SELF-DECEPTION

SELF-DECEPTION

By

LORNE EDWARD LOXTERKAMP, B.A.HONS

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Arts

McMaster University

October 1971

For

Edward, Agnes and Lynne

MASTER OF ARTS (Philosophy)

McMASTER UNIVERSITY
Hamilton, Canada.

TITLE: Self-Deception

AUTHOR: Lorne Edward Loxterkamp, B.A.Hons (Simon Fraser University)

SUPERVISOR: Evan Simpson

NUMBER OF PAGES: v, 82

SCOPE AND CONTENTS: After several competing analyses of self-deception are examined and found wanting, a model is presented that not only isolates central cases of self-deception but also distinguishes it from other related phenomena with which it is so often confused.

ACKNOWLEDGEMENT

I am indebted to Evan Simpson and James Noxon for their helpful criticism of the matter and manner of earlier drafts of this essay.

CONTENTS

INTRODUCTION

Frequently we speak of one person deceiving another, that A has been deluded, duped, taken in by B. Less often we speak of one person deceiving himself, that S is both the deceiver and the deceived, that S is duped, deluded, taken in by himself. 'He is deceiving himself if he thinks he is a great musician', 'She cannot really believe she will pass that test unless she is deceiving herself', 'His mother deceived herself into believing that he would reform his ways', 'He is fooling himself into thinking his marriage will survive', 'When he fails to admit to his intolerance he is not being honest with himself' are some examples of ascriptions of self-deception. In addition there are others which, if not perspicuous ascriptions, strongly suggest the activity of deceiving oneself: 'If only he did not want that so badly he could see his error', 'It is impossible for him to believe what he says because he knows perfectly well that such-and-such' and 'How could he be blind to the obvious?'.

The fact that all of these very different ascriptions may be used to refer to self-deception points to a vagueness and unclarity in the notion itself — a vagueness and unclarity which perhaps can be removed by an appeal to examples whose salient features reveal the notion's central elements. But however diverse the phenomena covered by the notion of self-deception are, referring to such phenomena in reflexive terms, i.e., '__ deceives __self',

implies straight off that self-deception is analogous to interpersonal deception and that an analysis of the former will provide an analysis of the latter. How successful this approach is in giving a satisfactory account of self-deception is the primary concern of the ensuing discussion.

Two commonly regarded elements of self-deception are: (1) that S (= self-deceiver) has a mistaken belief, and (2) that S has the capacity to recognize his mistake.[1] Without the first element, self-deception would be in some instances indistinguishable from a lucky guess, unbending faith or simple trust, where the belief in question is, as it happens or turns out, true. To be sure, the point in accusing someone of having deceived himself is that he is deemed responsible for holding an erroneous belief. Without the second, self-deception would be indistinguishable from psychosis and other forms of mental aberration in which the person has lost the ability to discriminate normally. Indeed, if the person has deceived himself, if he has the power to instil a false belief, then he must have the power to set things aright.

From a literal interpretation of what it means for a person to be in self-deception, the paradoxical version of the epistemic model arises. Beginning with an analysis of interpersonal deception, we find that A deceives B if (a) A believes (or knows) p, (b) B comes to believe q (which is incompatible with p) as a result of (c) A's intention to instil a mistaken belief, and (d) q is false. Now if self-deception mirrors interpersonal deception, it follows that the self-deceiver holds two incompatible beliefs one of which

it is his intention to hold. So, it appears that S is deceiving himself if and only if (a) S believes (or knows) p, (b) S succeeds in bringing it about that S believes q (where q is incompatible with p) as a result of an intention to do so, and (c) q is false.

According to this analysis, the self-deceiver must intend to hold a false belief which in turn requires that his activity to produce that result involves the recognition that the belief is erroneous. But a paradox looms. How is it possible for someone to successfully bring himself to have a false belief while recognizing or having access to the truth? Interpersonal deception is intelligible because the deceived is unaware of or lacks correct information. But in self-deception S is both deceiver and deceived, he both believes correctly and incorrectly while having good reason not to hold the false belief. How could anyone be fooled if he knows the truth? How could anyone be taken in by his own pretence? How, and this is the crux of the paradox, could anyone <u>intend</u> to hold a belief whether true or false?

If self-deception is analyzed in accordance with interpersonal deception, the paradox, how is it possible for someone to intentionally bring it about that he holds a false belief, seems unavoidable, and the difficulty in resolving it seems insurmountable.

The aim in interpersonal deception is the deceiver's desire to get the deceived to believe what is false. But in self-deception problems arise if the self-deceiver's aim is described as getting himself to believe what is false. If A believes p, then A believes that p is true, and if A discovers that p is false then A abandons

the belief. Beliefs represent, or claim to represent, how it is with the world. Since one can only believe what he takes to be a correct representation of the world, it is therefore perverse for someone to set out to believe what he takes to be an incorrect representation of the world. Moreover, the suggestion that someone could intend to believe something is equally unintelligible. The locution '__ intends to believe p' is fundamentally incoherent, because if a belief were the sort of thing which were acquired as a result of an intention to do so then beliefs would cease to be pictures of the world. If one could bring it about that one believed that p merely by deciding to believe p then a belief that p would no longer imply that p is true. We can twist someone's arm to release information about the truth of p; but we cannot, as it were, twist the world's arm to make a proposition true. As we cannot intend that something be true, so we cannot intend that we believe that something is true (or false).

Another objection to this kind of analysis is that the inclusion of incompatible beliefs among the conditions seems to bring the phenomenon so analyzed to be more correctly covered by other related mental terms. Thus intentionality and incompatible beliefs, it is argued, do not enter into the description of (central cases of) self-deception.

In Chapter I three attempts at an analysis of self-deception along the lines mentioned above, one contending with the paradox as stated and the others eliminating the most troublesome features, are discussed. A recent proposal and the most extensive treatment

to date, which rejects the epistemic model for a self-ascriptive model and purports to preserve the paradox in a palatable 'translated' form, is examined in Chapter II. Finally, in Chapter III, self-deception is distinguished from other phenomena to which it is closely related but with which it is frequently confused, and a non-paradoxical description of self-deception employing a revised epistemic model is presented which not only avoids the problems in the epistemic model as commonly interpreted but also attempts to explain the truth behind the metaphor that the self-deceiver is blind to the obvious.

CHAPTER I

THE EPISTEMIC MODEL[1]

1. In the paper which is perhaps most responsible for the recent interest in the problem of self-deception, Raphael Demos proposes an account of self-deception that incorporates a full-blooded explication of the epistemic model. While recognizing that "self-deception" has more than one use, he attempts to analyze what appears to be the central meaning. In doing so he combines analyses of deception and lying; since deception, while emphasizing the securing of the desired result, need not include intention, and lying, while emphasizing intention, need not include the securing of the desired result. (Thus to have deceived you, I need not have lied; and to have lied to you, I need not have deceived.) So, working with the elements of both deception and lying, the analysis of interpersonal deception pertinent to self-deception becomes: "'B lies to (deceives) C' means: B intends to induce a mistaken belief in C, B succeeds in carrying out this intention, and finally B knows (and believes) that what he tells C is false. All three: intention, results and knowledge, are included."[2] And, accordingly, "Self-deception exists, I will say, when a person lies to himself, that is to say, persuades himself to believe what he knows 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, perhaps the existence of a contradiction."[3]

Oddly enough, Demos locates the paradox not in the element of intention but in the self-deceiver's holding two incompatible beliefs. Apparently self-deception, baldly stated in terms of lying to oneself, is logically impossible because

> 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 respect. When B lies to himself he comes to believe what he knows to be false; to accept this as the description of a fact is to admit a violation of the law of contradiction. It would seem, then, that self-deception — lying to oneself — is logically impossible in the way it has been formulated. Perhaps, then, the description given is wrong.[4]

However, as he remarks, two obvious non-paradoxical redescriptions fail to significantly cope with the evidence. If the contrary beliefs occurred at different and successive times or the agreeable belief occupied the conscious mind while the disagreeable one were repressed into the unconscious, then, Demos contends, these suggested redescriptions would simply fall wide of the mark. For example, "The mother who has convinced herself that her prodigal son is a fine boy is haunted by a 'nagging doubt'. Here both the belief and the doubt are simultaneous and both seem to be in the conscious mind. Or take the man who deceives himself about his attractiveness to the ladies. Backed into a corner by his friends and confronted by past failures which he cannot deny, he confesses the truth, adding, "I knew all along I am no good"."[5] So, on Demos' view, self-deception must be characterized by an inner conflict which consists

in contrary beliefs held simultaneously in the consciousness of the person.

Since, according to Demos, a person cannot consciously hold contrary beliefs at the same time in the same respect and yet in self-deception both beliefs must be in the consciousness of the person, the problem becomes one of making intelligible the description of the self-deceiver as one consciously holding contrary beliefs in different respects.

He proposes a solution to this problem in arguing that contrary beliefs can exist as conscious beliefs held in different respects by attributing each belief to a different level of awareness: "one is simple awareness, the other awareness together with attending or noticing . . . I may be aware of something without, at the same time, noticing or focusing my attention on it. This comes about because I may be distracted by something else, or because I may deliberately ignore it, or because I may not wish to think about it. The not-noticing need not be something that just happens to me."[6] So, applying this to self-deception, when S believes p while in the state of knowing and believing not-p he is conscious of both beliefs but aware of them on different levels. Although this might provide a solution to the problem stated above, it does not come to grips with all the problems in the analysis Demos presents. There still remains the problem of handling the intentional feature. But before the adequacy of the proposed solution is examined, there is a misconception to be cleared away.

His claim that it is logically impossible for contrary

beliefs to exist in the same person at the same time in the same respect (for whatever the last rider may mean) is palpably incorrect. It is logically impossible for someone to know p and to know q (where q is incompatible with p) at the same time because knowing that it is raining entails that it is raining. If it is raining and A knows that fact then it is logically impossible for him to know that, say, the skies are clear (or, for that matter, that it is not raining). But whereas the truth of p is a necessary condition of knowing that p, no such condition exists for believing. It is not logically impossible for someone to believe p and believe q at the same time, although to say that it is logically possible to hold contrary beliefs simultaneously is not to say that it is also non-paradoxical. So, if holding incompatible beliefs simultaneously is paradoxical, the paradox is not a logical one. Perhaps the oddity of ascribing incompatible beliefs to someone at the same time derives not from logical considerations but from the failure to understand how someone could (rationally) think or act as if p is true and think or act as if q is true. We presume rational animals behave and wish to behave rationally; and rationality implies consistency in thoughts and behaviour.

So, contrary to Demos, there is no logical difficulty in attributing incompatible beliefs to one person at one time. Indeed, there are several possible explanations of how someone might consciously hold incompatible beliefs at the same time, provided he is not aware of their incompatibility. Some examples are: (1) A believes that Trudeau is married and that the Prime Minister of Canada

is a bachelor because A is not aware that Trudeau is the Canadian Prime Minister. (2) When the belief that Trudeau had married was activated in A (while he believed that Trudeau, the Prime Minister, is a bachelor) he failed to appreciate the relevance of the new belief to the old, and thus failed to make the required adjustment. (3) If p and q are unobviously incompatible (e.g., p = Next year each numbered day of the week in February will correspond to each numbered day of the week in March, and q = Any year whose last two digits are equally divisible by four is a leap year; and suppose the last two digits of the year in question are equally divisible by four) then it is possible for someone to be conscious of both and not recognize their incompatibility.

However, Demos may not be troubled by this because his chief concern seems to be explaining how it is possible for someone to be conscious of his incompatible beliefs and at the same time conscious of their incompatibility. Normally only the mentally defective or irrational person behaves in this fashion. But the self-deceiver is neither mentally defective nor irrational. To resolve this problem, Demos suggests that the self-deceiver is aware of both beliefs, but aware of each on a different level so that they do not clash. But, as we shall see, this solution has problems of its own.

Any solution to the problem as set by Demos must satisfy the crucial requirement that S, in deceiving himself, intends to induce in himself an erroneous belief while he knows the truth.

The solution Demos proposes implies that the paradox disappears when the deceiver is described as being aware of the contrary beliefs on two different levels. S is aware of his true belief on the level of simple awareness (and presumably aware also of the incompatibility of the true and false beliefs) and aware of his false belief on this level together with his attending or noticing it. However, the argument gets its force from an equivocation on the word "aware".

Now it is not entirely clear what is meant by a level of awareness because it is not entirely clear what the term "awareness" means.

To say that A is now aware that he is wearing blue trousers is to say either that A is now examining his trousers or attending to the fact that he is wearing blue trousers, or that while he is not at the moment paying attention to the colour of his trousers he is not ignorant of that fact. The use of "aware" and its cognates in the latter instance allows for substitution with either "know" or "believe" without loss of import. For example, when A blithely announces his intention to wear an orange shirt, B responds with 'You can't wear an orange shirt with blue trousers, the colours will clash'. A's rejoinder, 'I am aware of that but it is the only shirt that is clean', implies that although he was not at the time considering the fact that he was wearing blue trousers and the fact that the colours orange and blue clash, he knew both and if a suitably coloured shirt were clean he would wear it. A different use of "aware", according to Demos, is this latter 'simple awareness' plus attending or noticing. A is aware on this level if, say,

when announcing his intention, he manifests displeasure at the thought of the gaudy result.

'A is aware that p' is, then, ambiguous. It can mean either (a) A knows that p but is not now attending to p, or (b) A is noticing that or attending to p. If (a) is what Demos labels simple awareness then his position is that (b) consists of (a) "together with attending or noticing". But this is hardly the case. The use of "aware" in (b) does not entail that A either knows or believes that p; it entails only that A is entertaining p. I am presently aware of, i.e., attending to, the proposition that there existed a king named Arthur who led the Knights of the Round Table, but I neither believe nor disbelieve it. I merely entertain the possibility. And to have the propositional attitude "entertain" to p is not tantamount to having either of the propositional attitudes "know" or "believe" to p; to entertain p is not to commit oneself to an acceptance of p. Thus (b) neither consists of nor entails (a), and so the two senses of "aware" are distinct and separate. And hence there is no sense to be made of the contention that the self-deceiver is conscious (in case "conscious" here has a univocal meaning) of both beliefs but on different levels. But in any case, even if (b) would entail (a), we do not seem to have successfully avoided what Demos intends to avoid, namely, ascribing incompatible beliefs in the same respect or on the same level of awareness. S's being aware of a belief on level (b) entails that he is also aware of it on level (a). But if he is aware of the incompatible belief on level (a) as well, then he holds both beliefs

on the same level of awareness and in the same respect, which Demos claims to be logically impossible.

While it is true that "I may be aware of something without at the same time noticing or focusing my attention on it", it does not follow from the fact that "aware" is used in both senses (a) and (b) and that A can be aware of p in sense (a) and aware of not-p in sense (b), that "aware" means the same in both instances. Indeed the distinction between senses (a) and (b) precludes just such an inference. Demos cannot have it both ways. He cannot have the two levels of awareness to be so distinct and so similar that the self-deceiver can be said to be unambiguously aware of both beliefs but in different respects. And, as I shall argue shortly, his failure to establish a univocal sense of awareness for the different levels of awareness undermines his attempt to establish intentionality.

The point Demos tries to make about levels of awareness is illustrated by his contention that one's headache may continue to exist even if one does not notice it, because, say, one is engrossed in a movie. "While watching the story on the screen I don't "feel" my headache; but as soon as the play is over, I feel a headache once more . . . the headache continued existing, but I did not notice it."[7] Demos wants a case in which someone has a pain but does not feel it in order to produce an analogue to being conscious of a belief but not noticing it. But pains are so unlike beliefs that the analogy must fail. The differences between pains and beliefs that rule out the analogy are familiar enough and pursued elsewhere[8] as to warrant my not repeating them here.

What cases of self-deception does Demos' analysis purport to describe non-paradoxically? Demos explicitly restricts the scope of his analysis to those instances where intention, results and knowledge obtain. He distinguishes the use of the expression '__ is deceiving __self' which is analogous to the application of 'A is deceiving B' from other uses in which something other than deception is implied. A person who holds (or seems to hold) an erroneous belief as a result of wishing that something he has little or no control over were so, is more aptly described as engaging in wishful thinking. In such cases the person usually yields to an emotion or impulse which diminishes responsibility for holding the erroneous belief. And because "We must also be able to say of a person; he knew what he was doing, and he could have done otherwise", the use of "self-deception" and its accompanying ascription of responsibility should normally be withheld in just those cases. Delusions of grandeur while drunk, hypnotized or sick are strictly different from self-deception in that "the person having the delusions experiences no conflict; there is no countervailing belief." Self-deception is also distinct from pretending to oneself although the differences are more tenuous. Usually pretending to oneself very nearly approaches (total) belief and thus self-deception. For example, "those with a dramatic temperament, the enthusiast and the like, preserve some sense of reality; in some 'corner of the mind' they know that it is all an act."[9]

Thus, taking the line suggested in "pretending to oneself", self-deception is found in a "mother who has convinced herself that

her prodigal son is a fine boy is haunted by a 'nagging doubt'" and a "man who deceives himself about his attractiveness to the ladies [and when] Backed into a corner by his friends and confronted by past failures which he cannot deny, he confesses the truth, adding 'I knew all along that I am no good'."[10] But are these genuine cases of self-deception, let alone ones covered by Demos' analysis? I think not.

There is little information to go on, but let us see whether what there is measures up to the conditions Demos sets. To be in self-deception the woman would have to (1) know that her son is prodigal, (2) believe he is a fine boy as result of her intention to do so, and (3) 'notice' only the latter belief. It is questionable whether her nagging doubt amounts to knowledge: both the doubt and the belief that her son is a fine boy may be the consequence of conflicting pieces of evidence, such that although she has good reason to believe he is a fine boy there may be (mounting) evidence that he is not. But for the sake of argument, let us assume that the nagging doubt is a product of her knowledge that the boy is prodigal. Now, presumably, a nagging doubt is one that frequently comes to one's attention. But if this is so, by attending to the true belief, she then fails to meet condition (3).

The 'ladies' man' example undergoes a similar fate. To be self-deceived, he must (1) know that he is unsuccessful with women, (2) believe contrary to this knowledge as a result of his intention to do so, and (3) attend to only the latter belief. Suppose that his repudiation of his recent behaviour by his admission that he

knew all along fulfills condition (1). Suppose too that he was oblivious of his true belief until prodded by his friends. This should satisfy condition (3). Nonetheless there still is a problem. If he did at no time attend or 'notice' his true belief, then to say that he intended to believe mistakenly is without support. How could he have intended to believe mistakenly if he did not recognize the error he wanted to commit?

Of course, these cases might still be instances of self-deception, even though Demos' analysis does not fit. But there is no good reason to think they are. What the woman says publicly about her son is equally a manifestation of her love, an emphasis of his good traits, or a matter of wishful thinking rather than a product of self-deception. And it is implausible to suggest that a woman who wishes her prodigal son behaved better, or who believes that in spite of it all he is basically good and makes every effort to show her concern and love, is self-deceived. Similarly, the 'ladies' man' does not seem to differ significantly from a man who, through forgetfulness or the failure to attend to certain facts, comes to hold a mistaken belief about himself. Here we have a case of someone simply holding inconsistent beliefs, not someone in self-deception.

My criticism is not only that Demos has not produced a convincing example of self-deception conforming to his analysis, but primarily that no plausible case of self-deception whatsoever will fit his analysis, because the condition of intentionality cannot possibly be satisfied unless the person 'notices' his false

belief. If the person intends to believe falsely, then his recognition of the falsehood he wishes to believe must be simultaneous with the recognition of the truth of the matter. How else is he able to decide what is false? So if the condition imputing intention is to be made sense of, S must 'notice' his true belief and 'notice' that the belief he intends to hold is false. That requires that S is aware of both beliefs in sense (b), which contradicts Demos' contention that S is aware of the true belief in sense (a) and the false belief in sense (b). Of course, some of Demos' remarks suggest that part of the self-deceiver's intention is to remove the true belief from his attention. This may come about by deliberately ignoring it or not wishing to think about it.[11] But for either to obtain, S must attend to what is to be ignored in order to ignore it. And if to put oneself into a state of self-deception requires that one not attend to the true belief, then to deliberately ignore it is to defeat one's purposes.

According to Demos, either the self-deceiver is simply not noticing a belief or deliberately ignoring it. If the former, then self-deception does not seem to differ from someone holding an incorrect belief because he has forgotten what he knows, or from someone holding inconsistent beliefs and for any of a number of reasons not recognizing their incompatibility. If the latter, then self-deception on Demos' analysis is impossible. But perhaps the assumption that an analysis of interpersonal deception will throw light on self-deception is wrong.

2. Two subsequent attempts at giving a non-paradoxical account of self-deception avoid the seemingly insurmountable difficulties posed by its alleged correspondence to interpersonal deception. Frederick Siegler argues against the correspondence and formulates a motivational explanation of self-deception. Terence Penelhum, denying that self-deception involves motives, bases his analysis on the assumption that the self-deceiver believes in the face of strong evidence to the contrary, with the aim of producing an intelligible rendering of the supposed paradox.

Siegler rejects the possibility of the criteria for interpersonal deception providing an analysis of self-deception. He claims the criteria for self-deception which parallel the criteria for interpersonal deception, i.e., (1) S knows p,[12] (2) S believes not-p, (3) S believes not-p as a result of S's intentional procedure to believe not-p, cannot be fulfilled because there is a logical oddity in saying of someone that he believes what he knows to be false:

> For, as these words are normally understood, "A believes (knows) what he knows (believes) to be false" does not make sense. Since it was a feature of interpersonal deception that the deceiver know that what the deceived believes is false and this cannot make sense where the deceiver and the deceived are the same person.[13]

Indeed there is a prima facie oddness in saying of someone that he believes what he knows to be false, or someone sincerely asserting 'I believe what I know to be false'. (The latter would be odder

still if subsequent to the declaration no attempt was made to make the necessary adjustment.) But this is not a logical oddity. If A believes p and knows not-p and knowledge entails belief, then A believes p and believes not-p. Here we have someone holding inconsistent beliefs which, as we have seen, is not logically impossible. However, if it is allowed that believing not-p entails disbelieving p, then it seems that there is a contradiction. But again it must be recognized that incompatible beliefs, unlike incompatible ascriptions of knowledge, do not exclude one another. All that is required is to show that it is possible for someone to be in a certain belief-state in set of circumstances and in a contrary belief-state in a different set of circumstances over the same period of time. And that does not appear too difficult. We should not confuse the claim of ascribing a belief to someone and denying that he has that belief with the claim that he holds contrary beliefs: only the former is self-contradictory.[14]

While I think that that part of Siegler's criticism of the analysis à la Demos is slightly askew, Siegler does attack the central problematic feature, namely, that the self-deceiver is taken in by his own intention to believe falsely.

If S is to be correctly described as having intended to bring himself to believe what he knows to be false, it would have to be established that:

(1) S knows p,
(2) S intends to believe q, which requires that,
    (a) S recognizes both that he knows that p and

that p is incompatible with q, and,

(b) S has reason to bring it about that he believes q, i.e., there is a logical connection between the object of S's action and the answer to the question 'why did S do it?', [15]

(3) S believes q.

Could there ever be an instance of self-deception which would meet all the above conditions? I think not.

A typical case which the analysis in question should satisfactorily describe is something like the following.[16] S discovers from a reliable source (his doctor) that he will die from an incurable disease within ten days. The features of the case up to this point are that the evidence for the prognosis are conclusive, S verbally accepts the prognosis, and S's behaviour shows that he accepts the prognosis. After four days have elapsed, S's behaviour alters significantly despite no change in the medical prognosis; the evidence is still conclusive. S verbally rejects the prognosis and his behaviour shows that he rejects it. So it looks as if S believes what he knows to be false. Are there sufficient grounds to establish that S intended to believe erroneously?

On the first few days, S behaves as if he were going to die, i.e., making funeral arrangements, writing his last will and testament, etc. On subsequent days, S shows signs of stress and great anxiety. We find him muttering to himself, 'I want to live, I must live . . .' He begins to read stories about miraculous cures, unexplained recoveries from terminal diseases. On the fifth day, he tells us that he will be cured, cancels the funeral preparations,

and destroys his last will and testament.

Does S's behaviour establish an intention to hold a false belief?

> Here we can conclude that ⌜S⌝ did believe that he was going to die, and that now he believes that he is not going to die. He initiated a procedure which we could say resulted in the change of his belief. But so far we would be justified in concluding only that ⌜S⌝ had convinced himself that he is going to live. This is because there is no indication that ⌜S⌝ intended to induce in himself what he believed to be a false belief. His procedure may have included poor or silly, or otherwise unfounded considerations, but it was not a procedure by which ⌜S⌝ intended to produce a belief which he believed to be false.[17]

This seems correct. But more is needed to see exactly why we are unable to say of S that he intended to believe erroneously.

An objection might be urged along the following lines. 'S's mutterings expressing his desire to live and his searching out literature on miraculous cures are undoubtedly intentional. And since the false belief is acquired as a result of these activities, then, surely, S can be spoken of as having intentionally induced a false belief in himself.' There are two replies to this objection: The first is that although it can be admitted that S came to hold a false belief as a result of certain intended actions, it does not follow that he intended to believe what was false. Perhaps the belief was a fortuitous consequence of his activity even though it was not intended as such. Furthermore, cases in which someone does deliberately set out to hold a false belief by putting himself into

such a situation that he is led to hold the false belief and either drop or 'forget' the true one fail to qualify, because if the former then he does not hold two beliefs and if the latter, since the belief 'forgotten' is normally a source of discomfort, then the person's activity bears a striking resemblance to wishful thinking and repressing an unpleasant thought.[18] The second reply is that if S did intend to hold a false belief, then he must have had reason to do so. What reason, then, does S have which leads him to accept the stories and believe that he too will be miraculously cured?

> His reason for believing the stories is his own honest, non-observational explanation of why he believes them. He may confess that he believes the stories because they seem so plausible. But he could not say that he believes the stories because (plausible or implausible) he wished to believe them and that he accepted them as proof that he will live. It may be that the reason (which explains) why he believed foolish or implausible stories and why he accepted silly stories as proof that he failed to see just how silly or implausible the stories are. But this is not <u>his</u> reason for believing and accepting <u>the</u> stories or for believing and accepting <u>silly</u> and implausible stories. For his reason for believing and accepting <u>the</u> stories is that he finds them implausible.[19]

But even if the explanation Siegler gives is right, it does not rule out the possibility of S's intending to believe that he will be cured. To say that S had a reason for believing suggests that his coming to believe was intentional. However this need not be the case.

S tells us that he believes that he will be cured because he finds the stories plausible. Suppose that S is also in a position to acknowledge that some such stories are fabrications, and that many others have been scientifically explained, and that it is reasonable to assume in cases where an explanation is not available that there is a natural explanation nonetheless, and most importantly that no one with his disease has been known to be cured. Thus, it seems that although S has no grounds for believing that he will be cured he believes nonetheless. This shows, I think, that the "because" in 'S believes p because e' may signal only a causal relation between p and e and not a rational connection. The case of S believing that he will be cured because he finds the stories plausible can be construed as a case of someone holding a belief which lacks support, yet which the person was led to believe. Not every belief is held on the basis of rational support; some beliefs we have as a result of certain things which cause us to have them.[20] So it is somewhat misleading to say, as Siegler does, about S that since he believes p because e, e is his reason for believing p — as if e provided rational support for p. Indeed what appears to be true is that S believes p as a result of being in a certain mental state which permitted e to activate a belief that p in him. And to deny that S has or had a reason (i.e., one which provides rational support) strengthens the case against the claim that S intended to believe he could be cured.

The fact that intentionality has not been established in this particular case is symptomatic of the fact that the statement

'S intends (intended) to believe falsely' involves incoherencies concerning the nature of belief which do not seem readily, if at all, resolvable (see Introduction). But if it can be made coherent with respect to self-deception, then the onus is on the proponents of such a thesis to show what they have thus far failed to show. Nonetheless, in light of the difficulties, any endeavour to make the notion of intending to believe falsely coherent does not appear promising.

Now if Siegler's, together with my own, arguments against the 'strong' model (Siegler's terminolgy) of self-deception are correct, then what is the raison d'être behind the use of the term "self-deception"? If self-deception is not logically akin to interpersonal deception, is there nevertheless a point in employing reflexive deception ascriptions? Siegler contends that the analogy is somewhat weaker. To say someone is deceiving himself

> One talks <u>as though</u> he did not know better, <u>as though</u> somebody had <u>deceived</u> him into having false hopes. But he does know better (meaning <u>ought</u> to know) and nobody has <u>deceived</u> him and so he should reasonably give up such hopes. When he talks that way he is simply not taking into account what he knows (meaning "<u>ought</u> to know") about such matters. He should know better.[21]

To be sure, this general characterization of reflexive deception ascriptions is at least part of what we do mean when we accuse someone of being self-deceived. Unfortunately, Siegler's specific proposals as to a formulation of that meaning is less than

satisfactory.

Siegler eliminates one troublesome feature from the 'strong' (= Demos') model, namely, that S intends to believe what he knows to be false, and redescribes another, namely, that S holds incompatible beliefs. For the latter, the beliefs may occur at different times, or in one type of situation p is acknowledged while in another type of situation it is denied. For the former, S does not intend to believe what is false, rather "we could say that he actually believes that not-p as a result of a desire that not-p and a fear that p."[22] S 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.[23]

As set out, it appears to be a clearer model than that proposed by Demos and one which seems to admit readily of actual application. But are the phenomena covered by this analysis the kind of phenomena we are prepared to call central cases of self-deception? Siegler is careful in trying to distinguish his version of what it is to deceive oneself from other related phenomena, e.g., mistaken belief, jumping to a conclusion, and so on. But has he succeeded in finding instances of self-deception conforming to his analysis which are signicantly different from them?

According to Siegler, self-deception requires a belief that

not-p which is caused by both a desire that not-p and a fear that p through a misjudging of the evidence. Presumably, the case of the man who has a few days to live, with minor alterations, can be made to fit this analysis. The man accepts the doctor's prognosis that he has but ten days to live. Later he comes across stories relating unexplained cures of terminal diseases which lead him to believe that he will live. His belief that he will live is caused largely by his desire to live and his fear of death. Through the desire and fear he distorts the content and magnitude of the stories. When relating them he tells us that men exactly like him who had had the same disease were cured miraculously, and since he is no different from those already cured he too will live.

On the face of it, this seems to be a case of self-deception since it "differs from a stupid or naive mistake in that none of these features [conditions 1 - 4 above] need apply . . . from jumping to a conclusion caused by a desire in that feature 3 need not apply and the cause need not involve a fear that p . . . from wishful thinking in that there need be no fear that not-p in the causal explanation . . . from serious psychological disorders in that cravings and abhorrences rather than desires and fears cause a distortion at the level of perception rather than at the level of evidence and inference."[24] Nevertheless, Siegler's analysis will prove unsatisfactory if self-deception is a jumping to a conclusion coupled with an aversion to the opposite conclusion, or if self-deception is wishing that something were true compounded with a repression of the thought that it is not true, or, possibly, if it

is not distinct from mild psychological disorders. Certainly, instances of wishful thinking, jumping to a conclusion, etc. are frequently called self-delusory; we do say that someone is in self-deception when we could just as well say that he is engaged in wishful thinking or jumping to a conclusion. If Siegler is right, i.e., that self-deception is but one of these or a combination of these, then it appears that there is nothing which is uniquely called a deceiving of oneself. On Siegler's analysis these diverse applications of the term are not only loose applications but also manifestations of the term's completely surrogate nature. Accepting this way of looking at self-deception suggests that there is nothing to which "self-deception" uniquely applies. It is merely a convenient term serving to group together some related phenomena. But, surely, if this is true, "self-deception" does not appear to have a significant function. And the notion itself ceases to be of special interest.

The move to such a position can be checked by noting an important factor in self-deception to which Siegler gives scant attention. While it is probably true that "The person in self-deception either speaks or acts on the belief that p, but the state of self-deception is temporary and liable to be shattered by an unexpected event, e.g., an accusation of self-deception or a clear revelation of the falsity of p, etc.,"[25] by this Siegler seems to be suggesting (and his analysis does not rule it out) that the self-deceiver can be brought out of his self-deception by the mere mention of the error or by pointing out something which chal-

lenges the mistaken belief. Rather something to the contrary seems to be the case. What distinguishes self-deception from other closely related notions is that the self-deceiver is blind to the obvious. When the contrary evidence of which the self-deceiver is fully aware clinches the matter, he remains unmoved. Surely this resistance to the significance of the facts is what, if anything, distinguishes genuine cases of self-deception from cases where the term is loosely applied. There is something we cannot understand about the self-deceiver, 'how can he believe <u>that</u>!' We are impatient and intolerant because we know he is equipped to assess the situation as we do, but for some reason he does not. Perhaps this is all that the (alleged) paradox amounts to. In any case, Siegler's non-paradoxical analysis, though probably helpful in indicating the close relationship between self-deception and other mental phenomena, fails to capture this distinctive feature.

3. Terence Penelhum, acting upon a suggestion by John Canfield and Don Gustafson, develops an analysis which attempts to explicate the paradox in self-deception while avoiding the difficulties inherent in an analysis logically akin to interpersonal deception. Canfield and Gustafson argue that self-deception can best be understood as belief in the face of strong evidence to the contrary. They take as their starting point the treatment of "self-command". An analysis of self-command in terms parallel to that of interpersonal command produces a paradox. If, however, self-command is seen

not as a species of interpersonal command but as a 'making oneself act in the face of certain obstacles' the oddity disappears. Similarly, when self-deception is understood not as a species of interpersonal deception but as belief in the face of strong evidence to the contrary, the prima facie oddity disappears.[26]

While this seems to be true, it is not the whole story. Belief in the face of strong evidence to the contrary is a necessary condition of self-deception, without it "self-deception would be indistinguishable from intellectual indecision,"[27] but it is not a sufficient condition. "The self-deceiver must also know the evidence; or else we have not self-deception but ignorance," and, "if he knows the evidence yet does not accept what it points to, this might be because he does not see what it points to, and then we have stupidity or naïveté; so the self-deceiver must not only know the strong evidence, but see what it points to."[28] So the necessary and sufficient conditions of self-deception are:

(1) belief in the face of strong evidence to the contrary,
(2) the subject's knowledge of the evidence,
(3) the subject's recognition of the import of the evidence.[29]

But conditions (2) and (3) seem tantamount to the self-deceiver's holding an opposing belief to (1). Consequently, the paradox (as Penelhum sees it) of ascribing simultaneous incompatible beliefs to the person has merely been restated and not made more intelligible. To understand the paradox, we must, according to Penelhum, understand that the self-deceiver is in a 'conflict-

state' (a term borrowed from Demos) and that in "This way we can settle for consistent description of inconsistent behaviour. Someone in this state does partially satisfy the criteria for belief and also those for disbelief."[30] That is, in recognizing where the contrary evidence points he in effect accepts the conclusion implied and in sincerely declaring a belief which contradicts that acceptance he partly satisfies the criteria for disbelief. So there is reason to say that the self-deceiver believes what he asserts and reason to say that he does not believe what he asserts.[31]

Penelhum's analysis permits the following to be described as self-deception. Consider the case of a man S (a philosopher perhaps) who sincerely contends that Wittgenstein died in 1955. He acknowledges that everyone who knew Wittgenstein in the early 50's agrees that he died in 1951, that biographies list his death as having occurred in 1951, that no one is reported to have seen him in the years between 1951 and 1955, and so on. Despite the overwhelming evidence to the contrary, which S recognizes and accepts without question, S maintains that Wittgenstein died in 1955. Indeed he not only sees where the evidence points but he does not challenge the evidence in any way, nor does he offer evidence on behalf of his own contention.

Surely this man cannot be let off with an innocuous accusation of self-deception. Surely this man is behaving irrationally. He admits and accepts the evidence for p and yet at the same time denies it. If that is all there is to the case, then there should be an explanation of his outrageous behaviour. Does he not really

understand what he asserts and what the evidence means? Has the evidence slipped his mind? Is he unduly obstinate? Does he refuse to admit his erroneous claim? Is he simply irrational? If any answer to these questions is in the affirmative, then the man could not properly be said to be self-deceived; if all are negative, then the paradox has not been avoided — it could not be more blatant. Moreover, even if the case taken is not as clear cut as the preceding, say, one in which the evidence is not overwhelming, the difficulties still arise. While admitting that there are good reasons for believing that p and there are no good reasons for denying p, to claim not-p without reservation and with no attempt to assess or adjust the claim or the degree of conviction with which the claim is made is not to be merely obstinate but to traverse the bounds of rationality. Certainly, obstinacy requires at least some backing for what is claimed. Irrationality, on the other hand, requires none.

Where Penelhum seems to have gone wrong is that he includes, as one of the conditions, the feature that the self-deceiver must know the evidence and see where it points. But this is not in line with what is ordinarily said about the person in self-deception. The self-deceiver is one who is somehow blind to the obvious, for some reason he fails to see what we see, or indeed what he could see if he were not self-deluded. If it were the case that the self-deceiver accepts contrary evidence and the implications thereof without offering or having access to supporting evidence, then it seems that the self-deceiver differs not at all from the irrational

person. However, the self-deceiver is not criticized for denying what he sees, rather he is criticized for failing to see something he ought to see, something he could see if he were not in self-deception. Whatever the self-deceiver is, he is not irrational.

Despite the way Penelhum has handled the original suggestion by Canfield and Gustafson, that self-deception involves belief in the face of strong evidence to the contrary, it is, I think, if not completely correct, at least on the right track. The mistake Penelhum has made is to reintroduce a second opposing belief, or what might amount to an opposing belief. Both Siegler and Penelhum correctly eliminate the supposed intentional aspect in self-deception, namely, that the self-deceiver intends to believe falsely. But they retain the feature that self-deception, like interpersonal deception, consists primarily of two contrary beliefs. And the reason for the failure of both analyses seems to rest on the fact that if two incompatible beliefs are attributed to the self-deceiver at the same time or at different times, then descriptions other than self-deception are more appropriate. Some alternative descriptions are that S unwittingly holds incompatible beliefs, he has forgotten that such-and-such, he is engaged in wishful thinking, he has repressed an unpleasant thought, etc. Perhaps restricting self-deception to one significant mistaken belief will have more beneficial results.

But before embarking on a discussion of a possible improvement upon the previously cited attempts to formulate a satisfactory account of self-deception, a novel approach, which tries to circum-

vent the problems in the epistemic model but which also attempts to come to grips with intentionality by introducing a new vocabulary, deserves some attention.

CHAPTER II

THE SELF-ASCRIPTIVE MODEL

The failure of those working with what has been called the epistemic model in producing a satisfactiory analysis of self-deception has impelled at least one writer to abandon that approach altogether. In the most extensive and impressive treatment of the topic to date, Herbert Fingarette introduces a new model which, he maintains, offers a more comprehensive means of coming to grips with the essential paradox in self-deception and closely related psychological and moral issues imbedded in talk of the self-deceiver.[1]

Fingarette argues that the reason for the failure of the various analyses based on the epistemic model is that the model itself is a cul-de-sac. Paradoxes arise because the automatic comparison of self-deception to interpersonal deception and the usual description of the self-deceiver as perceiving mistakenly suggest that the basic element is belief. But there seems to be no way of reconciling the attribution of conflicting beliefs to the self-deceiver with the essential intentional nature of self-deception. Consequently, those attempts to make self-deception intelligible which employ knowledge and belief have, in one way or another, avoided the intentional aspect, and in its place have concentrated on the less crucial problem of attributing contrary beliefs. And, he says, any analysis which ignores the intention-

ality of self-deception must obviously miss a large part of the activity of self-deception.

His method of attack is to challenge the epistemic model's implicit, uncriticized assumption that self-deception consists primarily in a person's belief(s). His approach is to set aside, by demoting in importance, the role of knowledge and belief while attempting a non-paradoxical account of the intentional feature implied in describing the self-deceiver as one who brings it about that he believes what he knows to be false.

As the starting point, Fingarette proposes a dramatic shift of emphasis in our characterization of consciousness. However, this shift does not mean the elimination of the 'cognition-perception' family of terms, for which there is a natural tendency to say the self-deceiver is "one who doesn't perceive his own fakery, who can't see through the smokescreen he himself puts up, [who] sincerely believes the stories he tells while 'deep inside him' he knows it is not true, [who] makes it appear to himself that something is so, [or who] is unaware of his own deception."[2] Rather it is his aim to divorce "conscious" and its variant forms from this family and to show "by reinterpretation that they would be better treated . . . as members of the 'volition-action' family."[3] The reinterpretation offers a radical alternative to the traditional way of referring to consciousness as passive. If the prominent passive imagery is superseded by an active one, the resulting conception of consciousness when applied to the analysis of self-deception serves to make coherent its puzzling nature.

The easy substitution of "see" for "know", "be aware of" and "be conscious of" has led us to conceive consciousness as "the essentially passive registration and reflection to the 'mind' of what the world presents to our eyes."[4] Against this view, Fingarette proposes a more fertile conception of consciousness as something we do rather than something that happens to us:

> The model I propose is one in which we are doers, active rather than passive. To be specific, the model I suggest is that of a skill . . . The specific skill I particularly have 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 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.[5]

The key element in self-deception which replaces belief becomes, then, 'spelling-out' or 'becoming explicitly conscious of' one's engagement in the world. An individual's engagement in the world refers to

> what someone does or what he undergoes as a human subject; it is how an individual finds and/or takes the world, including himself. It is a matter of the activities he engages in, the projects he takes on, the way the world presents itself to him to be seen, heard, felt, enjoyed, feared, or otherwise 'experienced' by him. It is logically necessary that it should be typical of our description of an individual's engagement in the world that the description be cast in terms of such categories as aims, reasons, motives,

> attitudes, and feelings, of understanding and 'perception' of the world and himself.[6]

Thus, Fingarette contends, the things about which a person can be self-deceived expand to include aims, reasons, motives, etc. as well as beliefs. But aims, reasons, motives, etc. all involve beliefs. And, presumably, self-deception concerning any one of these also involves just those beliefs. So if his claim is right, another argument is required to establish it.

Part of what it means to be in self-deception is that the self-deceiver fails to spell-out a part of his engagement. Spelling-out is the "exercise of a specific skill for a special reason . . . Exercise of the skill requires sizing up the situation in order to assess whether there is adequate reason for spelling-out the engagement. And the corollary of this is that in exercising the skill we are also assessing the situation to see whether there is reason <u>not</u> to spell-out the engagement."[7] There will be instances, then, when there is an overriding reason not to spell-out one's engagement.

So the definition of self-deception is:

> In general, the person in self-deception is a person of whom it is a patent characteristic that even when normally appropriate he <u>persistently</u> avoids spelling-out some feature of his engagement in the world.[8]

But spelling-out is itself a way of being engaged in the world; we can spell-out or refrain from spelling-out the fact that we are spelling-out or not spelling-out. So, for the avoidance of the first-order spelling-out to be successful, the second-order spelling-out must also be avoided, since spelling-out that one is not

spelling-out requires that the engagement is spelled-out. Hence the self-deceiver's not spelling-out "is the adherence to a policy (tacitly) adopted."[9] And "the adoption of the policy of not spelling-out an engagement is a 'self-covering' policy."[10] The policy is such that the self-deceiver does not even spell-out the engagement to himself; "when the issue is raised, he does not, cannot, express the matter explicitly at all. He is in this respect in no better position than anyone else. He tells us nothing but what he tells himself."[11]

Having adumbrated the basic workings of self-deception, the next step is to explain why someone would avoid spelling-out an engagement when it is appropriate to do so. We have, up to this point, the external signs of self-deception; what we need now is an understanding of the self-deceiver's refusal to spell-out an engagement.

Throughout life, each individual identifies himself as a person in particular engagements. This spelling-out of an engagement constitutes one's personal identity. When an individual performs such a spelling-out, he is spoken of as having avowed the engagement as his. Since avowal is the defining of one's personal identity, there is a peculiar power associated with it:

> in speaking of avowal and acknowledgement we are concerned with an acceptance by the person which is constitutive, which is de jure in its force, which establishes something as his for him.[12]

And the ability to avow is what promotes the emergence of the person

from the individual; it is a necessary condition of being a person:

> If there were no such things as a person's acknowledging some identity as his and certain engagements as his, and disavowing other identities and engagements, there would be neither persons nor personal identity.[13]

As one matures, as the person emerges from the individual "there is a tendency for increasing correlation between what is avowed by the person and the actual engagements of the individual."[14] It is here that self-deception becomes possible. When tension arises between the engagement someone is in and the identity he avows, there is the potential for self-deception. We have, then, an answer to the question, why the self-deceiver adopts a policy of not spelling-out an engagement:

> an individual will be provoked into a kind of engagement which, in part or in whole, the person cannot avow as his engagement, for to avow it would apparently lead to such intensely disruptive, distressing consequences as to be unmanageably destructive to the person. The crux of the matter here is the unacceptability of the engagement to the person. The individual may be powerfully inclined towards a particular engagement, yet this particular engagement may be utterly incompatible with that currently achieved synthesis of engagements which is the person.[15]

Having freed ourselves from the epistemological paradox, we are also free from the moral paradox. It is no longer a consequence of imputing self-deception that the self-deceiver is accused of lying to himself, he is not both guilty deceiver and innocent

deceived. He is a victim, in much the same way a neurotic is the victim of his neurosis.

Fingarette discusses at length Sartre, Kierkegaard and Freud on the topics of self-deception, consciousness, personal identity and psychoanalytic theory. In each he sees a corroboration of his own position, which in turn, he claims, elucidates what tends to be unclear in them. However interesting the relationship is among them, his discussion of that relationship is, I believe, tangential to his main proposals. So I shall restrict the scope of my attention to the analysis as so far presented.

We now should have an outline of the model Fingarette proposes together with some related points concerning consciousness and avowal, i.e., avowed personal identity. Further details will be brought out as the adequacy of the model is examined.

The self-ascriptive model is undoubtedly an original approach to the problem of self-deception. But whether it succeeds at what it sets out to describe I am less certain of. Some obscurity surrounds the crucial notion of spelling-out, because, I think, Fingarette is not too careful in his explication of it. Nevertheless, I shall attempt to make it less troubling by ignoring some seemingly inconsistent remarks which serve to make it a more obscure notion than it need be. However, even if it can be made pellucid, the model does not preserve the element of intentionality, nor does the analysis seem to be one which we would be prepared to call unreservedly an analysis of self-deception.

To begin with, Fingarette claims to offer a radically new model of consciousness against the traditional 'passive-visual' model. This new 'active-skill' model, he says, will prove more fertile for philosophical analysis of mental phenomena. But apart from self-deception, one is left guessing what other benefits it might have, if any.

Fingarette is not incorrect in saying there is an 'active' element in consciousness, that conscious beings are not simply bombarded with sensory stimuli; because they have some control over what they percieve in selecting and attending to features of their environment. But whether this proposed model will satisfactorily characterize our notion of consciousness is another matter.

One difficulty with Fingarette's version which might prevent it from doing so is manifested in the confusing term "explicit consciousness". As described by Fingarette, the notion of explicit consciousness is somewhat muddled. Initially it looks as if it is a specific term focusing on one aspect of consciousness, but other remarks lead one to believe that Fingarette intends this special term to be not only directly applicable to the analysis of self-deception, but that it also coincide with the ordinary meaning of the word "conscious". And in attempting to make "explicit consciousness" both have a specific function and not deviate from the ordinary uses of "conscious", Fingarette fails to accomplish either.

We are told first of all that becoming explicitly conscious of something is the exercising of the learned skill of saying

what we are doing or experiencing. The kind of skill referred to is akin to a linguistic skill. As a description of what one is actually doing when becoming explicitly conscious, Fingarette intoduces the term "spelling-out", which is, he says, to be used synonymously with "becoming explicitly conscious of": "for an individual to become explicitly conscious of something is for him to spell-out an engagement of his."[16] "Spelling-out" suggests that explicit consciousness bears a close relationship to linguistic activity. But (and this is where understanding of explicit consciousness begins to break down) although "the phrase 'spelling-out' may refer . . . to the actual and elaborate saying out loud, or writing down, of that which one is becoming conscious of . . . one often spells-out something, without any evident utterance, even to oneself, or with only allusive or cryptic ones."[17] Spelling-out, then, is a linguistic activity yet it need not be. I may spell-out an engagement by saying it out loud (to someone else), saying it unvoiced to myself, fully or partially formulated, or by doing something else that is never specified. But how is it possible to attribute an engagement to oneself unless one does something resembling in one form or another a linguistic expression? An answer to that question is required in order to make "spelling-out" a workable notion. Moreover, unless Fingarette supplies a way of determining how spelling-out is possible without any form of linguistic expression obtaining, there appears to be no discernible difference between not spelling-out and spelling-out "without any evident utterance even to oneself". Indeed, what could count as

a criterion? Behaviour, perhaps. But if behaviour alone is sufficient to establish spelling-out, it is not too clear why the term was introduced in the first place.

Fingarette is deliberately vague about what the notion amounts to. Although "the phrase 'spelling-out' is intended to suggest an activity which has a close relation and analogy to linguistic activity . . . What the exact connection is between spelling-out and perfectly straightforward examples of linguistic activity, I do not know."[18] But that confession will hardly do to persuade one that the difficulties involved with spelling-out can be met. So, I shall attempt a reading of "spelling-out" which is consistent with some of Fingarette's remarks and, I think, with the specific use for which the term is intended, namely, in an analysis of self-deception.

It would seem to be a more coherent and useful reading of "becoming explicitly conscious of" and "spelling-out", in keeping with their intended application to self-deception, if linguistic expression, voiced or unvoiced, is regarded as both a necessary and sufficient condition. After all, if spelling-out is a learned skill, it seems to be quite naturally tied to the ability to formulate linguistic expressions.

If spelling-out is not restricted solely to linguistic expression there is the following difficulty to overcome. In a discussion of a distinction between 'strong' (= to be actually engaged in spelling-out) and 'weak' (= to be readily able to spell-out) senses of "conscious", Fingarette says both that "To spell-

out . . . is to be explicitly aware of; it is to pay conscious attention to"[19] and that to be conscious in the 'strong' sense is "essentially . . . the exercise of our skill in making explicit, in linguistic or closely related form."[20] But, for example, children and some adults who lack linguistic competence pay conscious attention to many things which they are totally unable to describe. If this is right, then we must drop either the equating of spelling-out with paying conscious attention to or that linguistic expression is a necessary condition of spelling-out. If the latter, then the usefulness of the term seems to have been lost, for it appears that the only method of determining whether someone is spelling-out would be to determine first whether he is paying conscious attention to something. If the former, then spelling-out fails to adequately account for a large and significant part of what it is to be conscious (of oneself).

The reason, I think, for Fingarette's equating spelling-out with paying conscious attention to, which conflicts with the relation between spelling-out and linguistic expression, is his yielding to the temptation of trying to make his 'active' model of consciousness cover more ground than it is capable of handling. But rather than reject Fingarette's proposal at the start, I shall glean what I believe to be essential to spelling-out in its application to the problem of self-deception. Contrary to Fingarette's objective, I do not believe his model of self-deception depends upon the 'active' model of consciousness. The latter seems to be unnecessary excess baggage.

Normally, there are reasons for spelling-out as there are reasons for speaking. "Skill in speech calls for assessing just when to speak, when not to speak, how to speak, what to say. Skill in spelling-out requires analogous assessments."[21] We know there are occasions when to refrain from speaking is preferrable to saying anything at all. Similarly, sometimes it may be beneficial to one's own well-being or peace of mind not to spell-out a feature of one's engagement. When one persistently avoids spelling-out some feature of his engagement, especially when it is appropriate to do so, he is, according to Fingarette, in self-deception.

Now I think we can understand what sort of phenomenon Fingarette has his finger on without appealing to the 'active' model of consciousness. He says in the paragraph where the definition of self-deception occurs that "The self-deceiver is 'unable' to admit the truth to himself."[22] And in a subsequent chapter he describes the self-deceiver further as "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."[23] The operative words here are "admit" and "acknowledge", not "conscious" and "consciousness".

It is not uncommon in our everyday mental activity to force an unpleasant thought from our attention. If some unpleasant thought about our past, present or future frequently comes to mind, it is normal to devise a method of getting rid of it. While such activity is not itself self-deception, it can easily develop into self-deception. When the person is so successful in repressing the

offending thought that he avoids it altogether, when it is no longer in his immediate repertoire of thoughts, and when it is important that he does acknowledge or admit it, then he is, according to Fingarette, self-deceived.

How spelling-out enters the picture is this: If thoughts relevant to self-deception are assumed to be intimately connected with language, then any thought of this kind that enters one's mind will take the form of a voiced or unvoiced linguistic expression. So, to think about one's engagement in the world is to describe it in words, i.e., to spell-out the engagement. The policy the self-deceiver adopts of not spelling-out an engagement refers to the method of avoiding the unpleasant thought; means are devised by which the thought never or very seldom has the opportunity to occur. This account is admittedly vague and not very helpful in fully explaining what sort of activity repression an unpleasant thought is. But the phenomenon is so complex and demanding of a discussion reaching beyond the scope of the issue at hand that it will not be dealt with at greater length here. Nevertheless, since it seems to be familiar enough, there should be no great difficulty in understanding what kind of thing is being talked about.[24]

I have two main objections to Fingarette's analysis, or at least my interpretation of it: (1) apart from certain unclarities in the account itself, there is no good reason to conclude that repressing an unpleasant thought is, as Fingarette argues, intentional, and (2) that there is good reason for ruling it out as an

analysis of self-deception per se.

Fingarette claims to have preserved the paradox in self-deception through a quite different way of describing it. He says that "If our subject persuades himself to believe contrary to the evidence in order to evade, somehow, an unpleasant truth . . . then and only then is he clearly a self-deceiver,"[25] "the deep paradox of self-deception lies . . . in the element of . . . intentional ignorance,"[26] "the essence [is] that the self-deceiver purposefully brings it about that he is self-deceived."[27] Notice the shift from expressing the paradoxical element as intentional to purposeful. This shift, I think, indicates the failure of Fingarette's model to come to grips with the supposed intentional aspect of self-deception. For although the two concepts overlap, not all purposive behaviour is intentional. However, that distinction will need some fleshing out.

Intelligent behaviour, although not always intentional, is subject to teleological description. For instance, I wish to make a habit of a routine set of movements so that I can perform it 'without thinking' to allow concentrating my attention on something else. In this case, the habit is acquired as a result of my decision. On the other hand, I may unconsciously acquire a habit simply by having performed the same set of movements any number of times until it becomes 'second nature'. Normally, habits are formed to free ourselves from attending to commonplace, oft-repeated actions. There is a point to the action which becomes habitual, because it is usually one which is routinely necessary.

And there is a point to acquiring the habit because the automatically performed action saves time by freeing our attention for other matters. Habits, then, even those acquired unwittingly, are amenable to teleological description, and thus are purposive bits of behaviour.

Now an example illustrating that not all purposive behaviour is intentional is the following. It happens occasionally that a set of movements is performed through force of habit when the object of the habit is neither appropriate nor desired. I resolve to smoke no more cigarettes until the next day. In a short while I become immersed in thought about, say, an upcoming event. Moments later, much to my chagrin, I discover myself smoking a cigarette which I have just lit unawares. The action was therefore unintentional. But since the action is an instance of a habit which is itself purposive, I was engaged in a purposive activity. So, even if Fingarette is successful in showing that his analysis of self-deception implies purposiveness, additional argument is needed to show that it also implies intentionality.

To return to self-deception. The self-deceiver avoids spelling-out some feature of his engagement in the world. The avoidance amounts to a tacitly adopted policy. That is, the self-deceiver, in assessing the situation, finds an overriding reason not to spell-out a particular feature of his engagement, and this commits him to the policy of not spelling-out. In general, the reason for not spelling-out stems from the self-deceiver's lack of "courage and a way of seeing how to approach his dilemma without probable disaster

to himself."[28] But the self-deceiver takes cognizance of neither the assessment nor the policy: "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 not to spell-out."[29] The policy adopted, then, is 'self-covering'; it is not made explicit. And thus the activity leading to self-deception is not one of which the self-deceiver is aware, for if he were aware of the policy about to be adopted it would "require spelling-out the engagement at issue, etc." But if this is so, then the self-deceiver could not be correctly described as deciding or intending to put himself into a state where he is unable to recognize something about himself. A condition of intending or deciding is that the person so engaged is aware of the intention or decision; or if unconscious intention or decision is possible, then the person could, in principle, be aware of it. But the policy the self-deceiver adopts precludes just such an awareness. So the state of being self-deceived is something that happens to the self-deceiver, not something he intends to or sets out to be in. This conclusion contradicts Fingarette's pronouncement that the self-deceiver has intentionally adopted the policy, but he offers little or no argument to establish that. Indeed from the way he describes self-deception, quite the opposite is implied.[30]

There is a further difficulty with the self-covering nature of the policy. Presumably, the policy steers the self-deceiver away from and around situations in which he would be forced to spell-out the displeasing engagement. The policy requires that the self-

deceiver be unaware of a feature of his engagement; but if that is so, then he must also be unaware of (the nature of) the situations which would force him to spell it out, for to recognize the nature of the situation would be to spell-out the engagement. Now if he does not recognize the situations to be avoided, how does he manage to avoid them? Perhaps the recognition is subliminal. But how this might occur remains a mystery.

There is more that is curious. It seems that Fingarette's model is meant (in part) to explain how the self-deceiver constructs a 'protective shield' against what might bring the unwelcome thought to his attention. But consider this. To test any candidate self-deceiver, we spell-out the offending engagement to him. If he understands the import of our remarks (and there is no reason to think he will not), then he must attend to what we have said. But if he does that, then he has violated the alleged policy. If it is a policy, it must be a very weak one indeed.

Although having failed to accomplish one of his aims, Fingarette has not by virtue of this failed to accomplish the aim of isolating a phenomenon which is clearly and uniquely self-deception. The accomplishment of this aim requires the absence of more suitable descriptions. But he obligingly supplies some apt alternatives:

> the self-deceiver appears before us as the neurotic, as the victim of the compulsive force of the unconscious, as a sufferer from mental illness.[31]
>
> . . .
>
> the inability to spell-out, because of its importance

> for sophisticated planning and assessment of complex engagements, leads to a profound further loss of self-control.[32]
>
> . . .
>
> the [self-deceiver] is to be pitied for the mental 'breakdown'. There emerges the 'medical' view of self-deception as helplessness due to 'mental pathology'.[33]

This interpretation of self-deception as a sickness forces us to view the self-deceiver non-judgmentally. We neither praise nor blame him, for he is someone to be pitied. Fingarette does not intend these descriptions to be figurative, because he encourages us to look at the self-deceiver as one in need of psychotherapy. In fact, seeing self-deception as the problem of self-acceptance, he considers, it seems, all forms of neuroses as analyzed in Freudian terminology to be cases of or border on self-deception. He acknowledges strong parallels between his account of self-deception and "character disorders, psychoses, and other kinds of pathology rooted in severe ego and superego deficiencies."[34] Thus, the self-deceiver as patient must be carefully liberated from his deception to the point where he accepts the personal identity he has, while avoiding the trauma associated with that emergence.

If the self-deceiver is indeed a neurotic or psychotic, it appears that Fingarette's original way of describing self-deception as an 'inability' to admit the truth to oneself[35] moves from the figurative to the literal. The person in self-deception would be not simply 'unable' to admit the truth, he would actually be <u>unable</u> due to rather serious psychological disorders. But this

violates the distinction made between the person who fails to see, is blind to the obvious, who can correct his mistake but for some reason does not, and the person whose incorrect judgments and delusion stem from mental imbalance. Surely, the person 'afflicted' by self-deception does not require therapeutic treatment. He does not suffer from a mental disorder; his judgments are systematically normal except that he does not recognize what we recognize and what he is capable of recognizing. The self-deceiver qua self-deceiver is capable of recognizing his error; the psychotic qua psychotic is incapable of recognizing his. This rather clear distinction between self-deception and mental disorders is pointlessly blurred if self-deception is considered a sickness and amenable to psychoanalytic interpretation.

Of course, accepting the self-ascriptive model does not commit one to Fingarette's view of the self-deceiver as one suffering from serious psychological disorders. But even on the baldly stated version of Fingarette's analysis, it does not appear to locate straightforward cases of self-deception.

In wishful thinking, pleasant thoughts are summoned up and dwelt upon. The kind of phenomenon Fingarette's analysis covers is, I think, complementary to this activity. Often they go hand-in-hand. Those uncomfortable facts about ourselves or inevitable distasteful experiences, when excluded from our attention make way for pleasant and comfortable ones. While this relation between repressing an unpleasant thought and wishful thinking is not itself a conclusive reason for denying the former status as a central case

of self-deception, it does suggest that Fingarette has not found a distinct kind of mental phenomenon rendering self-deception a unique mental state. But, of course, more support is needed for this claim — and that brings the discussion to the positive account to be presented in the following chapter.

CHAPTER III

THE EPISTEMIC MODEL REVISED

We have not progressed much further than the original assumption that self-deception consists, in part, of a mistaken belief and a capacity to recognize that mistake. Additional features proposed in various forms have, for one reason or another, failed to satisfactorily capture the essence of self-deception. Prominent among them are: (a) that the activity of self-deception is intentional, and (b) that the self-deceiver holds conflicting beliefs. Taking both (a) and (b) to be characteristic of self-deception is a result of seeing self-deception as analogous to interpersonal deception. It is thought that if self-deception is a deception then the analysis of interpersonal deception will shed light on intrapersonal deception. But this does not appear to be the case.

Any analysis of self-deception incorporating (a) will inevitably fail. No state of affairs could fulfill the condition that the self-deceiver intends to believe erroneously, not because of any special difficulty in determining the existence of intentions and beliefs, but because the proposition itself involves irreparable muddles concerning the nature of belief. It could be someone's wish that p were true, but it could not logically be his intention to believe that p is true while recognizing that p is false. Since beliefs are representations of reality, we can no more have beliefs which we take to be incorrect representations than want or intend

to have such beliefs. So if this feature, as some would maintain, were of the essence of self-deception, it would not and could not exist. Indeed, those analyses, notably Demos' and Fingarette's, that attempt to resolve the incoherency, have, necessarily perhaps, paid attention to problems slightly off the issue. Demos seems to lose sight of (a) and instead concentrates on (b). Fingarette, in trying to 'translate' (a) into a more coherent alternative to the formulation found in the epistemic model, seems to think that in establishing purposiveness he also establishes intentionality. He succeeds only in the former, but then all or most intelligent behaviour can be described teleologically.

The problem with (b) is somewhat different. While simultaneously ascribing two incompatible beliefs to someone is odd, if not paradoxical, it is not the type of oddity found in (a). There is a prima facie oddity associated with someone holding inconsistent beliefs, but it seems to be an oddity arising from the fact that logical consistency is a condition of rationality, and not that the holding of two incompatible beliefs is logically impossible. Consequently, the paradox or oddity in (b), unlike that in (a), is resolvable, providing an explanation of how the person came to hold incompatible beliefs is available.

Although (b) is easily assimilated as a feature, its inclusion, I believe, undermines what the analysis is intended to describe. Once the self-deceiver is said to hold inconsistent beliefs, there is a difficulty in keeping self-deception distinct from other related mental phenomena. It no longer appears to be self-deception

that is under investigation but something verging on, and more aptly described as, holding inconsistent beliefs, engaging in wishful thinking, etc.

(b) issues from the assumption that self-deception is a conflict state. It is typical of the analyses so far discussed, namely, those be Demos, Siegler, Penelhum and Fingarette, that the self-deceiver is pictured as fraught with mental conflict. Or if it is not a conflict he feels, it is one we feel in accounting for his beliefs. But since the mental duplicity feature does not lead to straightforward cases of self-deception, perhaps it too, along with (a), should be eliminated.

Dropping both (a) and (b) as elements of self-deception serves to remove self-deception that much more from interpersonal deception. And, as we have seen, the aspect of self-deception which distinguishes it from other closely related phenomena, is the self-deceiver's blindness to the truth or, as I shall interpret that metaphor, his resilience to recalcitrant evidence. Now the analysis suggested by Canfield and Gustafson's contention that self-deception is essentially belief in the face of strong evidence to the contrary avoids the pitfalls of both (a) and (b), and provides the basis upon which an explanation of the self-deceiver's resilience to recalcitrant evidence can be constructed. Beginning with and developing that suggestion will, I believe, produce the central elements of (via central cases of) self-deception.

It is the job of conceptual analysis to isolate those elements illustrating a central case. But in analyzing self-decep-

tion there is a special difficulty brought about by its loose application to diverse but closely related mental phenomena. Thus, it is incumbent upon any proposed analysis to not only present the central elements but also to explain by virtue of that analysis why so many different mental phenomena are readily, but not altogether accurately, labelled instances of self-deception.

Another significant reason for the analyses so far discussed of not having adequately explicated the concept is their almost total neglect of actual cases (or at least failing to treat the ones they do on more than a superficial level). Self-deception is such a slippery concept that failing to deal with fully worked out examples will permit theorizing to become detached from the practical use for which it is intended and, as we have seen in the previous analyses, to fall short of the goal. Close comparison between the analysis and fairly detailed examples should help prevent a loss of touch with the subject matter.

My strategy will be the reverse of the previous ones. Rather than working out a set of elements and then searching for a case it fits, I shall present what I take to be two non-paradoxical cases whose salient features will exemplify the essential elements of self-deception. And if the set of elements elicited by this method is correct, then it will show how self-deception bears a similarity to other mental phenomena with which it is often loosely equated.

Consider the following.

A man is in love with a woman who is also in love with him. At this time he not only believes but knows that she loves him. He is in possession of a great deal of evidence (much of which is only accessible to lovers) attesting to her sincerity and affection. So powerful and important is the relationship in his eyes that he stuctures his entire life around it. Her peronality, wants, desires, welfare, etc. dominate his every decision for the present and the future. This love affects his life so completely that every plan made takes his relationship with the woman into account. It is his chief and virtually sole concern.

Later, however, she suddenly and inexplicably falls out of love with him. Fickle, yet not insensitive to his feelings, she postpones telling the truth in order to spare him from what would most likely be a severe psychological blow. Unwilling to upset him, she tries as subtly as possible to convey her change of heart and break away. She uncharacteristically cancels a date, complaining of a headache. She claims to have forgotten his birthday. She becomes generally less affectionate. She accepts clandestine dates with other men, hoping he will somehow hear of them.

Meanwhile he is not oblivious of heranomalous behaviour. Though admittedly strange, it can be explained away. After all, people do have headaches occasionally, and they are prone to forget some otherwise well remembered events. It happens that sometimes one meets a member of the opposite sex in a public place. Why should she be any different? As for her seeming coldness, sometimes

the person of whom we are most fond is the person we least want to see. Just possibly she is in a state of malaise. Nothing to worry oneself about; she will be back to her old self in a few days.

Stage 1 At this point the man is fully self-deceived. He is in self-deception not because he sees where the evidence points and rejects it, but because the possibility of her having fallen out of love with him never enters his mind. While someone judging normally would treat the evidence as an ominous sign, he does not even begin to question her constancy. A friend, of both perhaps, who is in a good position to witness her behaviour towards the man notices a distinct change in her affection for him and is perplexed by the man's apparent unawareness; he says to himself, 'I can see it, why can't he?'. It can be said that although her behaviour does not entail a loss of affection, it means that there can be no question that she has fallen out of love with him. But it simply never occurs to the man that she might wish to sever their relationship. In fact, that possibility is so remote and unthinkable that he dismisses her uncharacteristic behaviour blithely, almost offhandedly. Her behaviour means nothing more than what it is, i.e., as he interprets it. And since he has overwhelming evidence confirming her love, there is no room for the countervailing hypothesis to establish itself.

Stage 2 He is untroubled until much stronger evidence comes to his attention. When he hears that she was seen not once but several times at a local entertainment spot accompanied by another man, he becomes plagued by a distressing thought. Maybe, though

unlikely, she loves him no longer. While he still tries to muster convincing explanations, they begin to lose their once unassailable adequacy. It becomes more and more difficult to devise plausible explanations as further evidence comes to the fore. Doubt begins to creep in. He tries to maintain his unbending faith in her fidelity, but his projected reasons for her behaviour becomes less and less persuasive. He tells himself (and possibly others) that she is still in love with him, and may publicly shun all thought of her having fallen out of love as pure nonsense. Yet whatever he does, he cannot avoid that frightening prospect. The mental torment and uncertainty continues until she breaks her silence and announces that their relationship has ended, or until he comes across conclusive evidence for her loss of love. Then, and only then, he chastises himself for having been such a fool in not having recognized it earlier, i.e., in Stage 1.

But self-deception need not always involve direct reference to personal relationships or the self-deceiver's character. An indication of this is the following case. A physicist holds a scientific statement P to be true because it accounts for a great deal of data which otherwise could not be easily handled, and there is no reason to think it is not correct. Operating on this assumption, he formulates new, comprehensive theories embodying P and inaugurates major research projects to investigate their implications. Nevertheless, it is discovered later that certain consequences entailed by P fail to obtain, and that P will have to be rejected and replaced by the more complex statement Q.

Stage 1 While his colleagues are excited by the recent findings challenging P, they are amazed that a fellow physicist could remain unperturbed. He suspects that the experiments showing negative results were somehow badly conceived and executed, and so he continues his research unruffled. The counterevidence is by no means conclusive, but the physicist does not even begin to appreciate that P is endangered.

Stage 2 But as more and more experiments are performed and more decisive evidence falsifying some statements following from P is accumulated, as it becomes less and less reasonable to hold on to P, the physicist belatedly begins to have slight doubts about P and so begins to emerge from self-deception.

This example differs from the previous one primarily in that the belief in question does not bear on one's personal life directly, and the belief is not at one time true but false all along.

In Stage 1, both men are in self-deception; in Stage 2, they are not strictly in self-deception but are, in fact, emerging from it. There is no hard and fast distinction between the Stages; they merge gradually into one another. But as the person moves from Stage 1 into Stage 2, descriptions other than self-deception become more appropriate. Thus Stage 1 will give the elements of self-deception, while Stage 2 will provide the links between self-deception and other related mental phenomena already mentioned.

As gleaned from the preceding examples, the elements of self-deception are:

(1) S knows p or is justified in believing p, that is, either,
  (a) p is true, and,
  (b) S is in possession of sufficient evidence for p, or,
  (c) S has good reason for holding p, and there is no good reason for denying p,
(2) S has something at stake in p; the belief is an important one,
(3) At a later time, p turns out or is shown to be false,
(4) S is presented with recalcitrant evidence which he interprets in accord with his mistaken belief; the possibility of error in the form of an alternative hypothesis is not considered,
(5) S is capable of realizing and would realize his error, or at least suspect that something is amiss, if he were in a different epistemic frame of mind.

But why does the self-deceiver not judge the way he should; what prevents him from seeing his error? We need a filling out of condition (5), which, as we shall see, will require an elaboration of condition (2) as well.

There are two candidate explanations for S's imperviousness to recalcitrant evidence — but only the second proves satisfactory.

1. When S knows that p, any state of affairs inconsistent with p is impossible; or if S has a justified belief that p, then not-p is highly improbable. Any piece of seemingly conflicting evidence must therefore be in some way misconceived. For if p is true, nothing whatever could show that it is false. So, as a possible account of the self-deceiver's state of mind: there is no

point in wasting time with evidence which conflicts with p, since p is true. S's mistake, then, is carrying over this assurance to the time when it has no basis.

But this explanation does not prove entirely satisfactory, because it makes no distinction between someone who holds on to a belief irrationally and someone self-deceived. The irrational believer, while acknowledging the evidence that undermines his belief, refuses to draw the appropriate inference. He may feel threatened and, possibly, vulnerable, but nothing could successfully challenge his belief because he cannot be wrong. The self-deceiver, on the other hand, does not feel challenged because he fails to appreciate the significance of the evidence. He is, up to a point, impervious to contrary evidence, inasmuch as he treats it with equanimity. It is, I submit, the nature of the resistance to recalcitrant evidence which differentiates self-deception from irrational belief and in which the essence of self-deception is to be found.

2. The element which must be employed in the explanation is the importance the belief carries for the self-deceiver. Indeed, if the belief in question were a trivial one, there would be no discernible difference between obtuseness, irrationality and self-deception. But what determines a belief's level of importance?

To help distinguish important from unimportant beliefs, the notions of core, intermediate and peripheral beliefs will be introduced. Core beliefs are, roughly, those beliefs upon which a great deal hangs, i.e., upon which many other beliefs depend. Peripheral

beliefs are those beliefs upon which little or nothing at all depends. Intermediate beliefs, as the name implies, are those beliefs upon which more than a little but less than a great deal depends.[1] And as we shall see, self-deception typically involves a firmly entrenched intermediate belief.

Every belief a person holds will fall into one of these three roughly distinct categories of belief. And, so, as level of importance is a matter of degree, we might expect that there be no rigid line separating core from intermediate from peripheral beliefs.

If one's system of beliefs could be mapped onto a circle, those beliefs having greatest importance would occupy the central region, those having least importance would line the perimeter, and those with varying degrees of importance filling in the intermediate area between the centre and the perimeter. A belief's position on that circle (i.e., its importance) is determined by the number and magnitude of other beliefs dependent upon it. An example which Jonathan Bennett uses to illustrate levels of sentence dispensability is helpful:

> Accepted sentences of the form (i) 'The temperature of such-and-such a star is such-and-such' depend, for those who accept them, on sentences of the form (ii) 'Temperature correlates with light-emission in such-and-such ways', and these depend on sentences of the form (iii) 'Temperature correlates with mercury-column readings in such-and-such ways', and these in their turn depend on sentences along the lines of (iv) 'Temperature has to do with the obtaining of

> such-and-such sensations.' Rejection of (ii) jeopardises (i) and all that depends on it; rejection of (iii) jeopardises (i) and (ii); rejection of (iv) jeopardises all the other three.[2]

(iv), then, is a belief occupying a place very near the centre; (i), on the other hand, occupies a place removed from the centre.

Whereas the rejection of a core belief ramifies onto many of one's beliefs, involving major readjustments and usually necessitating the adoption of an entirely new system of beliefs,[3] rejection of a peripheral belief involves a revision within a system of beliefs which is sharply limited and largely inconsequential. For example, my belief that A is in the room next door which is based on having heard a door slam a moment ago is immediately abandoned after having investigated and not found him there. Possibly several other beliefs held in conjunction with that belief concerning A and the room will be rejected at the same time. But the revision is restricted to a very small range of beliefs.

Core beliefs include both moral, metaphysical and religious beliefs as well as those beliefs involving less general and less abstract terms, but which are nonetheless as deeply entrenched in one's system of beliefs. Examples of the former are: 'The consequences of an action are always a relevant consideration when assessing an action's moral worth', 'Every event has a cause' and 'There is a God'. Examples of the latter are: 'My name is such-and-such', 'I have never been to the moon' and 'I have lived longer than five minutes'. Possibly less obvious core beliefs than the

first three, the latter three are suggested by Wittgenstein's discussion of knowledge and certainty.[4] While he does not make the distinction between core and peripheral beliefs per se, what I have to say about the distinction seems to be in line with his treatment of those propositions which are and are not subject to doubt. There are propositions amenable to doubt, he says, because there are propositions immune to doubt. Those propositions normally immune to doubt, or which it normally makes no sense to doubt (e.g., 'My name is such-and-such'), provide the bases for all language-games, indeed for all thought. They are immune to doubt because they make doubt possible. As groundless assumptions, they provide grounds by which all other judgments are made; they determine what grounds are, how reasoning is to go on. And without something presumed to be certain, there is only insanity. If it would make sense to doubt my name is L.L. when I am neither drugged, nor hynotized, nor psychotic, etc., then it would make sense to doubt that I know the meaning of the words used to express that doubt, and make sense to doubt everyone of my core beliefs. But if that were the case, it would no longer be possible to make sense of anything. Everything would collapse; the world would be rendered irrevocably incomprehensible.[5]

At first glance, the Wittgensteinian explication and my previous explication of the notion of a core belief may appear to differ on the nature of revision. The Quinean-type model, I initially proposed, holds that no belief is immune to revision; while Wiitgenstein seems to be saying the opposite, that the only

alternative to denying some core beliefs is total absurdity, which is no alternative at all. But perhaps they are not inconsistent. Other comments by Wittgenstein indicate that while it is logically possible that I do not know that my name is L.L., I cannot be wrong about this — we take it for granted that people know their own names with the greatest certainty. Although the previous explication agrees that it is logically possible for any belief to be rejected, it says next to nothing about the practical aspects of rejection, about which, I think, Wittgenstein says a great deal. His remarks show that some beliefs are indispensable to us as human beings engaged in certain activities (e.g., speakers of a language), and that there are limits to what will serve as core beliefs for us as creatures of a certain kind living in a world of a certain kind. So, we might add to what has already been said about the nature of core beliefs, that they function in upholding what else one believes, that they are indispensable for the preservation of one's present system of beliefs, that there are some core beliefs essential to any working belief system, and that core beliefs are not subject to doubt within a system.

However, the claim that core beliefs are totally indispensable seems to cover only the type of core belief that Wittgenstein mentions. Moral, metaphysical and religious core beliefs, I think, not only are not logically immune to revision but they and their contraries can be either accepted or rejected without absurdity. Certainly, the world is not rendered absurd when a theist becomes an atheist (or vice versa). And it is arguable that the

same applies to moral beliefs, though I am less sure whether it also applies to metaphysical beliefs. In any case, this points out, I think, a distinction between the core beliefs cited by Wittgenstein and moral, metaphysical and religious core beliefs, which is, in effect, that these two kinds of core beliefs are mutually independent.[6]

Now to apply the above to the problem of self-deception. The belief to which the self-deceiver misguidedly adheres lies near his core beliefs. The belief is predominant; it affects significantly a great many of his beliefs concerning himself, others and the world. As such, it controls a large part of the system of beliefs adopted. It provides, as it were, the foundation upon which less central beliefs are built. So, as core beliefs are immune to doubt, the self-deceiver's intermediate belief, by lying near his core beliefs, is to a degree immune to doubt.

To understand the self-deceiver's imperviousness to recalcitrant evidence, it is helpful to first understand how a theist, say, handles objections which can be considered analogous to recalcitrant evidence.[7] The believer in a benevolent God, when confronted with the horrific instances of evil in the world apparently challenging his belief, will, unable to satisfactorily reconcile them, invariably accept the existence of evil as the will of God, as a mystery, etc. Similarly, the self-deceiver's belief is not disturbed by some recalcitrant evidence. He is not disturbed because just as there is no room open for doubt in the theist's system of beliefs to upset his faith, so there is no immediate room

open in the self-deceiver's system of beliefs to allow entertaining the possibility that he is mistaken.

The peripheral beliefs are such that the possibility of a revision is always a lively option. The system allows for alternative beliefs to replace those peripheral beliefs found incorrect while the system itself remains intact, because the rejection of a peripheral belief does not directly endanger any core belief. However, alternatives to any core belief are not options within the system itself, for the revision of any one of the core beliefs necessitates the establishment of a new, distinct system. Considering the number and magnitude of beliefs connected to any core belief, this should not be surprising. Thus, a system of beliefs has a built-in resistance to that which challenges it, that is, it protects its core beliefs. Intermediate beliefs share the features of both core and peripheral beliefs, but to a lesser degree. Rejection of a belief which lies between the core and the periphery is possible within the system, but due to this position, revision is less easy than that found with peripheral beliefs. And though the rejection of an intermediate belief does not require the adoption of a new system, it does involve major adjustments. So, as with core beliefs but not to the same degree, there is a resistance to the change of any intermediate belief — the degree of resistance being directly proportionate to its proximity to the core.

An intermediate belief will be rejected or a system of beliefs replaced, when the belief or system of beliefs becomes so

beleaguered that the only recourse is to reject it in favour of another. This, I think, explains whay the self-deceiver (and the theist) is resilient to recalcitrant evidence but only up to a point. In Stage 1, the contrary evidence, not overwhelming, does not attack the belief, does not render the belief and those beliefs dependent upon it so inaccurate as to warrant alteration. In Stage 2, as the contrary evidence increases, as the person begins to emerge from Stage 1, the process of rejecting the erroneous intermediate belief and all that depends upon it is indicated by the perplexity and mental turmoil experienced by the self-deceiver.

But what is the nature of the emergence from self-deception? If the self-deceiver is at one time resilient to recalcitrant evidence, why isn't he always?

Each belief near the centre of a belief system will be the anchoring or intermediary link in at least one chain, but most likely very many chains, reaching the periphery. And just as removal of one of the links close to the centre will sever one or many chains, so to, if the links of the several chains leading up to one link, from the periphery to the centre, are removed, then that link ceases to hold anything together and therefore ceases to perform a function; it becomes otiose. (Cf. Wittgenstein's remark: "a wheel that can be turned though nothing else moves with it, is not part of the mechanism.") If a belief is rejected, then those beliefs dependent upon it must also be rejected. And if every belief dependent upon a belief is rejected, then that belief must also be rejected — a belief upon which nothing depends ceases to be a correct representa-

tion of the world.

The self-deceiver's emergence from his deception is characterized by the gradual elimination and replacement of peripheral and not-so-peripheral beliefs culminating in the elimination and replacement of the intermediate belief. At first, in Stage 1, every belief is retained, no relevant peripheral belief is revised, the recalcitrant evidence is treated in a way consistent with the beliefs which follow from the core belief. But, in Stage 2, as the person is confronted with mounting evidence attacking certain of the relevant peripheral beliefs, he gradually rejects them one by one until he must reject the erroneous intermediate belief. When this has been accomplished, the person has completely emerged from self-deception.

There is a temptation to say that the self-deceiver is incapable of recognizing his mistake while self-deceived. But it is also true that the self-deceiver <u>is</u> capable of recognizing his error. This apparent contradiction can be explained away with the help of an analogy. A long-distance runner, who has consistently run the mile in less than 4 minutes, finds that after a short time away from the track he fails on the first attempt to reach his standard. Yet, despite his failure now, we know that with proper conditioning he will soon return to form. We might say, then, that although he has the power to run the mile under 4 minutes, his physical condition at this moment will not allow him to do so. So, we might say that the self-deceiver, although he has the intellectual equipment necessary to realize his error, has a system of

beliefs which does not allow him to immediately recognize it.

So, assuming all the above is right, if there is a paradox in self-deception, it does not reside in intentionality or in conflicting beliefs. Rather the self-deceiver's state is puzzling because what for us appears to be, as a peripheral belief, amenable to uncomplicated revision, is for the self-deceiver a deeply entrenched intermediate belief, immune from doubt as long as the rest of his beliefs dependent upon it can be maintained reasonably and consistently. The tendency to characterize self-deception as paradoxical arises from the fact that we have difficulty understanding why, when both the self-deceiver and ourselves have access to the same information, the self-deceiver does not even begin to suspect his error. Much the same perplexity is experienced by theists and atheists alike, who, in treating the same data in diametrically opposed ways, are bewildered by each other's conclusions.

This way of picturing self-deception eschews talk of it in terms of intentionality and internal conflict, but it substantiates at least one of the many metaphors connected with self-deception, namely, that the self-deceiver is blind to the truth which he could see if only he would look. He does not see what we see in the relevant pieces of information. It is as if he were temporarily blinded by his belief, because although he is capable of appreciating the full significance of the evidence he does not recognize that what he believes is in jeopardy.

This analysis should also keep self-deception distinct from

and show its connection with closely associated mental phenomena.

While self-deception does consist in an erroneous belief, it is distinct from a simple mistaken belief. Holding a false belief may be the result of an error in judgment, stupidity, naïveté, etc. In such cases there need not be a resistance to correction; whereas in self-deception, the self-deceiver has a 'protective shield' against easy adjustment to his belief. Nor need the belief in question be an important one; whereas the self-deceiver has something at stake in what he believes, i.e., a lot depends upon it.

Of course, there are instances of persons holding on to incorrect beliefs and resisting the admission of error through sheer obstinacy. But the kind of resistance here is a product of, say, a refusal to admit fallibility, or a desire to save face. The self-deceiver, however, has no (ulterior) motive for preserving his belief. It would not be correct to say that the self-deceiver, as I have described him, wants to believe such-and-such. I argued earlier that the notion of wanting or intending to believe is incoherent.

Evident from the previous analyses considered in Chapters I and II, there is a definite tendency to speak of the self-deceiver as believing what he knows to be false. But as we have seen, holding inconsistent beliefs is not a feature of a central case of self-deception. Describing the self-deceiver in this fashion seems to arise from characterizing him as one who could or should know better. We feel he must know the truth because he has access

to the same information we have. But it gets more force from Stage 2, in which the person emerging from self-deception, is seemingly caught between two opposing beliefs. Thus, in referring to the process of rejecting the mistaken belief and acquiring the correct belief, it is suggestive (but inaccurate) to say that he believes what he knows to be false.[8] And to repeat, it does not describe self-deception as such, since it refers to Stage 2 in which the person is not in self-deception but emerging from it.

Stage 2 is culminated by the self-deceiver's having sucessfully completed the readjustment to his system of beliefs. But he may not reach that point directly (or ever). Stage 2 can be prolonged by engaging in an activity which avoids the turmoil and uncomfortable admission of error.

If, in the words of Fingarette, he adopts a policy of not spelling-out, not thinking about the turn of events, then he has succeeded in repressing an unpleasant thought. Although the term "self-deception" is applied often to phenomena of this sort, it is somewhat removed from the central features of the concept. It and its complementary, wishful thinking, are frequently taken to be instances of self-deception, but I think their differences warrant a clear distinction between them.

We have a facility for positing and dwelling on pleasant thoughts and a corresponding facility for 'forgetting' or obliterating unpleasant thoughts. When such pleasant thoughts are entertained one is usually not unaware of what is known to be true; in fact it is sometimes just because thought of the truth is so uncom-

fortable that more pleasant thoughts occupy one's mind. The degree to which the activity of wishful thinking commands one's attention ranges from the unimportant to the serious, from harmless daydreaming to a loss of touch with reality. The extent to which wishful thinking has a grip on one's life is indicated, I think, by the degree to which the corresponding unpleasant thoughts are repressed. And it seems the degree to which a thought is repressed is directly proportionate to the difficulty of bringing it to the person's attention and the degree of distress he experiences when he faces it.

Roughly, the elements of wishful thinking accompanied by repression of an unpleasant thought would be something like: (1) S believes (or knows) p, (2) the thought of p is a cause of discomfort, (3) consequently p is 'forgotten' or repressed, and (4) a pleasant thought q takes its place. The person who has repressed a thought may need help to come to grips with it, which would consist of efforts to overcome his mental defenses. But although self-deception is also characterized by resistance, the nature of the resistance is obviously very different.

Other mental phenomena bearing close resemblance to wishful thinking are also labelled self-delusory. Believing in someone or something despite adverse conditions suggests that the loyalty, love or respect for that person or thing supersedes awareness of its deficiencies — there is faith that the person or thing will eventually turn out all right. What differentiates believing-in from self-deception is that in the former the person recognizes the full significance of the unwelcome facts.

It is sometimes said that we deceive ourselves into liking what we truly dislike. For instance, if a condition for belonging to a certain organization is appreciating something a person does not presently care for, he may try to cultivate an interest. Though never quite succeeding with his project, he tells himself and others that he likes such-and-such in order to feel he belongs, in the hope that such declarations will help him to realize his aim. But while feeling more comfortable with the thought that he likes such-and-such and that he belongs, he is not unaware that his real affections belie his behaviour.

It is a mark of all the above mental phenomena connected with wishful thinking, that they are amenable to psychological interpretation and explanation, involving the attribution of desires and aversions. And this description in terms of motives is what primarily distinguishes them from self-deception.

NOTES

Introduction

1 The self-deceiver is also said to be incapable of recognizing his mistake, but this is to distinguished from his ultimate intellectual ability to discriminate correctly. This point will be taken up again in Chapter III.

Chapter I

1 Some of the papers discussed here have also been dealt with in the second chapter of H. Fingarette's Self-Deception, although our treatments and conclusions differ substantially.

2 R. Demos, "Lying to Oneself," p. 588. The combined analysis does indeed include results and intention, but I do not think knowledge need be. Surely all that is sufficient is that B believes that what he tells C is false and that it is in fact false.

3 Ibid.

4 Ibid., p. 591.

5 Ibid.

6 Ibid., p. 593.

7 Ibid.

8 F. Siegler, "Demos on Lying to Oneself," pp. 472-473.

9 Demos, op. cit., p. 591.

10 Ibid.

11 Ibid., p. 593.

12 The inclusion of this condition echoes Demos' analysis, but as mentioned earlier I think it is too strong.

13 F. Siegler, "Self-Deception," p. 32.

14 Cf. J. Canfield and P. McNally, "Paradoxes of Self-Deception," pp. 142-144.

15 Condition (a) has already been argued for, and I consider (b) to be an uncontroversial feature of an intention.

16 This is a slightly altered version of an example Siegler discusses.

17 Siegler, op. cit., p. 40.

18 More will be said about their relationship to self-deception in Chapters II and III.

19 Siegler, op. cit.

20 Another example which illustrates what I have in mind here is the fear some people have of flying in commercial airplanes. Although they may recognize that planes are safer than cars and perhaps even admit that such behaviour is unreasonable, the fear is so strong that it prevents them from travelling by air. They cannot help it despite themselves. We know that the causes of the fear are the stories and pictures of airplane disasters. They cause the fear but they do not at the same time provide a basis for that fear.

21 Siegler, "Self-Deception," p. 34 and "Demos on Lying to Oneself," p. 475.

22 F. Siegler, "An Analysis of Self-Deception," p. 161.

23 Ibid.

24 Ibid., p. 162.

25 Ibid., p. 160.

26 J. Canfield and D. Gustafson, "Self-Deception," pp. 33-35.

27 T. Penelhum, "Pleasure and Falsity," p. 88.

28 Ibid.

29 Ibid.

30 Ibid.

31 Ibid. Fingarette notes that the model as set out is inconsistent with Penelhum's remarks that the self-deceiver neither fulfills the criteria for belief nor disbelief. While this appears to be a result of carelessness on Penelhum's part, there is no possibility of misunderstanding the direction of the argument.

Chapter II

1 H. Fingarette, Self-Deception.

2 Ibid., p. 34.

3 Ibid., p. 35.

4 Ibid.

5 Ibid., pp. 38-39.

6 Ibid., pp. 40-41.

7 Ibid., p. 42.

8 Ibid., p. 47.

9 Ibid., p. 48.

10 Ibid., p. 49.

11 Ibid., p. 62.

12 Ibid., p. 71.

13 Ibid., pp. 86-87.

14 Ibid., p. 87.

15 Ibid.

16 Ibid., p. 41.

17 Ibid., pp. 39-40.

18 Ibid., p. 40.

19 Ibid., p. 45.

20 Ibid., p. 46.

21 Ibid., p. 43.

22 Ibid., p. 47.

23 Ibid., pp. 66-67.

24 Cf. ibid., pp 111-135.

25 Ibid., p. 28.

26 Ibid., p. 29.

27 Ibid., p. 31.

28 Ibid., p. 143.

29 Ibid., p. 48.

30 Indeed, lack of intention is in keeping with his description of the self-deceiver as victim rather than as instigator. See his Chapter VII.

31 Fingarette, op. cit., p. 140.

32 Ibid., p. 141.

33 Ibid.

34 Ibid., p. 144.

35 Ibid., p. 47.

Chapter III

1 My understanding of core and peripheral beliefs owes a great deal to Wittgenstein's On Certainty and the second section of J. Bennett's "Analytic-Synthetic," which is itself an interpretation and criticism of Quine's attack on the analytic-synthetic distinction.

2 J. Bennett, "Analytic-Synthetic," pp. 167-168.

3 But as we shall see below, not every core belief can be rejected without absurdity, allowing for a new belief system to be adopted.

4 L. Wittgenstein, On Certainty, passim.

5 To do justice to Wittgenstein's penetrating discussion, the reader is strongly recommended to refer to the aforementioned book.

6 Evidently these claims need further investigation. Unfortunately, I am not at the moment able to do so. But even if they are wrong, I do not think my model of self-deception thereby suffers.

7 This comparison is not meant to suggest that the theist is a self-deceiver. If the theist's belief is at the core of his belief system, i.e., a groundless assumption, then it is incorrect to speak of evidence for and against it. And since recalcitrant evidence is a crucial feature of self-deception, it follows that the theist is not a self-deceiver. (Cf. note 8 below) Of course, if someone's belief in a Supreme Being were held as an intermediate belief and thus less indispensable to his belief system, it seems that self-deception might be possible in a case like this.

8 In addition, it may be possible for someone to hold simultaneous inconsistent beliefs which are part of inconsistent belief systems. For example, those philosophers G. E. Moore accuses of denying in a philosophical context what they know with certainty in an everyday context suggests that they hold both ordinary and philosophical belief systems. And since both can, I think, be kept separate without impinging on one another, and each amounts to an adoption of either realism or idealism, for which qua philosophical theory there is no evidence for or against, it would be improper to speak of self-deception. Perhaps what Moore is getting at in his 'defense of common sense' is that the common sense belief system avoids the muddles inherent in other belief systems, and by virtue of this it is to be preferred. But if the alternative belief system is perfectly consistent, then the only means I can see of choosing between them is this: Part of what it is to be rational is that when faced with two opposing hypotheses, one more complex than the other, both of which adequately account for the same data, the simpler is taken to be the true or correct one (Occam's Razor — isn't it also a core belief?).

BIBLIOGRAPHY

Bennett, Jonathan. "Analytic-Synthetic." Proceedings of the Aristotelian Society, 1958-59, pp. 163-188.

Canfield, John V. and Gustavson, Don F. "Self-Deception." Analysis, vol. 23, 1962, pp. 32-36.

Canfield, John and McNally, Patrick. "Paradoxes of Self-Deception." Analysis, vol. 21, 1960, pp. 140-144.

Demos, Raphael. "Lying to Oneself." Journal of Philosophy, vol. 57, 1960, pp. 588-595.

De Sousa, Ronald B. "Self-Deception." Inquiry, vol. 13, 1970, pp. 308-320.

Fingarette, Herbert. Self-Deception. London, Routledge & Kegan Paul, 1969.

Paluch, Stanley. "Self-Deception." Inquiry, vol. 10, 1967, pp. 268-278.

Penelhum, Terence. "Pleasure and Falsity." American Philosophical Quarterly, vol. 1, 1964, pp. 81-91.

Pugmire, David. "'Strong' Self-Deception." Inquiry, vol. 12, 1969, pp. 339-361.

Sartre, Jean-Paul. Being and Nothingness. trans. Hazel E. Barnes, New York, Philosophical Library, 1956, pp. 47-70.

Siegler, Frederick A. "An Analysis of Self-Deception." Nous, vol. 2, 1968, pp. 147-164.

Siegler, F. A. "Demos on Lying to Oneself." Journal of Philosophy, vol. 59, 1962, pp. 469-475.

Siegler, Frederick A. "Self-Deception." Australasian Journal of Philosophy, vol. 41, 1963, pp. 29-43.

Wittgenstein, Ludwig. On Certainty. eds. G. E. M. Anscombe and G. H. von Wright, trans. Denis Paul and G. E. M. Anscombe, Oxford, Blackwell, 1969.