, or by putting it "away" from his children at home in such a thorough fashion that he cannot himself later find it). The subterranean approach is simply artificial.[9] A third major problem with the weak consciousness resolutions of self-deception's paradox is that they call for equivocation at the crucial point of specifying <u>what</u> the self-deceiver believes (knows). Perhaps the most clear explanation of the notion under discussion is Pugmire's. He says that S is aware of <u>something</u>, but <u>not</u> of its <u>details</u>. But this suggests that S is not after all conscious of the specific thing which puts the lie to his professed belief and which motivates his rationalizations. Some "detail" of a situation or some specific truth is dreaded by S; he either believes it, or he does not.

---

9   Champlin, "Self-Deception: Reflexive Dilemma," pp. 281-286.

To speak of S being aware of this situation or truth <u>without</u> its details is simply to say that S is aware of a <u>different</u> truth or object. The situation or truth of Johnny being caught around other students' desks is relevantly different from (although coinciding with) the situation or truth of Johnny being caught with his hand on the lunch money in other students' desks. If S is aware of the former, more abstract situation or truth, he still is not thereby aware of the latter, more detailed situation or truth. Thus the explanations of self-deception in terms of strong/weak, or explicit/non-explicit, or detailed/ abstract consciousness give the appearance of resolving the paradox at the price of equivocating on the object of S's awareness. Finally, all of the above proposals suffer the following defect. They all end up saying in one form or another that S does not explicitly notice or have detailed consciousness of the truth of p; they go on to say that, by various maneuvers in which S focuses his attention and rationalizes the evidence, S brings himself to believe not-p (or, as we suggested earlier, S comes to believe that he does not believe p). But then what are we to say about the alleged conflict state? How are we to make credible our attribution to S of a desire to avoid the truth of p? It turns out, that is, that these various proposals about weak consciousness go too far in their explanation; they resolve the paradox too neatly, not leaving any sense of perplexity about the situation finally. On these accounts, wherein S is not fully aware of the truth of p to begin with, there is really no problem to solve after all. If S does not notice (is not aware of) his unwanted belief, then it exerts no pressure on him to rationalize or to take account of an incompatibility in his system of beliefs.

6.2 The Self-Deceiver's Awareness of Truth, His Beliefs, and Motives

The analysis of self-deception which has been offered already in this study is adequate to resolve the perplexities which arise over the self-deceiver's awareness of various things about himself--an awareness Which seems to threaten the possibility of successful self-deception. To this point we have analyzed self-deception in the following manner: by rationalizing the adverse evidence, S brings himself to believe falsely that he does not believe that p, because he is motivated by that belief's distressing character to deny it. The self-deceiver believes that p, but he brings himself into a conflict state of incompatible beliefs by also believing that he does not believe that p. It is precisely because of the second-order belief that he disavows his first belief. The second-order belief prevents the ordinary disposition to

verbal behavior--to conscious formulations of p and explicit assertions of p (inwardly or externally)--from being actualized in the case of S's belief that p. These acts are resisted and blocked by S's alternative and conflicting avowal of disbelief. He cannot give assent to p precisely because he wishes to assent to something incompatible with it. This failure to give silent or public assertion to p, however, does not mean that S does not see p as true (does not believe that p). As discussed in chapter 2, his disavowal is not incorrigible.

So then, we can ask about the objects of the self-deceiver's awareness of conscious entertainment. What does he introspect about himself? We have said that self-deception begins with S's belief that p. Does that mean that he is aware of p's truth? The answer is obviously yes. S is aware of the evident-ness of p, for that is what we mean when we say that he takes p as true, or reliable, or supported by the evidence. Apprehension of p's truth is essential to self-deception, for without it S would have no need to avoid acknowledging p and would not engage in rationalization in order to see the evidence in a new way--in a way which will not appear adverse to what he prefers to believe. What marks the beginning of a period of self-deception is an occasion on which the self-deceiver is conscious of the truth (takes p to be true, whether it is or not) and, recognizing the psychic pain this produces, seeks to avoid making his awareness explicit--that is, avoids giving assent to p by rationalizing away the perceived evidence and convincing himself (on the basis of this new--reworked--"evidence") that he does not believe that p (or further, convincing himself that not-p). That is, S is aware of p as true, but he will not become aware that he believes p--will not assent and entertain in mind that he believes p. He believes p without awareness (entertainment) of his <u>believing</u>.[10]

By rationalizing the evidence and purposely overlooking affective symptoms of his belief (among which is the rationalizing activity itself) he comes to believe that he does not believe that p. Since he does not believe that he believes that p, he does not assent to p (inwardly or publicly). In this one respect--not asserting the truth--S may seem like a man who is simply ignorant of the truth. But the resemblance ends just about there. Unlike an ignorant man, the self-deceiver shows the slips and mistakes of "bad acting," obviously rationalizes, speaks in a strained voice or is less than calm under cross-examination, etc.; that is, the self-deceiver has the affective signs of trying to cover up something--to persuade himself that something is not

---

10   Szabados notes the necessity of the self-deceiver being aware of the truth so as to steer around the evidence ("Rorty on Belief and Self-Deception," p. 467), and Audi comments that the self-deceiver's belief (believing) need only be sufficiently veiled from consciousness to account for his disavowal of the belief ("Epistemic Authority of First Person," p. 10).

so. In self-deception S keeps himself from entertaining and assenting to his belief that p (keeps himself unaware of his belief) by systematically keeping his awareness of p's truth (i.e., his belief that p) covert. By believing through rationalization that he does not believe that p, he hides his awareness of p's truth. And of course, whenever anybody (falsely) believes that he does not believe something, thereby blocking out one of the key indicators of his awareness of p's truth (viz., explicit assent), others will naturally be in a much better position to observe his behavior and attribute that belief to him. This is just to repeat that believing p is not self-intimating; the first-order belief does not entail or automatically bring a second-order belief which is incorrigible. We can be mistaken about our beliefs, and our assenting with respect to such second-order beliefs can be fallible. Of course this produces a conflict state and makes incompatible requirements on behavior; the affective and cognitive elements of someone's behavior would be in tension, but there would likewise be an affective tension stemming from the belief that p along with the belief that one does not believe p (since the behavior appropriate to believing that one does not believe p is virtually identical with that appropriate to believing not-p). It is insightful to notice how later confessions of self-deception usually involve two elements: the claim that one did not entertain with assent (was not "aware" of) the thought of believing p, and yet was responsible for this "ignorance" because he should have known better (i.e., he believed--was aware of the truth of--p).

The above comments about the self-deceiver's awareness of the truth of p satisfy the necessary condition for <u>deceiving</u> himself that not-p (or that he does not believe p). The fact that the self- deceiver is not aware that he believes p (i.e., does not believe that he believes p) allows for him to assent sincerely to something incompatible with that belief, thereby satisfying the necessary condition for being <u>deceived</u> (i.e., for having a false belief). He cannot deceive himself into believing what he is aware is false.[11] The self-deceiver believes p, but must bring himself to believe falsely that he does not believe p. And only he is responsible for this state of affairs. He is his own deceiver. Just as in other-deception, the deceived, R, believes (at least implicitly) that the deceiver, S, does not believe p; otherwise, R could not be deceived by S into disbelieving p (say, by believing not-p). Further, S could not be deceiving R in this way if S was not aware of p's truth; however, this condition of being the deceiver does <u>not</u> require that S has paused to entertain in mind his belief that p with internal assent, much less external acknowledgment. He could be a thoroughly thoughtless, casual, or habitual

---

11   Champlin, "Self-Deception: Reflexive Dilemma," p. 284.

liar; being effective at his job, though, he never asserts the truth which he aims to keep hidden. Therefore, our analysis of self-deception parallels what can often happen in other-deception, and in both cases the deceived can be brought to false belief and assent, while the deceiver does not entertain in mind that he believes something or assents to it. The difference, of course, is that in other-deception (of a strong kind) the deceiver would and could assent (inwardly) to his believing of p, whereas in self-deception S could not do so and preserve the deception. However, belief without explicit assent is still belief, and thus self- deception is still genuine deception. It is well to remember here that our analysis of self-deception proceeds by way of addition, not subtraction. S believes that p (as the affective symptoms indicate to the cautious observer), but he <u>adds</u> to it an incompatible belief (that he does not believe that p); this added belief of a proposition which is preferable to S has behavioral implications which effectively cancel or block the assent he might give ,to p and the awareness that he believes p.

From here it is a relatively simple matter to decide whether the self-deceiver is aware of his incompatible beliefs or the deceived nature (falsity) of his acknowledged belief (that he does not believe p). First, it is necessary to self-deception that S appreciate the <u>incompatibility</u> of p and not-p (or of believing p and believing that he does not believe p).[12] If he was not aware of the incompatible nature of these pairs he would not be deceived, but merely confused, stupid, illogical, etc. Furthermore, if S did not recognize that his preferred belief was incompatible with believing the dreaded proposition, then he would not be motivated to rationalize the evidence. In strong self-deception S believes p, believes that he does not believe p, and is aware that <u>to hold such a pair</u> would be incompatible or irrational. However, he is not aware that he himself <u>does</u> hold to such an incompatible pair of beliefs, for the simple reason that he does not believe (is not aware) that he believes p (as discussed above). Logic prevents both beliefs from being true, but not from being held. Since S is not aware that he in fact holds both beliefs, S is not aware of an incompatibility within his belief system--an incompatibility which he <u>would</u> recognize as illogical.

In saying this we would differ with Rorty, who maintains that the self-deceiver recognizes that he has incompatible beliefs, recognizes their incompatibility, and thinks that there is some strategy for reconciling these beliefs.[13] In the first place, on this account S takes these beliefs as only apparently contradictory, in which case he views his beliefs as coherent--

---

12   Cf. Lerner, "Emotions of Self-Deception," p. 130; Bruce, "Investigation into Self-Deception," pp. 18-24.

13   Rorty, "Belief and Self-Deception," pp. 393-397.

which is contrary not only to Rorty's portrayal elsewhere, but also to the need for self-deception to be a conflict state. Moreover, if the self-deceiver were to recognize that he holds incompatible beliefs, as Rorty says, then it is clear that the self-deception would be punctured. It would then reduce to a case of admitted irrationality, or vacillation, etc. One cannot acknowledge the incompatibility of his beliefs and remain in self-deception; he must keep his belief that p hidden from view, so that no conflict is apparent to him. To be aware of the existence of such a conflict (the incompatible beliefs) would mean that S sees that one of his beliefs is false, and thereby S would not be deceived about it any longer. As Sartre was earlier quoted as saying, the lie would fall back and collapse beneath his look; it would be ruined from behind by his consciousness of lying to himself.[14] Thus through his rationalizing efforts S does not spell out his belief that p; he makes himself unaware of his belief by promoting the incompatible belief that he does not believe p. He does not recognize the truth about himself--that he holds incompatible beliefs and is thus being somewhat incoherent in his behavior. The incompatible beliefs are obvious to others (who observe the full range of S's affective and verbal behavior), but the idea is resisted by S himself. By obscuring his belief that p, although remaining aware of p's evidenced nature, S can simultaneously keep the fact that he holds incompatible beliefs out of immediate awareness; he can avoid the evidence for them, look away from them, and keep them hidden from himself.

Our next question, then, is whether the self-deceiver is aware that his professed belief that he does not believe p (or that not-p) is false. The answer is no, but he should. Herein lies the peculiarity of self-deception and its disturbing character. S has it within his power to undeceive himself--to become aware of the falsity of his professed belief; but he will not. We ask, is S aware that his belie that he does not believe p is false? If this is asking whether S entertains and assents to its falsity, then clearly the answer is no. By our analysis, S professes this belief with sincerity. If the question is asking whether S recognizes (is aware that, believes) that the proposition he assents to (that he does not believe that p) is false, then the answer is that he should. The reason for this is that S believes that p, and being a rational person (i.e., recognizing obvious contradictions, knowing that statements which are incompatible with the truth are false, and knowing that one should not believe falsities) he should not believe that he does not believe that p, or should not fail to see that such a second-order belief (as well as its first-

---

14   Szabados, "Rorty on Belief and Self-Deception," pp. 468-470; cf. Lerner, "Emotions of Self-Deception," p. 144; Gardiner, "Error, Faith, and Self-Deception," pp. 225-226, 238-240.

order concomitant of believing not-p) is false. If S could rid himself of his truth-distorting motive, see his rationalization for what it is, and observe the various affective symptoms of belief in his behavior, then he could see that he believes p. This would put the obvious lie to his belief that he does not believe p. Being aware that p is true, S should rationally see that not-p is false. If S believed, as he should, that he believes p, then he could not (as a rational being) lie to himself about it. However, because S is in the disturbing state of self-deception, he prevents himself from believing correctly about himself, his rationality seems to fail him, and he makes himself conspicuously unaware that his professed belief is contrary to the evidence (is deceived or false).

Against such considerations Hamlyn suggests that there are true cases of other-deception (and thus possibly self-deception) where the deceived person knows the truth (and by implication the falsity of what is portrayed to him or her by another). That is, the ignorance of the deceived person is as-it-were ignorance for he or she is not really made ignorant by the attempted deceiver. He gives an example of a husband deceiving his wife about a love affair, where she knows of the mistress anyway and he knows that she knows of the mistress.[15] This is an intriguing illustration, of course, but it does not support anything contrary to what we have said above. In the first place, if the ignorance of the deceived (viz., the wife) is only an as-it-were ignorance, then she is only <u>pretending</u> not to know; she does actually know, and thereby is not deceived. (She might, however, move <u>into self</u>-deception of the sort we have analyzed above.) The as-it-were knowledge of the husband, whereby he behaves as though his wife would be as unaffected as if she were not aware of his affair, is not genuine deception either; if he wishes to give her the impression that this love affair really makes no difference to their continued relation and believes it himself--then we simply have a case of mistaken or biased belief. Further, Hamlyn's example wrongly assumes that the husband is honest and dishonest about the same thing, when in fact there are two issues at stake: whether the husband has been unfaithful, and whether the unfaithfulness makes any difference in his relation to his wife (as long as it is not openly flaunted, even though she is aware of it). The husband may be honest about the first, but he is not nearly honest enough about the second. So because of such an equivocation it would not seem that Hamlyn has succeeded in giving us an illustration of deception where the deceived is "in the know." Moreover, this would not be strong deception anyway, but

---

15   Hamlyn, "Self-Deception," pp. 45-50; cf. critical interactions by Mounce, "Self-Deception," passim; Szabados, "Self-Deception," p. 62; Champlin, "Self-Deception: Reflexive Dilemma," pp. 286-287.

only deception in the sense of attempted deceit or dishonesty. And finally, the example would be quite difficult to transfer over to the case of self-deception since it rests on the assumption that the "deceived" wife actually knows the truth about her husband's affair without his telling her about it (e.g., by spying, perhaps). But what self-spying could be, or how the self-deceiver could actually know something about himself which is not ventured into the open, is not at all clear. We must conclude that, as discussed above, the deceived person does not believe that his professed belief is false; he is not aware of the falsity as he should be. The information is available to the self-deceiver by which he could acknowledge that his second-order belief is deceived or false, but he will not avail himself of it.

Our final question has to do with the self-deceiver's awareness or lack of awareness of his motive for the activity of rationalizing the evidence. Someone might attempt to create a paradox along these lines. The present analysis appears to portray the self-deceiver as both sincere and insincere at the same time. On the basis of reworked evidence S comes to believe that he does not believe that p; this belief, and S's assent to it, are deemed sincere--he acts in accordance with it, expresses himself with conviction, etc. Yet on the other hand S's behavior regarding the evidence (which, recall, is adverse to his cherished belief) is said to be motivated by a desire to avoid the psychic pain threatened by the truth to which this evidence points (and of which S is aware). Thus S's handling of the evidence--on which his sincere belief rests--is quite insincere. Hence the paradox.[16] By way of reply, let us note the mistaken assumption on which this line of argument rests. It is apparently thought that if S <u>acts on his desire </u>to avoid the discomfort caused by his awareness of p, then S must be aware that he is acting on that desire in these circumstances. If his activity (viz., rationalizing the evidence) is explained by an 'in order to' clause, it is thought that S must be conscious of the Y that X is unto (or for-which sake X is performed)--that Y is clearly before his mind. However, such assumptions are not true. If S's bringing himself to his cherished belief because it is in his interest (i.e., because of his desires) were to be construed as S's <u>reason</u> for acting in a rationalizing fashion toward the evidence, then of course he would necessarily be aware of his desire; it would be the grounds or justification he would offer for his activity. However, as explained previously, we are not to understand S's "motive" as his own reason for rationalizing. That S's rationalizing is motivated is to say that S has a desire which such activity would satisfy. And with respect to a desire for Y, it is not

---

16   Shea poses and responds to this dilemma in defending his own thesis: "Self-Deception," pp. 115ff.

the case that S must have Y clearly entertained before his mind while that desire is operative. Much of the time his mind will be directly taken up with other matters without the object of his desire appearing to him in mind or impinging on him as some kind of felt impulse. If I go to a wedding because of a desire to be courteous to a relative, there is no necessity for a preceding, conscious realization that going to the wedding would result in (or bring about) the being courteous to this relative. And if I, out of a desire to make a patio for my family, go to the builder's emporium to purchase cement mix, there is no necessity that I be entertaining explicit thoughts of that patio the whole time that I am driving, shopping, paying, returning, etc. Desires are not, even when aroused, <u>ipso</u> <u>facto</u> internally perceived by their possessors. Behavior can be motivated by desires of which the agent is not currently aware (entertaining in mind). Furthermore, if the entertaining of such desires would prove somehow uncomfortable to the agent, he may very well be prevented from such an entertainment all together. We should not think that just because S does X to satisfy a desire for Y he does X with Y clearly before his mind as an objective to be gained, for then it would make no sense for S to ask himself if he did X because of his desire for Y. Yet people do sometimes ask themselves such questions sensibly, and we take correct answers to them as the gaining of significant items of self-knowledge. Thus one can act out of a desire and not be aware that he is acting out of that desire.

Motives are not reasons of which S is necessarily aware (entertaining in mind), but nor are they to be construed as causes which force S's actions and of which he could not be directly aware. Motives do not <u>make</u> S act in a certain way, for it is quite conceivable that S could come to recognize that some unnoticed motive (say, that of envy) had been recently <u>unsuccessful</u> in getting him to act in a certain way; this is different from saying that the motive was non-existent. And when S does discern the character of his motive--which he is always free to do--he may even rid himself of it completely. Motives like envy of one's rivals can influence a person's actions, even though that person is not entertaining them in mind, or even though he disbelieves that he has them.- However, this is not to say that one's motives are inaccessible inner workings of the soul. Like cobwebs on the ceiling, motives are things which may become quite obvious once we notice them (come to believe that we have them). However, in self-deception S may be free to discern the character of that motive which accounts for his rationalizing, manipulation, and stifling of evidence, but he is preoccupied. He desires not to become aware of that shameful desire to avoid the truth. By his deception he is hiding

such a motive away from his conscious introspection.[17]

Therefore, with respect to the self-deceiver's motive, we have said that it (like other motives or desires) <u>need</u> not be entertained in his mind while it is operative. However, one is always free to discern his motives; they are not inaccessible. Yet in the case of self-deception a person <u>does</u> refrain from becoming aware of his motive for rationalizing the evidence. Indeed, if he did not do so, his effort at deceiving himself would be jeopardized. We do not say that the effort at self-deception would be automatically unsuccessful. S's awareness that he is acting on a desire to believe contrary to the evidence would not preclude self-deception. To be sure, believing that his cherished belief is not fully defensible—is rooted in a desire to avoid the painful truth, which desire is satisfied through rationalization—could threaten and endanger S's state of self-deception. But it would not necessarily terminate his self-deception, for at that point he could <u>extend</u> his self-deception (i.e., could redouble his efforts) to cover his belief about this motivation. That is, S could maintain his self-deception about believing p by creating a web of self-deception, so that he is now additionally self-deceived about his motives regarding the disbelief that p.[18] As Szabados remarks, "Self-Deception tends to breed further self-deception."[19] And this suggestion brings us to the topic and resolution of the next section's difficulty.

6.3 Can a Person Deceive Himself on Purpose?

We have just suggested that a self-deceiver is free to discern the character of his motive but does not do so. When and if he did come to recognize his motive (became aware of it, believed that he had it) he could, we have said, redouble his efforts and deceive himself about the motive as well. This all suggests that self-deception is an action that S can make up his mind to do. And of course that impression is reinforced when we recall that self-deception can be, and is, modeled on other-deception (wherein S can consciously decide to keep the truth from R). We have portrayed self-deception as involving a manipulation of oneself--manufacturing a cherished belief by means of rationalizing the evidence. The question naturally arises, however, whether this activity can be done intentionally.

---

17   That we are not necessarily aware of our desires is also noted by Valberg, "Rationality and Self-Deception," p. 180; that we are free to discern our motives if we were not preoccupied with other motives is also noted by Szabados, "Wishful Thinking and Self-Deception," p. 205.

18   Shea, "Self-Deception," pp. 170-171.

19   Szabados, "Wishful Thinking and Self-Deception," p. 205.

And if it is, does that not mean that S is aware of its character? And in that case, self-deception would seem to be impossible. The intended victim of the deception would be conscious of the deceptive character of his attempted deceiver's actions and words, and thus he could not really be misled after all. If the self-deceiver believes that his second-order belief that he does not believe p is actually false, then he does not believe it after all. Thus we are back to the fundamental question of whether genuine self-deception is even possible.

Can men <u>try</u> to deceive themselves? Can they do it <u>on purpose</u>? These are the sorts of questions with which we are ultimately concerned here. In addressing them we look upon self-deception as something an agent <u>does</u>; it is something he either brings about or achieves, rather than being something done to him. Accordingly, self-deception <u>should</u> be able to be done intentionally, and this observation could easily open the door to extended discussion of debates still raging in the philosophy of mind and action. However, the numerous questions that can be raised must, and can, be put aside for present purposes. The concepts of intention, motive, and reason are all closely tied to teleology, and there are great difficulties involved in any attempt to give an adequate account of the nature and interconnection of these concepts--often arising from their blurred edges and borderline cases. In the present situation it is especially the <u>purposiveness</u> of self-deception that is questioned, and we can attempt an answer without relying on subtle or specialized doctrines pertaining to related notions. We will here grant what should be unobjectionable to most schools of thought, namely: intentional actions are such that they <u>can</u> be taken "on purpose" and are in some sense under deliberate control. Of course, men do sometimes act for no reason at all, without purpose or aim; many things we do are done inadvertently, unknowingly, accidentally, or by reflex. And even when we decide to do something, that decision can be made without deliberation and resolution (as when there is no competing alternative which appeals to the agent, or when he decides on momentary impulse, or when he does not occurrently form an explicit intention). Moreover, one can distinguish between the question of an action's intentional character and its voluntary character: i.e. we may do things intentionally but not voluntarily (e.g., under duress, provocation, irresistible compulsion), and something may be done voluntarily but not intentionally (e.g., hitch-hiking a ride in what turns out .to be the wrong direction). And a commitment to do X here and now--a volition, if we like-- can be discriminated from an action (since one does not will to will to do X), a trying (since it may not involve doing something which you think will develop into doing X), a choosing (since it may not be over against alternatives), and

a deciding (since it may not be the culmination of process of deliberation). Therefore, being fully aware that to answer the relevant question before us is not simultaneously to answer questions over related matters, we can stay to the center of the field and simply inquire whether self-deception is the kind of thing a person <u>could</u> do <u>purposely</u>.

It is a common thing to speak of human activity in a teleological fashion: men are said to do things in order to achieve some aim or purpose. And when it is alleged that S did X intentionally, this can ordinarily mean that X was done on purpose (or for some reason). It is important to note, though, that the same activity (event, happening, movement) may be intentional under one description, but not under another. For instance, imagine that an arm is raised. This could be described as 'attracting the speaker's attention'(intentional), 'reflex to an electrical shock' or 'to cracking one's funny-bone' (unintentional), 'making a bid at an auction' (intentional), 'being pulled up the cliff' (unintentional), etc. Furthermore, actions and movements can bear a one-many relationship to each other. One movement may amount to the accomplishment of many intended actions (e.g., guiding a pen over a paper may be not only the signing of a name, but the issuing of a decree, the ritual of a monarch, etc.). And some intended action may call for many movements (e.g., reading a book). According to some theorists, the very criterion for individuating "the same action" is that the movements in question all have the same point or intention. The question before us is whether self-deception can be an intended action. Can the activity of rationalizing evidence so as to bring oneself to a false belief be described as intentional, as being done purposely?

There are those who state that self-deception <u>need not</u> be purposeful, and some who claim that it <u>cannot</u> be. In the latter group we find those who claim that the self-deceiver does not purpose to be unconscious of the truth, for such a policy on his part would lead to a paradox; by his purpose S would be conscious of the very thing that he decides to be unconscious of. Therefore, self-deception is not intentional, but comes about by forces beyond S's control. Such a perspective, however, simultaneously moves us away from self-deception into the realm of mental disorder, occult coercion, brainwashing, or some such thing. Self-deception is no longer an action of S's, but is rather something that happens to him; it is thus hard to specify how this would be self-deception. Others have argued that self-deception is the unintentional result of an emotion.[20] S becomes aware of a truth, which in turn produces a negative emotion in him; this negative emotion is a disposition

---

20   E.g., Lerner, "Emotions and Self-Deception," chapter 6.

which results in his ignoring of the truth and manufacturing a belief through rationalization. Thus one cannot choose to deceive himself; self-deception is not really purposive, but is a kind of inner reflex--a mental flinching from an unpleasant truth (just like the natural reaction of the body in cringing, blinking, turning away from feared or revolting things). Self-deception serves a function, but not a purpose (like the human heart); it satisfies a need of S's. The reason offered for saying that self-deception <u>cannot</u> be intentional is that it would then require a regress of intentions (each intention having its own intention, etc.). Because emotions are dispositions which do not call for explanatory intentions, they would escape such a regress and could function as the cause of self-deception. But against this proposal it might be said that, because S would still have to prevent himself from becoming aware of his emotion (since awareness of it would make him aware of the truth, and without some effort at prevention S would always be able to say what he is feeling), a second self-deception is required--which <u>ex hypothesi</u> results from an emotion, which itself would have to be covered by a further self-deception, etc. It seems, then, that the regress has not been avoided; it has simply become a regress of emotions <u>ad infinitum</u>. Moreover, the present proposal is hard pressed to explain how the various elements which constitute self-deception could merely coincidentally come together and produce the right outcome; such a coherence, unity, cohesiveness, or cooperation of different things might easily appear to require some unifying intention, purpose, or conspiracy on the part of an agent. In answer to these two criticisms Lerner says: (1) unlike intentions, emotions can have a double object, thereby allowing for a single emotion to cause both the deception about the truth and the deception about the emotion itself; and (2) self-deception only appears purposive, in the same way that a heart might appear purposive, and neither requires us to hypothesize an intention behind them which is fulfilled by its occurrence.

But problems remain. The claim that a human heart might appear purposive can be defeated by noting that purpose is evident when <u>independent</u> entities are directively correlated, and this is not what we have in the case of a functioning heart (i.e., the pumped blood, infused arteries, opening valves, etc., would not be such independent of each other)--whereas the condition is met by self-deception (i.e., false beliefs, adverse evidence, rationalization, etc. can be what they are independent of the others). Secondly, the charge of false analogy is more than feasible with respect to Lerner's illustration of bodily flinching; what is the nature of a <u>mental </u>flinching? Are there degrees of it (as in the case of the body, where we can move from blinking, to cocking the head, to ducking)? What view of the "mind" does this analogy require, etc.? Thirdly, as is quite evident, Lerner's escape from the infinite regress

of emotions depends on his thesis that emotions are dispositions. Various problems with that thesis would naturally weaken his defense, and thereby his insistence that self-deception must originate with an emotion. However, for whatever problems there may be with the thesis under consideration, we can allow that perhaps self-deception in <u>some</u> cases does stem from an emotion. What is of more direct relevance to us is Lerner's stronger claim that self-deception <u>must</u> stem from an emotion, for an intentional self-deception cannot escape an infinite regress. This claim will be considered in the next section and hopefully defeated.

Apart from the claim that self-deception <u>cannot</u> be intentional, some have written that self-deception <u>need not</u> be intentional. For instance, without meaning to do so we can slip into self-deception unguardedly, just as we can slip into sleep without intending so. Adherence to self-deceiving actions can become habitual or second-nature to us, thereby dispensing with conscious intentions.[21] Just like S can mislead R from the truth without purposing to do so, so S can unintentionally mislead himself from the truth; there are many kinds of deception in other-deception (including unwitting deception), and that variety need not be denied to self-deception.[22] Self-deception is not <u>wholly</u> made up of intentional elements, for belief is necessarily involved in it; people cannot simply decide to believe as they choose. Thus self-deception is not fully intentional.[23] There are difficulties with some of these suggestions. The analogy suggested in "slipping" into self-deception needs to be worked out with some clarity. Even habitual self-deception started with a first instance, wherein S might have intended to deceive himself. Champlin's interesting distinction between 'deceive' (<u>tout</u> <u>court</u>) and 'deceive about (that, by, etc.)' helps us to see the various kinds of deception of which we may speak, including unintentional misleading; however, when he claims that one can practice deceit upon another person, but not ever upon himself (i.e., there can never be strong self-deception), we look in vain for the supporting grounds. Finally, Szabados may be correct in noting that self-deception is not wholly comprised of intentional actions; however, the fact remains that the intentional components allow for deliberate or purposive engagement in self-deception, especially when we note that the non-intentional beliefs can result from the intentional actions (viz., rationalization, manipulation of the evidence, distorted focusing). But difficulties aside, we can again grant that perhaps some instances of self-deception are not intentional. Furthermore, we can take note that many of the above perspectives still allow for holding

---

21   E.g., Bruce, "Investigation of Self-Deception," p. 115.
22   Champlin, "Self-Deception: Reflexive Dilemma," pp. 292-294.
23   Szabados, "Morality of Self-Deception," p. 32.

the self-deceiver responsible or blameworthy for his self-deception. People can be held responsible for their mistaken beliefs when they are due to neglect of the steps and precautions necessary for avoiding error. The plea "S did not intend to deceive himself" may mitigate his blame, but S is still open to the lesser criticism, "Well, he should not have been so careless." His -continuance in self-deception or his beliefs may be criticized for insufficient attention on S's part. One can, it is argued, discover that an emotion is controlling him and then terminate its dominance.[24] He is free to discern the character of his motivated manipulation of the evidence and thereby cease to engage in it further.[25] The self-deceiver is being dishonest with himself, even if unintentionally.[26] So if he would be more attentive, more careful, more honest, more self-examining, the self-deceiver could avoid his condition. In that sense even unintentional self-deception might be deemed culpable.

However, apart from unintentional yet culpable cases of self-deception, the question remains whether self-deception <u>can</u> be engaged on purpose. Even if not all cases of self-deception are intentional, are any? Might the self-deceiver be blamed, not only for not coming out of his deception, but for getting there in the first place? Is strong self-deception (self-<u>deceit</u>) possible? While some current writers have backed away from this, saying that self-deception is too subtle or ingenious to be deliberate,[27] others have been bold to assert that self-deception is the purposive refusal to face the truth or interpret the evidence realistically.[28] Reasons to endorse the second viewpoint were advanced in the last chapter. We may recall here that in many cases self-deceivers are held responsible for doing something which is deemed their own fault to begin with. Some of the separate elements of self-deception as a process are skills (e.g., avoiding, manipulating) and as such could be subsumed as means to an end under an over-riding intention to believe something (contrary to one's awareness of the truth). Furthermore, usually self-deception must stem from a strong enough purpose (e.g., to repress awareness of the truth, to avoid evidence contrary to one's cherished belief) that one is not engaged merely in sham, pretending, inattentiveness, intellectual laziness, etc.[29] Finally, in considering confessions of self-deception, or in reflecting on one's own experiences of it, we can find a willingness to reconstruct the vents leading up to self-deception in a way which includes

24 Lerner, "Emotions of Self-Deception," pp. 180-182.
25 Szabados, "Morality of Self-Deception," pp. 32-33.
26 Champlin, "Self-Deception: Reflexive Dilemma," p. 297.
27 E.g., Saunders, "Paradox of Self-Deception," pp. 567-568.
28 E.g., Fingarette, *Self-Deception*, pp. 28-29, 146-149.
29 Cf. Hauerwas and Burrell, "Self-Deception and Autobiography," p. 101.

something like a resolve (e.g., "I shall bring myself to a contrary opinion, and find a way to justify it"). Therefore, there are some reasons to think quite naturally that self-deception could be taken as an intentional action in some cases. The issue boils down to whether there are telling criticisms of that as a possibility. And the central criticism of this kind maintains that if S deceives himself on purpose, then he is aware that what he comes to believe is a lie; since it is impossible to believe what you take to be a lie, intentional self-deception is impossible.[30]

In response, some would defend the possibility of intentional self-deception by maintaining that intentions can be something of which their possessors are unconscious. Since S can intentionally do something without being aware of his intention, deceiving oneself on purpose need not be excluded as a possibility. In some cases the assertion that there are unconscious intentions envisions an extraordinary kind of intention. For instance, one might speak of dynamically unconscious intention as a "twilight purposiveness."[31] Or one might think of an intention that a rational agent should have, given his behavior, but does not--as when S does not intend the implications of his intention (e.g., be intends to smoke, and he is aware that smoking brings about cancer, but he does not apparently intend to get cancer).[32] However, such suggestions only underscore that the description of self-deception is problematic or that human behavior is not in all respects rational. A more helpful explanation of unconscious intentions would point out that doing things on purpose need not reflect full-fledged reflection and conscious deliberation. We are all familiar with behavior that is informed by distinctive purposes, and yet the agent gives no indication of being aware of that purpose; for instance, S may have become involved in a complicated goal-oriented activity (e.g., walking out to his car and starting the engine) and forgotten completely why. Or better, we can note that many things are done intentionally without any noticeable preceding conscious episode of choosing, deciding, purposing, resolving, etc.--for instance striking the keys in typing, moving your legs in walking, turning the pages in reading, scratching an itch. Such things are done "straight away" or immediately, without a conscious choice or deliberation as preliminaries. Instead, it would seem that intentional actions are sometimes--but not always--preceded by the formation of a disposition or readiness to act appropriately under proper conditions and stimuli And as a disposition an intention may be formed as effortlessly and as unconsciously as any proneness, tendency, habit, skill, belief, etc.

30 E.g., Drengson, "Self-Deception," pp. 92-93.
31 E.g., Saunders, "Paradox of Self-Deception," p. 565.
32 E.g., Jordan, "On Comprehending Free Will," p. 192.

However, not only are there objections to be encountered in a dispositional understanding of intention (and thus relying simply on it to resolve the perplexity surrounding intentional self-deception would be too easy and unconvincing to critics), the fact remains that a teleological description or understanding of intention is still common and not incompatible with a causal (dispositional) explanation of behavior according to some dispositionalists. Others, of course, do think that a causal explanation in terms of dispositions is unnecessary or incompatible with a teleological one and will therefore still have to be satisfied regarding the possibility of intentional self-deception.[33]

Accordingly let us press on, and consider the possibility of intentional self-deception where the kind of intention involved implies that the agent is aware of it. This will certainly be the strongest case of self-deception, and if it can be shown possible, then the paradox of self-deception should be resolved to our complete satisfaction. There are reasons which support the idea that when S does X on purpose, he is aware of the nature of X as it is described. Of course, there is a sense in which X can be an "action" of S's even when he did not intend it under that description (e.g., he can shoot the president even when he did not realize that it was the president that he shot); S can bring about Y even when he did not realize that doing X would accomplish Y (e.g., he can expose the photographic film even though he did not intend to do so by opening the darkroom door). But when we consider an <u>intentional</u> action it is crucial to ascertain the way in which the <u>agent</u> would (have) <u>describe</u>(d) it. Without meaning to take a position on the compatibility or incompatibility of a causal explanation we can observe that in cases of purposive action explaining and describing can amount to the same thing. The movement of extending one's hand can be explained as reaching for the salt (i.e., 'S extended his hand in order to pick up the salt across the table') and described as reaching for the salt (i.e., 'The action of S was that of reaching for the salt'). It is appropriate (and some would say necessary) to give intentional actions such a teleological description-explanation, describing a movement (happening, event) by giving its point or rationale for some agent (appealing to his ends or goals). It happens that multiple descriptions of some movements are available, as when doing X can be designated 'signing his name' or as 'enlisting in the Marines.' Whether these are equally appropriate and true descriptions of some movement will normally depend on the agent's purpose--how <u>he</u> would describe the movement (e.g., simply as 'making a noise' which happens to result in annoying a neighbor, or as 'annoying a neighbor'). Accordingly, an action or a description of a movement may be

---

33 E.g. Richard Taylor, *Action and Purpose* (Englewood Cliffs: Prentice-Hall, 1966).

intentional under one description, but not under another. When a third party describes something as 'annoying a neighbor' we may not know whether this is the result of an action or the intended action itself; the testimony of the agent (if sincere) would be the best way to be certain of the matter. An agent's intention is basically what he takes himself to be doing or about to do. It is constitutive of the described action, a central part of what makes the movement <u>this</u> (e.g., 'annoying') rather than <u>that</u> (e.g., 'making a noise'). If the identity of a man's intention is fundamentally determined by the way in which he conceives and will describe it, then having an intention is (in part at least) a matter of envisaging an action or state of affairs, being ready to do what one believes (hopes) will bring about its realization. Thus when S does X intentionally, we may say that he does so because X has a place in his "plan of action" (given the way in which S conceives of his situation). S does X purposefully when he conceives of it as a means for achieving a certain end or result--when S believes that X is a means to Y. If X is the raising of his hand, and Y is the shooing away of a fly, then this action was purposeful as long as S believed that-X was the means to an end, Y. As an end or goal of his activity, Y is something which S can be said to try and accomplish by X. And to say that S tried to do Y, or that he deliberately did Y, or that he did Y on purpose, would all seem to depend on S seeing his movement in light of a certain description (e.g., 'shooing away a fly') or as a means to an end (e.g., S takes the raising of his arm as a way to accomplish something else).

The foregoing portrayal of intentional action has been put in a strong enough way to show why some writers have held that self-deception could not be done on purpose.[34] If S purposes or tries to deceive someone, then he accordingly describes his actions in that way or conceives of them as having the aim of misleading someone from the truth. Such a description (or means-end conception) would not be correct unless S believed contrary to what he tries to get his victim to believe, and unless S believed that what he tries to get his victim to believe is mistaken. For instance, if S did not believe that his testimony was false, he could hardly be rightly accused of perjury. If he really believed that he was the person named on the check, he would not be guilty of forgery. That is, in order for S to deceive someone intentionally, he must conceive of his actions or words as misleading from the truth; he must be aware of the character of his deeds (i.e., could describe them as 'deceptive). Consequently, if S <u>intentionally</u> deceives someone, he could not be said to be taken in by the deception himself; he cannot believe that a proposition is false, and believe as well that it is true (short of utter irrationality, anyway). In that

---

34   E.g., Lerner, "Emotions of Self-Deception," pp. 162, 176.

case intentional self-deception would seem to be an impossibility in practical outworking. It would call for S being aware that his words (or actions) were deceptive, and yet being taken in by them. In intentional self- deception S would have to be aware that he is resisting the truth (i.e., be aware that it is the <u>truth</u> which he aims to distort in order to believe contrary to it).

Given such a strongly self-conscious sense of intention, can S deceive himself on purpose? Can he try to do so? In defense of that possibility one might (again) resort to a new sense being given the words which describe intentional self-deception; he might speak of an extraordinary "twilight" purposiveness which is reminiscent of full-blown purposiveness, but inaccessible to full consciousness. However, the futility of such an approach has been indicated already; it simply linguistically masks the persisting problem. A different, more bold, approach would be to accept the implication that, if S is intentionally deceiving himself, then he is aware that he is telling himself a lie (resisting the truth, etc.). Hamlyn gives an initial impression that he thinks of self-deception as being compatible with the agent knowing what he is up to.[35] When a person admits to himself how he stands on a given issue, but behaves (and claims) as if it made no difference, he is using pseudo-rational devices to isolate his true feelings about the matter from his awareness of the truth (perhaps even his super-honesty about the truth). Hamlyn rightly sees this as one form of self-deception. And in the case of intentional self-deception, says Hamlyn, one knows that he knows about himself--apparently meaning that S knows (in virtue of his intention) that he knows about his feelings (since in self-deception he knows what he isolates from himself with an as-it-were ignorance). But when Hamlyn then says that the self-deceiver at this point makes himself unconscious of his knowledge (diverts his attention from it)--thereby preserving the state of self-deception, as I understand him--it becomes clear that Hamlyn is not after all defending the possibility of a self-deceiver being aware that he is misleading himself from the truth. Consciousness of that fact is after all covered over or removed by rationalization (concentrating one's attention in a way which excludes certain things). As far as I know, the only writer who goes so far as to claim that a person can intentionally deceive himself in the sense of being aware that he is telling himself lies is King-Farlow.[36] King-Farlow contends that one does in fact fool himself by consciously chosen lies, and he shows this by means of an apparent illustration. This illustration is intended to demonstrate that S can deceive himself (i.e., knowing that

---

35   Hamlyn, "Self-Deception," pp. 46-52.
36   King-Farlow, "Self-Deceivers and Sartrian Seducers," pp. 132-133.

what he gets himself to believe is false) into doing something he would never otherwise do. For example, S is quite conscious, that he hates the thought of finishing his Christmas correspondence, and he would never do so unless he deceived himself (prior to each letter) that it would be a pleasant duty. So S consciously tells himself lies to this effect, all along knowing just how false his statements are in order to assure his continued writing. We are probably all familiar with such circumstances from personal experience. The question is whether King-Farlow has accurately and fully described such a common experience. If he here attempts to prove his thesis by an <u>unfamiliar</u> experience, then we would naturally fall into question-begging over the description of it or the imposition of preconceived opinions as to what is possible (e.g., "this event has not been described correctly because such things just cannot happen"). And surely an appeal to an extraordinary experience could not be expected to have the argumentative strength which King-Farlow assigns it, especially noting his brief and undefensive rehearsal of it. But if he offers us a common experience with which we are familiar, and if this is presented as evidence that men actually do believe what they recognize as lies, then King-Farlow has exaggerated or misconceived the situation. First, it is a dubious assumption that such a duty (e.g., writing tedious Christmas letters) would in fact never be followed unless S believed that it were pleasant activity. Second, the things which S says to himself in such situations are not conceived as falsehoods, but rather as attempts to concentrate on and emphasize aspects of the activity which offer the only pleasure involved, albeit slim (e.g., "I will remember the good time we had together back when. . ." or "Finishing this last letter will be rewarded with the feeling of relief, commendation for a duty performed, etc."). Third, even when S actually tells himself downright lies (and not merely focuses on insignificant areas of truth) and goes through with the letter-writing, there are more plausible interpretations of what he is doing than that of self-deception. In particular, this situation is more naturally seen as a man involved in play-acting, in pretending in responding as though he believed the duty pleasant (even though it is not). Surely in an actual play (or motion picture, etc.) genuine actions can be taken which are indistinguishable from similar ones "in real life" (especially with respect to their effects); for instance, when the actor drinks his glass of water in a play, the glass is emptied just as surely as if he drank it during intermission. And if a man writes a letter, pretending he is someone who enjoys that sort of thing (i.e., pretending to himself), then the letter is no less actually written, even though it was produced through playacting. It is not insignificant that we deem a person who writes this kind of note (viz., saying thank-you when he is not really grateful) a "hypocrite" (from the Greek word for 'actor'). Finally,

there is nothing in King-Farlow's illustrations to support the supposition that S <u>simultaneously</u> believes a statement to be true (as deceived) and false (as deceiver)--i.e., that S recognizes that he believes an unpleasant duty to be pleasant. If anything, what we actually would have is a case of S vacillating between different attitudes on the activity (e.g., resisting the writing, but then continuing, etc.). Therefore, it does not appear that the perplexity over self-conscious self-deception (due to it being purposeful) has been resolved by the claim that men can believe their consciously chosen lies.

6.4 Intentional Self-Deception as Self-Covering

In the case of strong self-deception S deceives himself on purpose; it is intentional, and accordingly S would be aware of the character of his actions. But if S believes that he is deceiving himself, then he could not be taken in by his efforts. He would (and not just should) believe that the proposition he comes to believe is false, and in that case could hardly be said to believe it then. Intentional self-deception thus appears to be impossible. In posing this perplexity we have attempted to use an understanding of intention which, whether the one to be endorsed or not after the resolution of various debates in the philosophy of mind and action, is most unfavorable to the possibility of successful, strong deception. Gardiner may very well be correct in pointing out that we are often mistaken in thinking of purpose as full-fledged conscious deliberation, and that purposeful self-deception may not entail fully reflective behavior,[37] but our discussion cannot end on that note unless we can demonstrate that all self-deception is in fact never purposeful in any stronger sense. I think we should explore further. There has been no want of philosophers who maintain that deception, including self-deception, involves intentional behavior and candor.[38] And as Champlin observes, "when it is said that self-deception is an intentional activity I think the subsequent discussion sometimes shows that the real issue is not <u>intention</u> at all but <u>awareness</u>."[39] If it should be shown true that self-deception is possible when there is full-fledged conscious deliberation--and thereby a full awareness of what one is up to--then the strongest form of self-deception will have been defended and the apparent paradox removed from any and all forms of the phenomenon (even weaker versions). The problem is thus whether a person

---

37 Gardiner, "Error, Faith, and Self-Deception," p. 241.

38 Cf. Hamlyn, "Self-Deception, p. 18; Lerner, "Emotions of Self-Deception," p. 159; Shea, "Self-Deception," chapter 9; and for a discussion of Freud, Marx, Mannheim, Kierkegaard, Sartre, and Fingarette, see Shapiro, "Self-Deception," chapter 5.

39 Champlin, "Self-Deception: Reflexive Dilemma, p. 293.

can deceive himself when he is fully aware that this is his purpose.

However, the perplexity is not fatal to successful self-deception, even on these terms. We have noted above that a general test for the intentional character of some action is asking the agent how he would describe it--how he conceives of its explanation, purpose, or aim. Thus we can imagine ourselves asking the self-deceiver, "Why did you treat the evidence in that way?" And if he was involved in intentional self-deception, he would be able to answer "to escape recognizing the truth of p" (or something else indicating the recognized deceptive character of his behavior). However, although the self-deceiver may be able to give such an answer, he will be <u>unwilling</u> to do so. In the first place, there may be many answers available to such questions (viz., "Why did you X?"), all of which are equally true, but not all being equally relevant in various contexts. For instance, when there is a series of purposes involved, S may answer to why he was moving his arm in

a particular way by saying 'to operate the water pump' or 'to supply the house with water' or 'to poison the inhabitants' (since S knows the water is poisoned). Likewise, in self-deception S may be able to give as his purpose 'the distortion of the evidence,' but he may in fact give other, equally true answers (e.g., 'to explore all interpretive possibilities before making up my mind')--but answers which are not equally relevant in assessing his behavior for what it fully is. Moreover, there appear to be cases where S may be sincere in citing his reason for X (or his key reason for X) and still be mistaken. Although he could "normally" state in unmistaken words what his intention was, there are abnormal cases where he does not seem to do so. For instance, the agent may not state his intentions clearly to himself, being an inarticulate person.[40] Or more relevantly, the agent may be motivated in his mistaken answer about his intention, or so thoroughly flustered as to be unable to speak the truth explicitly (even to himself)--as when he is ashamed to acknowledge his real intention. For example, a guilty child caught in the act of disobeying his parents may be so flustered that, when asked "What do you think you are doing?" he cannot admit the truth but cannot get out a convincing alternative either. Or an adult may, without being insincere, falsely describe his action due to being ashamed of the truth about himself:

> There are abnormal cases where I may mistake my intention, as when I deceive myself because I would be ashamed to acknowledge my real one. 'I didn't mean to smack the baby's hand, I was only

---

40   R. L. Franklin, *Freewill and Determinism*. International Library of Philosophy and Scientific Method, ed. Ted Henerich (New York: Humanities Press, 1968), pp. 95-96, 101.

pushing it away from the jam.' I might then be brought to agree that I had 'really' intended to smack it in my annoyance, though I had not actually lied in denying it.[41]

"It is a notorious fact that people can often deceive themselves about their own reasons for acting."[42] A person may not want to confront and admit the real factors at work in his actions. Therefore, in one sense an agent may be in a privileged position for assessing his actions, but in another sense (which all must recognize who have later realized their own errors in describing their intentions) an agent <u>can</u> have great difficulties to overcome in order to describe his aims correctly. This is obviously the case when one is in the process of rationalizing the evidence or deceiving himself. We would not here expect the agent to give the correct answer to "Why are you treating the evidence in that manner?" even though the information is accessible to him. His answer can very well be mistaken, even though it is only later that he realizes the mistake himself (i.e., explicitly admits it to himself).

What this indicates, of course, is that the self-deceiver is also self-deceived about his purposes. When he intentionally deceives himself he is aware (believes) that his behavior is deceptive (or is aware that his handling of the evidence is rationalization), but he will not assent to it; be comes to believe something incompatible with his belief about his intentions, thereby blocking assent to them and not entertaining them in mind. He purposively distracts himself from recognizing his intention for what it is. He constructs an intricate account of his actions which--while out of full keeping with them--facilitates that behavior by distracting him from recognizing what he is up to. That is, S purposely hides his belief that p from himself <u>and</u> hides the <u>hiding</u> of his belief. The self-deceiver has an unacknowledged intention to hide his hiding from himself.[43] He has an end in view as he reacts to the evidence, but he will not acknowledge that purpose as his own. For example, we can imagine a middle-aged mother with one child, her only pride and joy in life. One day Mts. Jones receives a phone call from the principal of Johnny's school, who says that her son has been caught stealing lunch-money out of other students' desks. Johnny will be suspended from school for two weeks and kept on probation until the end of school. Mrs. Jones is stunned--nobody in her family has ever stooped to dishonesty before. She dreads the thought of seeing the principal, the thought of her son's punishment, the

---

41 Ibid., p. 95.

42 Jerome A. Shaffer, *Philosophy of Mind*. Foundations of Philosophy Series, ed. Elizabeth and Monroe Beardsley (Englewood Cliffs: Prentice-Hall, 1968), p. 103.

43 Wilshire, "Self, Body, and Self-Deception," pp. 422-423.

thought of living with the self-image of a mother with a delinquent child. She wants things to be as they were before the troubling phone call, when she believed that her Johnny was an honorable and honest son. She decides she will try to make herself believe those things again. Therefore she decides that nothing will be said of the incident at home, and she transfers Johnny to another school. Indeed she moves to a new neighborhood altogether so as not to run into her current inquisitive neighbors. The incident can now be put out of mind and eventually forgotten altogether; there will be nothing to remind her of it. Moreover, Mrs. Jones plans steps to provide her with fresh evidence of Johnny's honesty. She encourages him to do honest acts and rewards them with high praise. She provides him with a large allowance and many clothes, so that he will feel in need of nothing. She downplays success at school as personally important. She places her son in acknowledged positions of responsibility—no curfew, sending him to the bank for her, etc. Finally, she even takes steps to avoid any future evidence that would be adverse to her desired view of her son (and herself). No dishonest acts on Johnny's part will be known. She avoids contact with school officials, and even refuses to own a phone; in casual conversations she shies away from topics which could lead into a discussion of Johnny's past school performance. Finally, faced only with the' evidence she wishes to see, Mrs. Jones has forgotten about the stealing incident and believes that her son is as virtuous as ever. She cannot believe that Johnny has ever done a dishonest deed in his life. But just to be sure that her efforts are fully successful, Mrs. Jones insures that whatever facts may be brought to her attention in the future about Johnny will not be such as to "prove" any dishonesty on his part. She pays great attention these days to her own carelessness and forgetfulness; thus when money turns up missing from her purse, she always has some plausible explanation without pointing an accusing finger at Johnny. We can finish this story by adding that the various efforts of Mrs. Jones have their intended effect. She comes to be deceived about her son's dishonesty (or her belief about that dishonesty), and she does so on purpose. However, she would not acknowledge that as her purpose. So then, we may say that in self-deception, S resolves to perpetrate a deception upon himself, but the characteristics required at the beginning of this project (lucid planning and shrewd execution of elaborate maneuvers so as to avoid the truth, thereby being aware of his actions as rationalization) are precisely those which would be fatal to it at the end (where opacity with respect to this intention and his belief that p is necessary in order to be deceived).[44] S

---

44 Ibid., p. 44.

can purpose to deceive himself, but when he is finished he cannot notice that he had that purpose. As deceived S would then be unable to pronounce upon his own success. S can intend to deceive himself or do-it deliberately, but his success precludes recognition of the fulfillment of that intention. In speaking of the process of self-deception Audi says, "Presumably, the process often begins with S's consciously putting the evidence against p out of mind. . . ."[45] Likewise, Price has noted that- "we may sometimes consciously decide to assent or to act unreasonably."[46] At the outset an agent may be fully aware of what he is up--namely, unreasonably opposing the evidence which supports a dreaded proposition. However, this is only at the outset; the success of the project involves the loss of this awareness as the process advances.

> One can quite deliberately formulate a plan to deceive oneself, as one can blind oneself, surprise oneself, hoodwink oneself, etc. Only a recipe for self-deception is like a recipe for putting one's eyes out. At some stage one will not be able to supervise proceedings by looking to make sure everything is going according to plan. The man who successfully deceives himself may have begun by forming the intention of doing so, but success renders impossible any recognition on the part of the self-deceiver of the fulfillment of his intention.[47]

A self-deceiver is hereby portrayed as aware of his purpose to deceive himself at the outset, but in the process becomes distracted from and unwilling to assent to that awareness; he now mistakenly describes his actions. Note should be made here that S does not move from believing p to replacing that belief with another (incompatible) one; he Simply adds to his belief one which is incompatible with it, thereby falsely dissenting from it. However, with respect to his awareness of his intention, S moves from entertaining it in mind (conceiving of his aim in a particular way) to no longer entertaining it in mind; what he could formerly assent to (in principle) he no longer will because he has become deceived as to his intention. He makes himself forget what he was up to by overcoming the obstacles to forgetting, by avoiding and rationalizing the reminders, by dwelling on anything which will buoy him up in a different interpretation of his behavior regarding the evidence. Szabados remarks in passing that "self-deception tends to breed further self-

---

45   Audi, "Epistemic Disavowals and Self-Deception," p. 383; cf. "Epistemic Authority of First Person," p. 12.

46   Price, *Belief*, p. 231.

47   Champlin, "Self-Deception: Reflexive Dilemma," p. 293.

deception."[48] Not only is the self-deceiver deceiving himself about his belief, but when he engages in this purposefully he also comes to deceive himself about his intention.

However, an obvious objection might be raised against this account of how intentional self-deception may be possible.[49] If when S deceives himself intentionally he must make himself unaware of that intention, then he must deceive himself about his deceiving himself. If he purposely (rather than accidentally) prevents himself from assenting to and entertaining his intention in mind, then he is deceiving himself not only about his original belief but also about his intention to obscure that belief. But ex hypothesi this intention to deceive himself regarding his intention would also have to be obscured through a purposive self-deception, and likewise ad infinitum. It then appears that intentional self-deception becomes possible only at the expense of an infinite regress of intentions. It would be tempting, in response, to say that S need not actually intend to deceive himself about his intentions until some occasion arose which demanded it. Because nobody pays close attention to all of his intentions, S's intention to deceive himself might rest undisturbed and out of sight until it is questioned; special measures would not be called for until that time. Likewise, he would not have to deceive himself about the second self-deception until it in turn was questioned. And likewise with other intentions. The extended web of self-deception is created piece by piece, not all at once. In that case the infinite regress is mildly possible, but not actual. Accordingly the preceding criticism is not telling; we no longer have an infinite series of actual intentions, but rather a disposition to counter as many as may become necessary one by one. The infinite series is now only made up of possible intentions. While this answer would seem to defeat the criticism, it does so at the price of positing a disposition to self-deception beyond the conscious intention or purpose to deceive oneself--a disposition to deceive oneself if and as often as the question of his purposes arises. It is such a disposition which accounts for the claim that S would formulate new, further intentions to deceive himself as the need may be; to consciously purpose to pursue the second- level self-deception, and the third-level self-deception, and fourth-level self-deception, etc. would bring back the destructive infinite regress. Thus a disposition must be posited as explaining the self- deceiver's further attempts to deceive himself. But once this is admitted, there would

---

48   Szabados, "Wishful Thinking and Self-Deception," p. 205.

49   Lerner discusses this matter at some length in "Emotions of Self-Deception," pp. 162-169; cf. Fingarette, *Self-Deception*, p. 114; Bruce, "Investigation of Self-Deception," pp. 162-165.

appear to be no reason why such a disposition could not eliminate the original self-conscious purpose to deceive oneself as well. Why limit this dispositional explanation only to further self-deceptions? Accordingly, since our present effort is to utilize an understanding of intentions which does not resort to them as unconscious dispositions (thereby making the resolution of our perplexity concerning intentional self-deception too easy or questionable from the standpoint of other schools of thought) such an answer to the charge of infinite regress is not presently available to us. We still need to explain how someone can deceive himself on purpose--with the conscious awareness that he is planning to engage in deceptive maneuvers.

Fingarette gives us the lead on how to overcome the perplexity here, but his version of the answer suffers from its attribution of a "tacit" and "automatic" element in how S can intentionally deceive himself. Fingarette says that the self-deceiver adopts a policy of not spelling out his engagements in the world; by adopting such a policy the self-deceiver must also commit himself to not spelling out the policy itself, for then his attempt at not spelling out such engagements would be defeated. Thus the policy of refraining from spelling out one's engagements must be self-covering-- that is, the policy must include the provision that S not spell out the policy either.[50] This is a helpful lead. However, emit stands, the natural question is whether the self-deceiver intentionally adopts such a policy, in the sense that he self-consciously enter-tains the policy and assents to it (to himself).[51] This repeats the question that has been before us throughout this section of our discussion. It would seem clear that if Fingarette's self-deceiver intentionally adopts such a policy (and can one have a "policy" without consciously adopting it?), then he is aware of the very thing that the policy seeks to prevent. Again, intentional self-deception would be foiled. Accordingly, it is not surprising that in his description of this "policy" Fingarette speaks of it as tacit and automatic. In so saying he removes self-conscious adoption and preserves the success of the program--but at the price of confusion over the sense of 'policy' and of removing the possibility of engaging in self-deception on purpose (in the sense now under consideration). Moreover, if the policy is automatic, in what sense is the agent taking an action in self-deception at all?

However, Fingarette is correct in suggesting that the possibility of intentional self-deception is preserved, without succumbing to an infinite regress of intentions, by seeing that the self- deceiver deceives himself about believing p and about his purposes with one and the same intention. We

50   Fingarette, *Self-Deception*, p. 49.
51   Bruce, "Investigation of Self-Deception," p. 162.

have had occasion above to comment that one intention can encompass many actions, and in the case of self-deception which is engaged on purpose this is apparently what we have--<u>one</u> intention to <u>two</u> deceptions (about his awareness that p is true, and about his entertained purpose to rationalize away the evidence for p). This one-many relation between intention and action would defeat the criticism that intentional self-deception must resort to an infinite regress of intended deceptions of oneself. The intention to deceive oneself is considered self-covering because the self-deceiver simultaneously deceives himself about the character of his rationalizing activity (which he had purposely engaged, being aware of the character of his deeds) and the evident-ness of the proposition which he wishes not to believe. Indeed, in the process of rationalizing S would not accomplish the deed--would not convince himself with the distorted evidence--without distorting <u>many</u> of the things of which he could be aware (observations, memories, inferences which support p), and among those things would be the direct and indirect evidences of his aim; otherwise the rest of the distorted evidence would fail to be convincing, for S would still take it to be <u>distorted</u> in character. By rationalizing his intention S can convince himself on the basis of rationalized evidence that he does not believe p (or that he believes not-p). Intentional self-deception would allow for an agent to begin with an awareness of his purpose to deceive himself (by rationalizing evidence, etc.) because, as self-covering, in the process of successful accomplishment that awareness would be obscured as well. In the nature of the case, successful intentional self-deception precludes' the agent's awareness that some plan of action of his had been fulfilled. If he were later to come to realize that such a purpose to deceive himself had been successful, that plan would no longer be successful. As long as the intention to deceive oneself remains self-covering it can be successful.

Lerner has criticized the suggestion that intentional self-deception can be self-covering in this way, and he has insisted that all self-deception must ultimately be caused by emotions (which are not purposive actions, but dispositions).[52] His argument is that, while one emotion can have multiple results, intentions must have only single objects; a complex intention (one which has an object involving many actions, such as a self-referential or self-covering intention would be) can always be analyzed into its component parts. An intention which has a complex object amounts to many intentions, while an emotion which has multiple results does not break down into multiple emotions. If one is angry (an emotion) and both shouts and stamps

---

52   Lerner, "Emotions of Self-Deception," pp. 174-177.

his feet, he is still only angry once. However, if he has an intention to do something involving X and Y, then he actually has an intention to X and an intention to Y (two intentions). Therefore, according to Lerner, an intention must always have a simple object and cannot then be self-covering. Such a criticism plainly rests on a doctrine regarding the <u>individuating</u> of intentions and some doctrine regarding the criterion of <u>simplicity</u>.[53] Lerner does not explicitly state or defend his assumptions here. It is questionable whether he is correct on this matter.

If I have <u>two</u> intentions and succeed with only one of them, it would make sense to say that I have fulfilled half of my intentions. For instance, I may purpose to go to dinner, and I may purpose to go to the show; if it turns out that I was able only to go to dinner, then I would have accomplished half of my plans. According to Lerner's perspective, when a person intends to do something which involves doing X and Y, then he has two intentions (not one). In such a case, then, it would be conceivable (or sensible to say) that he could have accomplished half of that something (his plan). But this is wrong, I think, for there are some complex intentions (complex aims or purposes) where it makes no sense to speak of partial success or partial fulfillment (except in a morbid or joking manner). While the aim or intention of reading a book may be complex, involving the turning of pages, focusing of eyes, and an effort to understand a series of symbols, we would not say that a student who intends to read a book is partially successful if he only turns the pages. A man who intends to drive a car has not partially fulfilled that intention when he gets only so far as sitting in the driver's seat and turning the ignition. Yet if Lerner's implicit doctrine of how we should count intentions (individuate them) were accurate, our would-be-reader and would-be-driver should be said to complete their purposes partially. A complex intention (one whose object involves many elements or actions) need not be a mere divisible collection of intentions or series of intentions. If someone specifically intends to poison himself to death (having come from a long line of melancholy writers who gained reputations for poisoning themselves to death) he does not merely have the joint aims of (1) taking poison and (2) dying today. If he should take his arsenic, but then be shot to death by a jealous brother, he would have both taken the poison and died today. But he would not have fulfilled his intention to poison himself to death. A complex intention is not always analyzable into its simple components or elements, in that an intention to do many things is

---

53   The difficulty in holding to and defending a set doctrine of simplicity has recently and notoriously been exposed in Wittgenstein's *Philosophical Investigations*, sections 39-64, 215-216, 253-254. However, the problem of identity is an age-old and vexing metaphysical problem.

not many intentions. It was suggested above that intentions are individuated by the aim, purpose, or goal under which they are described or conceived. On such an approach it would be acceptable to unify many actions under one intention (e.g., typing a paper, walking to the library, building a house). On that basis the intention to deceive oneself could be self-covering: being constituted by a number of distorting maneuvers, including the distortion or obscuring of the original intention.

The self-deceiver intentionally does something which will remove even his awareness of the intention to do so. He intends <u>once</u> to deceive himself, which in itself would involve the obscuring of the intention as well as the obscuring of his (dreaded) belief that p. Is such a self-referential intention possible? Can a self-deceiver be <u>aware</u> (by way of his self-conscious intention, purpose, or resolve) that he aims to become <u>unaware</u> of something? Can one's intention defeat awareness of itself as well as something else? Can S intentionally take steps which will block his awareness even of the steps he took? Such questions naturally arise, but their critical thrust is easily put to rest. Of course there <u>are</u> very familiar and well known intentions that do this very thing. A prime example would be the intention to go to sleep. S can purposely go to sleep; he can intend to do so, and be quite aware that this is his intention (i.e., he would readily explain his actions under that description). And if he is successful in his plan, that very success will prevent him from being aware of his former intention; while asleep he will not be occurrently aware of anything (i.e., he will not entertain anything in mind, apart from the irrelevant possibility of dreaming), much less his purpose to reach that condition.[54] Likewise, in self-deception a person may very well entertain in mind, and be self-consciously aware, that he is about to deceive himself by rationalizing the available evidence which is adverse to his cherished belief; he can also be aware that this candid plan will not be successful apart from the deception covering his own intention as well. Yet if the purpose should be completed successfully, the awareness of this purpose will itself be obscured in his mind. It is likely an empirical truth that, for most people anyway, a person cannot reflect very long and hard on his intention (whether to go to sleep, or to deceive himself) and be successful in his project. If S is very explicitly conscious and dwells on the fact that he is counting sheep or trying to forget the day's problems because he wants to go to sleep, he may defeat

---

54   The suggestion that self-deception might be likened to falling asleep originates from Wilshire, "Self, Body, and Self-Deception," p. 442. Cf. "It is in our power to wake up, to become self-conscious and clearly aware of what is going on in us. . .   It is important to emphasize that we do have the power of waking ourselves up. . . ." (Price, *Belief*, pp. 230, 238).

himself; he may quickly need to <u>see</u> his efforts <u>as</u> something other than they are (i.e., distort their character or aim) in order for the efforts to be successful. In a similar way, the self-deceiver may very soon need to distort his awareness of his intention if his self-deceiving purpose is to be successful, but this fact does not preclude the possibility that someone may entertain in mind quite clearly that he intends to deceive himself about something. It only requires that the intention be self-covering, and we have seen that such intentions do have a place in our common experience.

The comparison between self-deception and the intention to go to sleep may be rehearsed once again briefly, incorporating various insights with respect to deception in general. When S attempts to deceive R, this can take the form of drawing R's attention away from good evidence--obscuring it, hiding it, distorting it, etc. Further, R comes to believe contrary to what S has the capacity to show as true--the deceiver has an unrealized capacity to undeceive his victim. Now in self-deception, S attempts to draw his own attention away from good evidence that supports a dreaded proposition. One way in which a person can purposely distract himself from painful realizations or psychic discomfort is by falling asleep.[55] Indeed, one can try to fall asleep--to do things unto that end, even though success precludes prolonged attention on the character of such efforts to distract one's attention (to help him fall asleep). Now, while asleep, one may still be said to believe the unpleasant truth, even though he is unaware of it and does not assent to it (to himself or others obviously). To make that dreaded belief manifest--to become aware of it, and of one's capacity to use it to keep himself awake—one can simply awaken from his sleep. Likewise, in self-deception one can purpose to distract himself from awareness of a dreaded truth by rationalizing the evidence connected with it (rather than literally falling asleep), even though success would likely preclude prolonged dwelling on the character of his intention. If he brings himself to believe that he does not believe the dreaded truth, then he (as when asleep) will not entertain his dreaded belief in mind or assent to it (to himself or others). Nevertheless, he will still believe the dreaded truth (as symptoms like his affective behavior indicate), and he has the capacity to undeceive himself. However, the former will not be assented to, and the latter will not be realized, apart from the self-deceiver "awakening" and putting an end to his deception, thereby exchanging a false belief about himself for a true one. Self-deception may be viewed as one unified phenomenon: the belief which is the <u>object</u> of self-deception (S's awareness of p's truth) is

---

55   E.g., consider Jack Burden's "great sleep" in *All the Rings Men* by Robert Penn Warren (Harcourt, Brace, & Co., 1946).

also the <u>cause</u> of S's attempt to deceive himself, and the intention to deceive himself about his <u>belief</u> includes the deceiving himself about the <u>intention</u> itself.[56]

---

56   Cf. Lerner's similar effort to characterize the cohesiveness of self-deception in "Emotions of Self-Deception," pp. 172-173.

Chapter Seven
Summary of the Solution and its Adequacy

7.1 The Analysis of Self-Deception

Self-deception has been a very familiar notion throughout ancient and modern literature, as observed in chapter 1, and yet recent analytical scrutiny of the notion has suggested that it is paradoxical. Questions have been raised as to the possibility of one person's being both deceiver and. deceived, thereby coming to believe what he simultaneously disbelieves. Nevertheless, since people so often refer to self-deception and seem to know what they are talking about, the notion most likely makes sense but has yet to be given an adequate analysis which could resolve its apparent paradox. Subsequent study has indicated that the phenomenon can be accurately described without contradiction, and thus the paradox can be resolved. Although there is a variety of forms of self-deception, including borderline and secondary uses of the term 'self-deception,' the basic or paradigmatic sense has been explored herein.

Chapter 2 argued that belief is an essential part of a proper analysis of self-deception. While a technical analysis of belief could not be given, it was characterized as a positive, intellectual, propositional attitude (not excluding false propositions) that is constituted by a continuing, intentional, mental state which, as stimulus-independent, is a potential contributing cause in one's mental, verbal, or bodily behavior (under suitable conditions)--such that he relies upon the believed proposition in his theoretical inferences and/or practical actions and plans. Accordingly, the mental state of belief may be manifested in a large variety of symptoms. Moreover, it was noted that not all of our beliefs are formed consciously, rationally, and with the giving of internal or external assent. The corrigibility of doxastic avowals and disavowals was also noted. Upon investigation of certain suggestions in the current literature on the problem we concluded that there is no good reason to omit reference to belief in the analysis of self-deception, and the

belief involved was seen to be genuine or full-fledged. Consequently a person who is self-deceived has (minimally) induced in himself, by controlling his attention to the relevant evidence, a belief in a false proposition.

The tendency is for people to model self-deception on interpersonal deception, but some writers, as discussed in chapter 3, have either challenged the legitimacy of that procedure or-have gone to the other extreme of making self-deception a literal case of other-deception. Both of these reactions were criticized for leaving us with inadequate characterizations of self-deception and merging it with related but distinct notions. Common ground can be found between self-deception and other-deception which justifies using the latter as a model for the former; the fact that it is a model allows for elements of discontinuity or difference between the two. Common elements were discovered, as follows: the deceiver's actions influence the deceived to believe falsely, there is a set of incompatible beliefs involved in the deception, the deceived must hold (at least implicitly) a false belief about the deceiver's beliefs, and the deceiver responds to discrediting evidence in a characteristically awkward fashion.

Chapter 4 discusses the incompatible beliefs which are involved in self-deception. The self-deceiver has come to a belief which he dreads, cannot face up to, or wishes were otherwise since it brings some unpleasant truth (as he perceives it) before him. Hence he brings it about that he believes that he does not hold that belief (or alternately, believes the contradictory of the proposition believed). To his first-order belief he adds a second-order (false) belief about his beliefs. The incompatible beliefs are ascribed to the self-deceiver on the basis of his varied and inconsistent behavior. The incompatibility between these beliefs is not logical (even when contradictory first-order beliefs are held) but practical or behavioral; they create conflicting actions on the part of the self-deceiver (e.g., his verbal and affective behavior is not uniform with respect to some proposition). Having answered objections to incorporating incompatible beliefs into an analysis of self-deception, we conclude that the self-deceiver is in a conflict state wherein he believes that p, but his assent to it is blocked by acquiring a second-order belief that he does not believe p. His disavowal is sincere but deceived, and he bears the responsibility for bringing about that deceived belief in himself.

The way in which the self-deceiver accomplishes his own deception is taken up in chapter 5. There it is argued that self-deception is a resultant state, stemming from certain rationalizing actions (e.g., controlling attention to the evidence, distorting it). The conflict state of mind comes about when, in the face of evidence adverse to his cherished belief, the self-deceiver engages

in pseudo- rational attempts to use the evidence in support of what he is motivated to believe. He goes to extreme measures to hide his belief that p from himself.' The explanation for this behavior must be motivational: he does it in order to satisfy the desire to avoid the discomfort and pain associated with believing that p. By means of this motivated rationalization he enters into and maintains self-deception, believing that he does not believe that p after all. Actions which have the effect of achieving this special state of incompatible beliefs are referred to when we say "S is deceiving himself regarding p."

Finally, as human actions, these actions may be taken intentionally. It is possible to purposefully engage in self-deception. This is the subject of chapter 6. Little help or clarity is offered, we see, in analyzing self-deception in terms of different kinds of consciousness. Rather, in terms of the analysis previously developed, we can say that the self-deceiver is aware of the truth of p, but not aware of believing p. He would be able to detect an incompatibility between beliefs if he saw it, but he does not think that his beliefs are incompatible. He disavows believing p. He is not aware that this disavowal is false, but he should be. Moreover, he could be aware of his motives in rationalizing the evidence, but he is unwilling to become so. These observations preserve the possibility of successful self-deception. However the question remains whether a person can deceive himself on purpose, just as he can obviously deceive others on purpose. If he did it on purpose, then it would seem that he would be aware of its deceptive character and could not be taken in by his actions after all. In contrast with this hasty conclusion it was argued that S's awareness of his aim to make his belief that p covert (by believing something incompatible with it) does not undermine the success of his deception effort as long as his intention to deceive himself is self-covering. In purposeful self-deception an agent not only tries to hide some belief from himself, but he also purposes to hide the hiding of it. The fact that self-deception can be self-covering prevents an infinite regress of deceptions. That there are intentions which aim to accomplish something and additionally to obscure themselves is demonstrated by observing the intention to go to sleep.

Therefore we conclude that the common notion of self-deception can be analyzed in this fashion: by rationalizing the adverse evidence, S brings himself to believe falsely that he does not believe that p, because he is motivated by that belief's distressing character to deny it. In terms of the characterization of belief offered previously, this analysis of self-deception amounts to the following: through motivated rationalization S brings it about that in his theoretical inferences and/or practical actions and plans he

mistakenly relies upon the proposition that he does not rely upon p in his theoretical inferences and/or practical actions and plans. Through motivated rationalization S brings himself to an incorrect belief about himself; he mistakenly uses it as a premise in his theoretical and practical inferences. Or to put it another way, through motivated rationalization S comes into a mental state of believing that he does not believe p, even though in fact he is in the mental state of believing p. S relies on a conception of himself as not relying on p; however his observed behavior indicates that p is part of his theoretical or practical inferences (i.e., he is in, the action-guiding state of mind of believing p). When he engages in such self-deception on purpose, the deception is self-covering; it obscures not only the dreaded belief but also the intention to rationalize away evidence in favor of it. Self-deception involves deception <u>by</u> the self, <u>of</u> the self, <u>for</u> the sake of the self, and <u>about</u> the self. The paradox of self-deception is thus only apparent and can be given a coherent resolution.

7.2 The Adequacy of the Solution

In chapter 1 certain criteria were set forth as defining what kind of solution to the paradox of self-deception would be adequate. In the first place it must supply the truth conditions for 'S deceived himself into believing p' (where p is the false proposition involved, amounting here to the belief that one does not believe something in particular). Accordingly we can render the truth conditions as follows on the basis of the above study:

1. S believes p, without assenting to it.

2. S is motivated to ignore, hide, deny (etc.) his belief that p.

3. By rationalizing .the adverse evidence S brings himself to believe that he does not believe that p (or alternatively, to believe not-p).

The second criterion laid down is that the solution must be true to the ordinarily recognized or paradigm illustrations of self-deception and be true to the ordinary language of 'self-deception.' This test is also passed by the solution proposed herein. Whether we look at the tragedies of *Oedipus the King* or *King Lear*, we see the above-mentioned conditions manifested as the protagonist moves through the painful steps of learning the truth in the end. A point comes where he believes the worst, but by misconstruing the

evidence he convinces himself that he does not believe it; it would be tragic to face. Or consider Tolstoy's portrayal of Count Rostov in *War and Peace*, as quoted in chapter 1:

> The count saw clearly that something had gone wrong during his absence; but it was so terrible for him to imagine anything discreditable occurring in connection with his beloved daughter, and he so prized his own cheerful tranquility, that he avoided asking questions and did his best to persuade himself that there was nothing very much wrong or out of the way. . . .

Rostov clearly sees something, but not wishing to admit it, he focuses his attention in such a way that he can bring himself to believe something incompatible with what he originally believed; moreover he purposefully engages in this act of self-persuasion. Likewise, the character of Strether in James' *The Ambassadors* finally awakens to what he had been doing in deceiving himself, and we read: "He recognized at last that he had really been trying all along to suppose nothing." He had intentionally acted to obscure a belief which he held without acknowledging it. The account of self-deception offered in this study can also accommodate the clear examples of self-deception offered in philosophical discussions of the subject. Pugmire tells the story of an incompetent official who loses job after job for administrative malfeasance. He continues to say to himself that he is going to face his boss and tell him off, to indicate that he realizes that jealousy (or some such thing) really lies behind his dismissal(s); and yet he never pursues the confrontation, making excuses for his failure to do so. Here we have a man who believes that he is incompetent, which explains why he never challenges those who "unjustly" dismiss him. And yet it would be too dreadful to admit that he holds that belief about himself; such a self-conception would be a crushing emotional threat. Therefore the official disavows that belief, maintaining all along that he is an innocent victim of a system that will not recognize his gifts and ability. By rationalizing the evidence and telling himself excuses over and over again, by focusing away from adverse evidence that his employers could give him and which he could find in his own behavior, he induces in himself a false belief. Or consider again the case of Mrs. Jones from chapter 6. The principal calls her and says that her son Johnny has been caught stealing lunch money out of students' desks. The evidence is plain: Johnny is a thief. (We can add to the story that this is the third or fourth time she has received such calls, she has noticed money missing out of her own purse at home, she is aware that Johnny has been coming home with expensive items from

the store, etc.) Mrs. Jones shows all the affective symptoms of believing the proposition that Johnny is a thief. She tries to avoid situations in which she is likely to be reminded of his dishonesty: moves to a new neighborhood, transfers Johnny into a new school, does not have a phone put in her home, keeps an unusually attentive eye on the boy, will not discuss why she does these things, etc. Yet on the other hand, since nobody in the Jones' family has ever stooped to dishonesty, she persuades herself that Johnny has never done a dishonest deed. She forgets past evidence and supplies "more credible" explanations of present evidence (e.g., the money is missing from her purse because she is so careless and forgetful). She expresses confidence in her son to others, makes a show of giving him mature responsibilities, and tries to do only what one who believed in Johnny's virtue would do. She freely confesses that she has a fine boy who is a joy to her, a paragon of virtue. And yet she flies off the handle at him over trifling matters, unlike the way she used to relate to him prior to the afternoon phone calls from the principal. She seizes upon every oblique reference to personal integrity to deliver an extended speech on the importance of honesty among the younger people of our day; she astonishes and embarrasses others by culminating her discourse with an emotional outburst that her Johnny is the living example of an honest child. When neighbors get curious over her missing cash and Johnny's new, unusually expensive acquisitions, Mrs. Jones fidgets, slightly blushes, looks away, answers in a halting fashion, etc. She treats the broached evidence in a very unusual and distorted fashion--all the while apparently satisfying herself that her interpretations are quite plausible. The scenario could be extended. But even from this much we should be willing to say that the affective symptoms of believing that Johnny is a thief justify us in attributing that belief to Mrs. Jones. She cannot stand the thought that her son, however, could stoop so low; she is motivated to hide this information and dissent from it. She has a desire to rid herself of the psychic discomfort caused by believing that Johnny is a thief. Consequently, she rationalizes the evidence ("the school officials have a vendetta against Johnny; they are framing the poor boy . . ."), leans on implausible interpretations of the facts, ignores the best and most obvious indicators, etc. and thereby brings herself to believe that she does <u>not</u> believe in Johnny's dishonesty. She fools herself about her awareness of the truth. The symptoms of this false second-order belief are nearly identical with believing the negation of the proposition 'Johnny is a thief.' She does not conceive of herself as not trusting her son, and so she assents to his virtue, makes a show of relying on him, etc. She meets all of the criteria proposed above. Accordingly we say quite naturally that she is deceiving herself. Therefore the analysis of self-deception which has

been offered is sufficient to describe accurately the common and obvious cases of self-deception to which we often hear reference. Although the term 'self-deception' is not always used in this way, the various other uses can be traced to this paradigmatic understanding of the notion. For instance, in non-central cases where every feature of the full account is not found, people can still use 'self-deception' in a secondary fashion to describe them.[1] We often say of a person that "he would have to deceive himself in order to believe what he does." In such cases we may simply be using the term for a reprimand, indicating that he (in common with genuine self-deceivers) is not dealing properly with the evidence involved.[2] Such ordinary usage picks out one feature of paradigmatic self-deception and uses it to make a dramatic point in a conversation. This kind of secondary or idiomatic use of the term 'self-deception' is no more a counterexample to our analysis than the statement "You'd have to be insane to believe that" is a counterexample to some psychological analysis of insanity.[3] Linguistic usage here reflects the central concept without intending to capture all of its features. Therefore, the present analysis of self-deception conforms to the second criterion of adequacy.

The third criterion requires that the analysis avoid logical contradiction, which the present proposal does. Likewise, in conformity with the fourth criterion, the present analysis goes to some lengths to distinguish self-deception from related notions such as wishful thinking, obstinance, faith, etc. Throughout the preceding discussion it has been precisely the failure to draw these distinctions which has been used to fault alternative accounts of the phenomenon. The fifth test requires that an adequate analysis not appeal to notions which are even more problematic than self-deception itself in explaining it. The account offered herein has abided by that standard, not resting on notions such as the "unconscious," etc. but confining the elements of the analysis to belief, motivation, and rationalization. The present analysis, in line with criterion 6, gives an entire chapter (3) to account for the fact that the term 'deception' can be appropriately used both in cases of interpersonal and interpersonal deception.

Finally, the analysis of self-deception which has been proposed can pass the test posed by criterion 7, that of incorporating the credible insights of alternative solutions to the paradox. For instance, the necessity of analyzing self-deception in terms of belief (cf. Rorty, Szabados), in terms of a conflict

---

1   Audi, "Epistemic Disavowals and Self-Deception," p. 384.

2   Siegler, "Demos on Lying to Oneself," pp. 473-475; "Self-Deception," pp. 32-43.

3   Cf. Gardiner, "Error, Faith, and Self-Deception," p. 231; Bruce, "Investigation of Self-Deception," pp. 99-101.

state (cf. Penelhum), in terms of first-order and second-order beliefs (cf. de Sousa), in terms of the presence of adverse evidence (cf. Canfield, Siegler), in terms of motivated rationalization (cf. Shea, Shapiro), and in terms of a self-covering policy (cf. Fingarette) are all elements which, along with others, have been recognized by various writers on the subject, and their insights have found their way into the proposal which this study advances.

Therefore, the proposed analysis of self-deception gives a coherent description of the activity and state of self-deception which renders it a venture that can be successful. It analyzes self-deception without contradiction, yet "saves the phenomenon." My argument has not consisted in some single demonstrative proof, but in a constructive elaboration of a perspective or account against the background of the failure of other approaches and the lack of critical refutations of my observations and proposals regarding particular elements of the analysis. The argument will have been successful if it has escaped negative attacks (as they might be found in the literature) on particular points in the central analysis, and if--as it seems--it has passed the tests for adequacy laid out previously.

Self-deception is a common notion which has appeared to some philosophers as paradoxical. However that paradox is only apparent, for a coherent account of self-deception can be given which is descriptively accurate. Nevertheless; the proposed analysis does not so dissolve the perplexity over self-deception that we now wonder why anyone should have been troubled over it in the first place. We have rendered a consistent account of inconsistent behavior, but the phenomenon is still disturbing to us. It has not been defined out of existence. Rather than the analysis being troublesome, however, it is the fact of self-deception itself which perplexes us so. In self-deception a person's rational powers seem to fail him so obviously, and that reminds us of our own frailties just here.[4] Consequently, self-deception forces us to revise our idealized picture of ourselves as consistent and integrated in what we believe. The phenomenon seems paradoxical because we tend to think of men's beliefs as rational.[5] But what ought to be, often is not. Self-deception thus continues to confront us with the disturbing fact of man's capacity for irrationality and duplicity. Given a paradox-free analysis of the common phenomenon, we must either adjust our self-conception or willingly engage in further acts of self-deception itself.

---

4   Shea, "Self-Deception," p. 172.
5   Shapiro, "Self-Deception," p. 9.

Selected Bibliography

Books and Dissertations

Ackermann, Robert J. *Belief and Knowledge.* Garden City, New York: Anchor Books of Doubleday and Co., 1972.

Ammerman, Robert R. and Singer, Marcus G., eds. *Belief, Knowledge and Truth.* New York: Charles Scribner's Sons, 1970.

Annis, David Boyden. "Knowledge and Belief: The Relation of the Two Concepts." Ph.D. dissertation, University of Illinois at Urbana-Champaigne. Ann Arbor, MI: Xerox University Microfilms, 70-781, 1969.

Armstrong, D. M. *Belief, Truth and Knowledge.* London: Cambridge University Press, 1973.

Aune, Bruce, *Knowledge, Mind and Nature: An Introduction to Theory of Knowledge and the Philosophy of Mind.* New York: Random House, 1967.

Bain, Alexander. *The Emotions and The Will.* 3rd ed. New York: D. Appleton and Co., 1875.

Bruce, Charles Dawson. "An Investigation of Self-Deception." Ph.D. dissertation, Michigan State University. Ann Arbor, MI: Xerox University Microfilms, 75-20,815, 1975.

Buchler, Justus, ed., *Philosophical Writings of Peirce.* New York: Dover Publications, 1955.

Butchavorov, Panayot. *The Concept of Knowledge.* Evanston, IL: North-western University Press, 1970.

Casey, John Perry, Jr. "Knowledge, Belief, and Evidence." Ph.D. dissertation, University of Iowa. Ann Arbor, MI: Xerox University Microfilms, 75-23,022, 1975.

Cochran, John Buchanan. "An Investigation into Self-Deception: Three Dialogues." Ph.D. dissertation, University of Oregon. Ann Arbor, MI: Xerox University Microfilms, 76-19,379, 1976.

Dilman, Ilham and Phillips, D. Z. *Sense and Delusion.* New York: Humanities Press, 1971.

Drengson, Alan R. "Self-Deception." Ph.D. dissertation, University of Oregon. Ann Arbor, MI: Xerox University Microfilms, 72-14,723, 1971.

Dretake, Fred I. *Seeing and Knowing*. Chicago: University of Chicago Press, 1969.

Dyke, Daniel. *The Mystery of Selfe-Deceiving*. 5th ed. London: Thomas Snodham, 1617.

Feldman, Richard Harold. "Non-Propositional Analyses of Belief." Ph.D. dissertation, University of Massachusetts. Ann Arbor, MI: Xerox University Microfilms, 75-27,513, 1975.

Fingarette, Herbert. *Self-Deception*. New York: Humanities Press, 1969.

Geach, Peter. *Mental Acts: Their Content and Their Objects*. Studies in Philosophical Psychology Series. New York: Humanities Press, (1957) 1960.

Ginsberg, Mitchell. *Mind and Belief: Psychological Ascription and the Concept of Belief*. New York: Humanities Press, 1972.

Grant, C. K. *Belief and Action*. Durham, England: University of Durham, 1960.

Griffiths, A. Phillips, ed. *Knowledge and Belief*. Oxford Readings in Philosophy. London: Oxford University Press, 1967.

Hampshire, Stuart, ed. *Philosophy of Mind*. Sources in Contemporary Philosophy. New York: Harper and Row, Publishers, 1966.

_____. *Thought and Action*. London: Chatto and Windus, (1959) 1960.

Helm, Paul. *The Varieties of Belief*. Muirhead Library of Philosophy Series. New York: Humanities Press, 1973.

Hintikka, Jaakko. Knowledge and Belief: An Introduction to the Logic of the Two Notions. Ithaca: Cornell University Press, 1962.

Laird, John. *Knowledge, Belief and Opinion*. Hamden, CT: Shoe String Press, 1972.

Lerner, Eric Joseph. "The Emotions of Self-Deception." Ph.D. dissertation, Cornell University. Ann Arbor, MI: Xerox University Microfilms, 75-27,038, 1975.

Martin, R. M. *Belief, Existence and Meaning*. New York: New York University Press, 1969.

Needham, Rodney. *Belief, Language and Experience*. Chicago: University of Chicago Press, 1972.

Newman, John Henry. *An Essay in Aid of a Grammar of Assent*. New York: Catholic Publication Society, 1870.

Oser, Janice Auritt. "Invitations to Self-Deception: An Application of Jean-Paul Sartre's Analysis of Self-Deception to Views of Self-Knowledge, Self-Identity, and Self-Deception Advanced, Respectively, by Gilbert Ryle,

Terence Penelhum, and Herbert Fingarette." Ph.D. dissertation, New York University. Ann Arbor, MI: Xerox University Microfilms, 76-19,531, 1976.

Peale, John Stafford. "The Theory of Belief." Ph.D. dissertation, University of North Carolina at Chapel Hill. Ann Arbor, MI: Xerox University Microfilms, 75-15,687, 1974.

Price, H. H. *Belief.* Muirhead Library of Philosophy Series. London: George Allen and Unwin, 1969.

Provence, Donald Lee. "Is Seeing Believing?" Ph.D. dissertation, Stanford University. Ann Arbor, MI: Xerox University Microfilms, 69-271, 1968.

Quine, W. V. and Ullian, J. S. *The Web of Belief.* New York: Random House, 1970.

Russell, Bertrand. *Philosophical Essays.* London: George Allen & Unwin Ltd., 1910 [1966).

Russell, James Michael. "Self-Deception." Ph.D. dissertation, University of California at Santa Barbara. Ann Arbor, MI: Xerox University Microfilms, 73-8105, 1971.

Ryle, Gilbert. *The Concept of Mind.* New York: Barnes and Noble, 1949.

Shaffer, Jerome A. *Philosophy of Mind.* Foundations of Philosophy Series. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1968.

Shapiro, Daniel. "Self-Deception." Ph.D. dissertation, City University of New York. Ann Arbor, MI: Xerox University Microfilms, 76-1491, 1975.

Shea, James Marvin, Jr. "Self-Deception." Ph.D. dissertation, Cornell University. Ann Arbor, MI: Xerox University Microfilms, 67-1411, 1966.

Sartre, Jean-Paul. *Being and Nothingness.* Translated and with an Introduction by Hazel E. Barnes. New York: Washington Square Press, 1966 [1956).

Sprigge, Timothy L. S. *Facts, Words and Beliefs.* New York: Humanities Press, 1970.

Valberg, Eugene. "Rationality and Self-Deception." Ph.D. dissertation, State University of New York at Buffalo. Ann Arbor, MI: Xerox University Microfilms, 73-29,146, 1973.

Articles

Alston, William P. "Self-Warrant: A Neglected Form of Privileged Access." *American Philosophical Quarterly* 13, no. 4 (October 1976):257-272.

_____. "Varieties of Privileged Access." *American Philosophical Quarterly* 8, no. 3 (July 1971):223-241.

Aqvist, Lennart. "A Solution to Moore's Paradox." *Philosophical Studies* 15, nos. 1-2 (January-February 1964):1-5.

Audi, Robert. "The Concept of Believing." *Personalist* 53, no. 1 (Winter, 1972):43-62.

_____. "The Epistemic Authority of the First Person." *Personalist* 56, no. 1 (Winter 1975):5-15.

_____. "Epistemic Disavowals and Self-Deception." *Personalist* 57, no. 4 (Autumn 1976):378-385.

_____. "The Limits of Self-Knowledge." *Canadian Journal of Philosophy* 4, no. 2 (December 1974):253-267.

Austin, J. L. and Anscombe, G. E. M. "Symposium on Pretending." *Proceedings of the Aristotelian Society, Supplementary Volume* 32 (1958):261-294.

Barker, John A. "Audi on Epistemic Disavowals." *The Personalist* 57, no. 4 (Autumn 1976)076-377.

_____. "Socratic Ignorance Vindicated." *Philosophical Studies* 28, no. 1 (July 1975):71.

Barrett, C. "Not Exactly Pretending." *Philosophy* 44 (October 1969): 331-338.

Bonny, W. L. "Mr. Deutscher on Saying-and. Disbelieving." *Analysis* 26, no. 1 (October 1965):17-20.

Braithwaite, R. B. "Belief and Action." *The Aristotelian Society Publications* 20, n.s. (1946):1-19.

_____ . "Symposium: Half-Belief II." *Proceedings of the Aristotelian Society, Supplemental Volume* 38 (1964):163-174.

_____. "The Nature of Believing." In *Knowledge and Belief*, pp. 2840. Edited by A. Phillips Griffiths. London: Oxford University Press, 1967.

Burns, S. A. M. Review of *Self Deception*, by Herbert Fingarette. In *Philosophy* 45, no. 171 (January 1970):72.

Butler, Joseph. "Upon Self-Deceit." *Sermons*, pp. 113-124. New York: Robert Carter & Bros., 1860.

Canfield, John V. and Gustavson, Don F. "Self-Deception." *Analysis* 23 (December 1962):32-36.

Canfield, John and McNally, Patrick. "Paradoxes of Self-Deception." *Analysis* 21, no. 6 (June 1961):140-144.

Cargile, James. "On Believing You Believe." *Analysis* 27, no. 6 (June 1967):177.

Castaneda, Hector-Neri, ed. "On Knowing (or Believing) That One Knows (or Believes)," *Synthese* 21 (1970): 187-203.

Champlin, T. S. "Double Deception." *Mind* 85, no. 337 (January 1976): 100-102.

_____. "Self-Deception: A Reflexive Dilemma." *Philosophy* 52, no. 201 (July 1977):281-299.

Chisholm, Roderick M. "Sentences About Believing." *Proceedings of the Aristotelian Society* 56, n.s. (1955-56):125-248.

Clifford, William Kingdon. "The Ethics of Belief." In *Belief, Knowledge and Truth*, pp. 39-44. Edited by R. R. Ammerman and Marcus G. Singer. New York: Charles Scribner's Sons, 1970.

Coburn, Robert C. "Believing Things," *Canadian Journal of Philosophy* 1, no. 1 (September 1971):93-103.

Collins, Arthur W. "Unconscious Belief." *The Journal of Philosophy* 66, no. 20 (October 16, 1969):667-680.

Daniels, Charles B. "Self-Deception and Interpersonal Deception." *The Personalist* 55, no. 3 (Summer 1974):244-252.

Danto, Arthur C. "Beliefs as Sentential States of Persons," *Language, Belief and Metaphysics*, pp. 10-23. Edited by Howard K. Kiefer and Milton K. Munitz. Albany: State University of New York Press, 1970.

Darmstadter, Howard. "Consistency of Belief." *The Journal of Philosophy* 68, no. 10 (May 20, 1971):301-310.

Demos, Raphael. "Lying to Oneself." *Journal of Philosophy* 57, no. 18 (September 1, 1960):588-594.

Deutscher, Max. "A Note on Saying and Disbelieving." *Analysis* 25, no. 3 (January 1965):53-57.

Drengson, Alan R. Critique of *Self-Deception* by Herbert Fingarette. In *The Canadian Journal of Philosophy* 3, no. 3 (March 1974):475.

_____. Review of *Self-Deception* by Herbert Fingarette. In Dialogue

no. 1 (1973):142.

Eberle, Rolf A. "A Logic of Believing, Knowing, and Inferring," *Synthese* 26 (1974):356-382.

Edgington, Eugene S. "On the Possibility of Rational 'Inconsistent' Beliefs." *Mind* 67, no. 308 (October 1968):582-583.

Exdell, John and Hamilton, James R. "The Incorrigibility of First Person Disavowals." *Personalist* 56, no. 4 (Autumn 1975):389-394.

Ezorsky, Gertrude; "On Danto's 'Belief as a Sentential State of Persons.'" In *Language, Belief and Metaphysics*, pp. 141-145. Edited by Howard E. Kiefer and Milton K. Munitz. Albany: State University of New York Press, 1970.

Factor, R. Lance. "Self-Deception and the Functionalist Theory of Mental Processes." *Personalist* 58, no. 2 (April 1977):115-123.

Fain, Haskell and Griffiths, A. Phillips. "On Falsely Believing that One Doesn't Know." *Studies in the Philosophy of Mind*, pp. 10-23. American Philosophical Quarterly Monograph Series, no. 6. Edited by Nicholas Rescher. Oxford: Basil Blackwell, 1972.

Feibleman, James K. "On Beliefs and Believing." In *Tulane Studies in Philosophy*, vol. 4: *Metaphysics and Belief*, pp. 11-30. New Orleans: Tulane University Press, 1966.

Fell, Joseph P. Review of *Self-Deception* by Herbert Fingarette. *Philosophical Quarterly* 20, no. 8 (July 1970):290.

Fohr, Samuel. "The Non-Rationality of Beliefs and Attitudes." *Personalist* 23, no. 1 (Winter 1972):63-70.

Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68, no. 1 (January 14, 1971):5-20.

Fuller, Gary. "Other-Deception." *The Southwestern Journal of Philosophy* 7, no. 1 (Winter 1976):21-31.

Gardiner, Patrick. "Error, Faith and Self-Deception." *Proceedings of the Aristotelian Society* 70, n.s. (1969-1970):221-243.

Graham, Keith. "Belief and the Limits of Irrationality." *Inquiry* 17, no. 3 (Autumn 1974):315-326.

Griffiths, A. Phillips. "On Belief." In *Knowledge and Belief*, pp. 127143. Edited by A. Phillips Griffiths. London: Oxford University Press, 1967.

Hamlyn, D. W. "Self-Deception." *Proceedings of the Aristotelian Society*, *Supplemental Volume* 35 (1971):45-60.

Hartland-Swann, J. "'Being Aware of' and 'Knowing.'" *Philosophical Quarterly* 7, no. 27 (April 1957):126-135.

Hauerwas, Stanley and Burrell, David. "Self-Deception and Autobiography: Theological and Ethical Reflections on Speer's *Inside the Third*

Reich.” *Journal of Religious Ethics* 2, no. 1 (Spring 1974):11-117.

Imley, Robert A. “Sayward on Assertion and Belief.” *Philosophical Studies* 19, nos. 1-2 (January-February 1968):76-79.

Jones, O. R. “Knowing and Guessing--By Examples.” *Analysis* 32, no. 145 n.s. (October 1971):19-23.

Jordan, James N. “On Comprehending Free Will.” *The Southern Journal of Philosophy* 11, no. 3 (Fall 1973):184-201.

King-Farlow, John. “Self-Deceivers and Sartrian Seducers.” *Analysis* 23, no. 6 (n.s. no. 96) (June 1963):131-136.

_____. Review of *Self-Deception*, by Herbert Fingarette. *Metaphilosophy* 4, no. 1 (January 1973):76.

Kiteley, Murray. “The Grammars of ‘Believe,’”*The Journal of Philosophy* 61, no. 145 n.s. (October 1971):19-23.

Linsky, L. “Deception,” *Inquiry* 6 (1963):157-169.

Lycan, William G. “Hintikka and Moore’s Paradox.” *Philosophical Studies* 21, nos. 1-2 (January-February 1970):9.

McLean, Murdith. “Episodic Belief.” *Philosophical Quarterly* 20, no. 81 (October 1970):389.

Mannison, D. S. “Lying and Lies.” *Australasian Journal of Philosophy* 47 (August 1969):132-144.

Martin, Richard M. “On Knowing, Believing, Thinking.” *The Journal of Philosophy* 59, no. 21 (October 11, 1962):586-600.

Mellor, W. W. “Knowing, Believing and Behaving.” Mind 76, no. 303 (July 1967):327-345.

Miri, M. “Self-Deception.” *Philosophy and Phenomenological Research* 34 (1974):576-585.

Mounce, H. O. “Self-Deception.” Proceedings *of the Aristotelian Society, Supplemental Volume* 45 (1971):61-72.

O’Connor, D. J. “Beliefs, Dispositions and Actions.” *Proceedings of the Aristotelian Society* 69, n.s. (1968-1969):1-16.

Paluch, Stanley. “Self-Deception.” *Inquiry* 10, no. 3 (Autumn 1967): 268-278.

Penelhum, Terence. “Pleasure and Falsity.” *American Philosophical Quarterly* 1 (1964):81-91. *Philosophy of Mind*, pp. 242-266. Edited by Stuart Hampshire. Sources in Contemporary Philosophy Series. New York: Harper and Row, Publishers, 1966.

Place, U. T. “The Infallibility of Our Knowledge of Our Own Beliefs.” *Analysis* 31, no. 6 (June 1971):197.

Pole, D. “The Socratic Injunction.” *British Society for Phenomenology* 2 (May

1971):31-40.

Powell, Betty. "A Note on Deceiving." *Analysis* 17 (1957):93-95.

Price, H. H. "Belief and Will." In Philosophy of Mind, pp. 91-115. Edited by Stuart Hampshire. New York: Harper and Row, Publishers, 1966.

_____. "Some Considerations about Belief." In Knowledge and Belief, pp. 41-59. Edited by A. Phillips Griffiths. London: Oxford University Press, 1967.

Pugmire, David. "A Doubt about the Normative Theory of Belief." *Mind* 81, no. 324 (October 1972):584-587.

_____. "'Strong' Self-Deception." *Inquiry* 12, no. 3 (Autumn 1969): 339-346.

Radford, Colin. "Does Unwitting Knowledge Entail Unconscious Belief?" *Analysis* 30, no. 3 (January 1970):103-107.

Radford, Colin. "Knowing But Not Believing." *Analysis* 27, no. 4 (March 1967):139-140.

_____. "On Sticking to What I Don't Believe to be the Case." *Analysis* 32, no. 5 n.s. 149 (April 1972):170-173.

Ramsey, Frank Plumpton. "Last Papers." *In The Foundations of Mathematics and Other Logical Essays*. Edited by R. B. Braithwaite. London: Routledge and Kegan Paul,-(1931) 1954.

Reilly, Richard. "Self-Deception: Resolving the Epistemological Paradox." *The Personalist* 57, no. 4 (Autumn 1976):.391.

Rorty, Amelia Oksenberg. "Belief and Self-Deception." *Inquiry* 15, no.4 (Winter 1972):387-410.

Saunders, John Turk. "The Paradox of Self-Deception." *Philosophy and Phenomenological Research* 35, no. 4 (June 1975):554.

Sayward, Charles. "Assertion and Belief." *Philosophical Studies* 17, no. 5 (October 1966):74.

_____. "More on Assertion and Belief." *Philosophical Studies* 22, nos. 1-2 (January-February 1971):20.

Scott-Taggart, M. J. "Socratic-Irony and Self Deceit":*Ratio* 144 no. 1 (June 1972):1-15.

Sellars, Wilfrid. "Belief and the Expression of Belief." In *Language, Belief and Metaphysics*. Edited by Howard E. Kiefer and Milton K. Munitz. Albany: State University of New York Press,"1970.

_____ . "Some Problems about Belief," *Words and Objections: Essays on the Work of W. V. Quine*, pp.. 186-205. Edited by Donald Davidson and Jaakko Hintikka. Hingham, MA: E. Reidel Publishing Co., 1975.

Shwayder, David S. "Self-Defeating Pronouncements." *Analysis* 16, no. 4 (March 1956):74-85.

Siegler, Frederick A. "An Analysis of Self-Deception." *Nous* 2 no. 2 (May 1968):147-164.

_____. "Demos on Lying to Oneself." *The Journal of Philosophy* 59, no. 16 (August 2, 1962):469-475.

_____. "Lying." *American Philosophical Quarterly* 3 (April 1966): 128-136.

Siegler, Frederick A. "Self-Deception." *Australasian Journal of Philosophy* 41, no. 1 (May 1963):29-43.

_____. "Self-Deception and Other Deception." *The Journal of Philosophy* 60, no. 24 (November 21, 1963):759.

Sousa, Ronald B. de. "How to Give a Piece of Your Mind: or, The Logic of Belief and Assent." *The Review of Metaphysics* 25, no. 1 (September 1971):526-533.

_____. "Knowledge, Consistent Belief, and Self-Consciousness." *The Journal of Philosophy* 67, no. 3 (February 12, 1970):66.

_____. "Review Discussion: Self-Deception." *Inquiry* 31, no. 3 (Autumn 1970):308-321.

Szabados, Béla. "The Morality of Self-Deception." *Dialogue* 13, no. 1 (1974):25-34.

_____. "Rorty on Belief and Self-Deception." *Inquiry* 17, no. 4 (Winter 1974):464-473.

_____. "Self-Deception." *Canadian Journal of Philosophy* 4, no. 1 (September 1974):51-68.

_____. "Wishful Thinking and Self-Deception." *Analysis* 33, no. 6 (June 1973):201-205.

Urmson, J. O. "Parenthetical Verbs." *Essays in Conceptual Analysis*, pp. 192-212. Edited by Antony Flew. New York: St. Martins, 1956.

White, A. R. "Belief Sentences." *Mind* 67, no. 268 (October 1958): 527-532.

Willis, Richard. "Prof. Black on Saying and Disbelieving." *Analysis* 14, no. 1 (October 1953):24-25.

Wilshire, Bruce. "Self, Body, and Self-Deception." *Man and World* 5, no. 4 (November 1972):442-451.

Woozley, A. D. "Knowing and Not Knowing." *In Knowledge and Belief*, pp. 82-99. Edited by A. Phillips Griffiths. London: Oxford University Press, 1967.