

WHAT IS THE ROLE OF THE SELF IN SELF-DECEPTION?*

RICHARD HOLTON

The orthodox answer to my question is this: in a case of self-deception, the self acts to deceive itself. That is, the self is the *author* of its own deception. I want to explore an opposing idea here: that the self is rather the *subject matter* of the deception. That is, I want to explore the idea that self-deception is more concerned with the self's deception *about* the self, than with the self's deception *by* the self. The expression would thus be semantically comparable to expressions like 'self-knowledge' (which involves knowledge *about* the self) rather than to expressions like 'self-control' (which involves control *by* the self).¹ On this approach, what goes wrong, when we are self-deceived, is that we lack self-knowledge; or, more accurately, since one can lack knowledge without falling into error, what goes wrong is that we have *false* beliefs about ourselves. Not any kind of false belief about oneself; I am not self-deceived when I mistake my shoe size. Rather, self-deception requires false beliefs about the kind of subject matter that, were one to get it right, would constitute self-knowledge. It is an interesting fact about current English that, though we talk freely of self-knowledge, we have no common term to designate its absence. Seventeenth century writers talked of self-ignorance; but the term has fallen from use. I suggest that 'self-deception' is the nearest we have.

Lack of self-knowledge is clearly central to many cases that we describe as cases of self-deception or self-delusion (following the OED I take the two as synonymous). For an illustration in a non-philosophical context, here is Cowper:

How many self-deluded nymphs and swaines,
Who dream they have a taste for fields and groves,
Would find them hideous nurs'ries of the spleen,
And crowd the road, impatient for the town²

*An early version of this paper was given at a workshop on self-deception at Stirling. Thanks to the audience there, and to those who heard similar material at Aberdeen, Keele, Birmingham and Edinburgh. Special thanks to André Gallois, Eve Garrard, Rachana Kamtekar, Rae Langton, Al Mele, and Tim Williamson.

¹But note that I am not giving an argument about grammatical form. I am not saying that

(1) x was self-deceived

has the grammar of

(2) x was deceived about x.

rather than of

(3) x was deceived by x,

Evidence against that claim comes from observing the greater acceptability of

(4) William was self-deceived about his mother

than of

(5) William was self-deceived by his analyst.

²William Cowper, *Task* iii 316–19

The contention here is that the nymphs and swains lack self-knowledge; they do not know what their tastes really are. Similarly, much philosophical discussion of self-deception is centrally concerned with lack of self-knowledge. This is the main focus of the discussions given by both Joseph Butler and Adam Smith; and much earlier, by Daniel Dyke.³ A passage from Smith will give a sense:

The opinion we entertain of our own character depends entirely on our judgements concerning our past conduct. It is so disagreeable to think ill of ourselves, that we often purposely turn away our view from those circumstances which might render that judgement unfavourable. He is a bold surgeon, they say, whose hand does not tremble when he performs an operation upon his own person; and he is often equally bold who does not hesitate to pull off the mysterious veil of self-delusion, which covers from his view the deformities of his own conduct.⁴

However, it is one thing to say that cases of self-deception typically *involve* error about the self; it is another to say that such error is somehow *essential* to the concept. It is this latter claim that I want to investigate. So what exactly would the claim be? The weakest position is that the ordinary concept of self-deception is ambiguous: it can mean either deception by the self or about the self. This is the position taken by certain dictionaries.⁵ But I want to go further than that. I want to claim, at the very least, that mistake about the self is *necessary* for self-deception. And this already will be controversial. For it might seem as though a counter-example is ready to hand in that mainstay of the self-deception literature: the trusting cuckold who maintains a belief in his wife's fidelity despite enormous evidence to the contrary. Isn't he mistaken about his wife, but not mistaken about himself? I shall suggest that things are not as simple as they might seem. Even here mistake about the self is required if we are to have a case of self-deception: the victim must be in error as to how justified his belief is. Cleared-eyed wishful thinking, involving no mistake as to the warrant for the belief, and hence no mistake about the self, does not constitute self-deception. In the case of the cuckold the point is hard to test, for it is hard to see how his belief in the fidelity of his wife could coexist with an awareness that his evidence was against it; we are into the difficult terrain of the impossibility of deciding to believe. But other cases will make my point.

Some might object already. I said that I was going to treat self-deception as mistake about the self; and mistake about one's wife hardly counts as that, even if it is underpinned by mistake about oneself. It might seem then that I am now offering a much less appealing disjunctive account: self deception requires either mistake about the self, or else mistake that is underpinned by mistake about the self. I think in fact that this appearance of heterogeneity is

³Butler, *Fifteen Sermons*, Sermon vii, 'Upon the Character of Balaam'; Sermon x 'Upon Self-Deceit'; Smith, *Theory of Moral Sentiments* iii. 4; Dyke, *The Mystery of Self-Deceiving* (Revised Edition, London: Richard Higgenbotham, 1630)

⁴Smith, *ibid.*

⁵The 1993 edition of the SOED for instance, or the 1987 edition of the Random House Dictionary

misleading. The crucial point is that ‘self-deceived’, like ‘deceived’ does not take a sentential complement. One cannot say that someone was self-deceived *that* p. Rather, when one is self-deceived, one is deceived *about* something; and here what replaces the ‘something’ is a noun phrase: one is deceived about one’s motives say, or, one’s desires, or the qualities of one’s child, or the regard that others have for one. Self-deception thus concerns a *subject matter*, and not a single belief. This is not to deny the central role of false belief here: of course, if a person is self-deceived about a subject matter then some of their beliefs about that subject matter are false. Here my proposal is that some of these false beliefs, indeed, some that are sufficiently central, must be beliefs about the self. The cuckold’s beliefs about the status of his own assessment of his wife’s fidelity are part of his set of beliefs about his wife’s fidelity. It is true that once we come to speak of particular beliefs being the result of self-deception—when we say that an individual was self-deceived in believing that *p*— we seem to be forced back into a disjunctive characterization: either that belief, or a belief on which it based, are about the self. But we need not think that the account is *fundamentally* disjunctive.

Such a conception will form the basis of my argument that mistake about the self is *necessary* for self-deception. Should I go further still, and argue that it is necessary *and sufficient* for self-deception—that all there is to self-deception is mistake about the self? The temptation is strong; yet, reluctantly, I resist.

Why is the temptation strong? There is a disreputable reason, of course: the lure of the bold and sweeping claim. But there is a legitimate reason too. If we were to claim that all there is to self-deception is mistake about the self, then we would at a stroke clear up the main perplexities that we find with the concept. For we would no longer be forced to confront the question of how the self could succeed in engineering its own deception. And this is a tricky question which has been at the heart of the philosophical discussion of self-deception. If self-deception is modelled on the deception of others, then it seems impossible that the self could succeed in being both the deceiver and the deceived: for it would either have too little knowledge to play the former role, or too much to succumb to the latter. The only route seems to be to split the self into two; but this brings with it a host of other problems.⁶

Accordingly, many writers have argued that in accounting for self-deception, we should maintain the idea that the self deceives itself, but that we should drop various of the features that arise in the deception of others. In particular, it has been suggested that we should drop the idea that the deception occurs intentionally; this gives rise to an account that sees self-deception as something like *motivated biased judgement*.⁷ Indeed, on a simple head-count of recent philosophical articles this sort of view seems to be well on the way to becoming the new orthodoxy. Yet if self-deception amounted to

⁶For an amusing discussion of the problems, see Mark Johnston ‘Self-Deception and the Nature of Mind’ in B. McLaughlin and A. Rorty eds, *Perspectives on Self-Deception* (Berkeley: University of California Press, 1988) 63–91.

⁷This is the essence of Mele’s account that is discussed in detail below.

nothing more than mistake about the self, we could side-step this whole debate. We would no longer need to give an account of how the self can bring about its own deception, even unintentionally.

So why resist? Simply because one can be in error about oneself without being self-deceived. Thus suppose a psychologist becomes convinced that self-knowledge is hardly ever to be achieved by introspection. We might call him B.F. Desirous of self-knowledge he submits himself to a battery of standard psychological tests. Somewhat surprisingly the results tell him that he is the submissive type; but B.F. is convinced by the method and believes them. It turns out though that he should have taken more notice of his prior beliefs; these were someone else's results, sent to him by mistake. His own results revealed him, correctly, to be domineering. Clearly B.F. has false beliefs about himself, of just the kind that, had they been true, might have counted as self-knowledge. Yet we would not naturally talk of self-deception here.

At least part of what is lacking is that B.F. is in no way culpable for his error; he has not brought it upon himself. And that brings us back to the idea that self-deception requires the self to be a deceiver, an idea which is even more to the fore when we speak of an individual *deceiving himself*.⁸ I think it has to be conceded that the idea of the self as an active manipulator of its own beliefs has been around for a very long time. As we shall see, it is there in the New Testament, and in much of the Christian discussion since. It would be foolish to deny that such ideas are part of the concept of self-deception. Moreover there are other idioms—*kidding yourself*, and *fooling yourself*—which bring much the same idea.⁹

I have two responses. The first is simply to retreat to the claim about a necessary condition. In so far as there is a way of making coherent sense of the idea that one can (perhaps non-intentionally) deceive oneself, it seems to me that it is much more plausible as an account of our ordinary concept of self-

⁸Might it be supposed that these two locutions are not equivalent? Thus suppose that we gained control over the mental life of some unfortunate. And suppose we used our power to give him false beliefs about himself. Mightn't we say of him:

(N) He's self-deceived, though he isn't deceiving himself.

(We might have said something similar for the case of B.F.) The problem with putting much weight on such considerations is that, when faced with a construction such as (N), we do our best to provide it with a sensible interpretation. It could be that our success in giving an interpretation to (N) says more about our ability here than about what (N) antecedently meant. The strongest conclusion that could possibly be drawn is, I think, that 'self-deceived' is ambiguous between a sense in which it means 'mistaken about oneself' and a sense in which it means 'deceived by oneself'; but I am sceptical.

⁹Though they are often used to mean something like *pretending to oneself*. 'For a while there, you try kidding yourself that you're going with an unmarried man' says Fran in Billy Wilder's 1960 film *The Apartment*; 'Then one day, he keeps looking at his watch, and asks you if there's any lipstick showing, then rushes out to catch the 7:14 to White Plains'. Fran never really believed that her lover was unmarried; she just pretended to herself that he was. For a compelling argument that such pretence can be morally important, see Stephen Darwall's illuminating article 'Self-Deception, Autonomy and Moral Constitution' in B. McLaughlin and A. Rorty eds, *Perspectives on Self-Deception* (Berkeley: University of California Press, 1988) 407–30. But unlike Darwall, I do not think that it forms the core of what we think of self-deception. And, despite his persuasive arguments, I am not convinced that this is what Butler meant by it either.

deception if we add a further necessary condition concerning subject matter: it must involve deception *about* the self. The second response is more speculative. It is that once the idea of self-deception as involving lack of self-knowledge has been put back into centre stage, we can put the conception of the self as a genuine deceiver back in its proper place: namely as a bit of theorising, an attempt to explain how such a lack of self-knowledge is possible, and why we should be held responsible for it. I shall discuss this at the end of the paper.

THE NECESSITY OF MISTAKE ABOUT THE SELF

Let us start then with the first point, and with the idea that mistake about the self is necessary for self-deception. I shall begin by pointing out what I think is wrong with Alfred Mele's influential account.¹⁰ Mele is giving an account which tries to make sense of the idea that the self can deceive itself; but in a way that avoids the paradoxes to which such an approach is prone. Crucially this involves rejecting the idea that the a process is intentional. Instead he proposes four conditions as jointly sufficient for, as he puts it, *entering self-deception in acquiring a belief that p*:

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

Note that Mele simply says that these conditions are *sufficient* for self-deception; they are not claimed to be necessary. We can return to this issue of necessity shortly. For now I just want to argue that the conditions are not sufficient. Consider this:

Jean-Marie is a racist. He thinks that blacks and Arabs are not as good as whites: not as clever, or as imaginative, or as brave, or as trustworthy, or whatever. Take just about any property that Jean-Marie might regard as a virtue, and he will think that whites have more of it. Let us assume that his beliefs here are, by and large, false. But he holds them sincerely. And were we to challenge them, he would provide evidence: reams of it, taken from the magazines and newspapers of the kinds of organization to which he belongs. He is aware of the opposing view; indeed he has reams of that too, collected to document the conspiracy which he thinks pervades the liberal establishment that controls the mainstream press and publishing houses.

Jean-Marie meets all four of Mele's conditions. He is bigoted and prejudiced. Yet he is not obviously self-deceived. The reason, I suggest, is that his false beliefs are about the wrong sort of thing. He is not making a mistake about himself.

¹⁰Mele has presented these in a number of places. See, for instance, 'Real Self-Deception', *Behavioral and Brain Sciences* 20 (1997) 91–102.

Or at least, he is not *obviously* making a mistake about himself. But suppose we focus on certain aspects of Jean-Marie. Suppose we ask him about his own epistemic practices: why he believes this bit of evidence rather than that, why he thinks he should listen to this person and not that one. I think that it is very likely that we will very soon stumble upon some areas where he is making a mistake about himself: he thinks he is giving both sides a fair hearing, weighing the evidence equally, when he is not. He will be mistaken about whether he is living up to his own belief forming standards. Here then will be his mistakes about himself. But it is exactly as a result of considering these factors that we might hesitate over the initial judgement that he is not self-deceived. For once we think that he is bound to be making a mistake about himself in coming up with the racist judgements he comes up with, it is plausible to think that he *is* self-deceived in making those very judgements.

So we now in a rather difficult position. I suggested that Mele's four conditions were not sufficient for self-deception; I had an example that seemed to bear this out; but on closer inspection it appears that it will not work as a counter-example precisely because in meeting Mele's four conditions it meets the very condition that I was claiming he had left out. And it might seem that this shows we will never get a test case that meets Mele's conditions without meeting this further one. For it might seem that whenever we form biased opinions, we will be bound to be making a mistake about ourselves: once we recognize that we have formed our beliefs by a misapplication of our own belief forming methods, we will no longer be able to maintain those beliefs.

Now perhaps that last point is right: we cannot maintain beliefs together with the realization that they stem from a misapplication of our own methods. But that does not mean that we can never form biased beliefs without making mistakes about ourselves. Consider this:

Catherine has applied for several jobs recently, and has been unsuccessful each time. She has also been horribly disappointed each time. She puts her disappointment down to too much thought. On each occasion she had spent a great deal of time thinking about the job, had, as a result, imagined just what it would be like to get it, and so had been devastated when she didn't. She has just decided to apply for another job. She thinks that it is clearly better than her current job; otherwise she wouldn't be putting in for it. But she has resolved not to think too deeply about what it is like, or to examine the evidence that she has; at least not until or unless she gets an offer. She knows that, were she to think more about the job, there is some chance that her opinion of its merits would change; but she thinks that the possibility of error here is worth risking to maintain her equanimity. As it happens she is radically wrong about the job. It is a terrible job, far worse than her current one, as a little more reflection would have shown her.

What should we say about Catherine? She has formed a false belief; and she has formed it in a non-deviant way as a result of treating evidence, which would have supported the opposite conclusion, in a way which is motivationally biased: she ignores it, because she does not want to get her hopes up. She meets all four of Mele's conditions. I think that we might say, exploiting a legal phrase, that she is wilfully blind. But is she self-deceived?

Surely not. She is simply, and quite rationally, refusing to look at certain evidence that she has.

How could we turn Catherine's case into a case of self-deception? Suppose that she had been less aware of her own strategy. Suppose that she had in fact been avoiding finding out about the job, but that she would have sincerely and vehemently denied it if someone had suggested that that was the case. Then we plausibly have a case of self-deception. Alternatively suppose the story is as we initially described it, but that when Catherine is offered the job, flushed with her success, she forms the belief that she knows all that there is to know about it, and promptly accepts. Again it is plausible to think that we now have a case of self-deception. In each case what makes the difference between wilful blindness and self-deception is the absence or presence of Catherine's false beliefs about herself.

In these cases the difference between wilful blindness and self-deception will result in a difference in behaviour. If Catherine is merely wilfully blind then we would expect her, on being offered the job, to make some more enquiries about it; if she is self-deceived we would not. But not every case will bring such differences in behaviour. Juliet's past disappointments have concerned holidays rather than jobs. Convinced that her past disappointments had been the result of too much time spent poring over the brochures, she now skims them quickly, makes her choice, and puts them out of her mind till the holiday starts. Her behaviour might differ not at all from that of her self-deceived alter-ego who has no conception that she is avoiding the evidence.

We have the kind of test case that we need. Mele's four conditions are not sufficient for self-deception. We need to add a further condition that brings the requirement of error about the self. Getting a little clearer on the exact nature of that condition is the subject of the next section.

REFINING THE NECESSARY CONDITION

It is easy enough to come up with necessary conditions for the application of concepts. Being a thinker is a necessary condition on being self-deceived, but that is scarcely informative. Necessary conditions become more interesting as they become more restrictive. And one way to explore whether we can make a necessary condition more restrictive is to ask a question about sufficiency. That is: suppose a condition N is a necessary condition for the application of a certain concept. Then we can ask whether N, together with some other conditions $C_1 \dots C_n$, is *sufficient* for the application of that concept. If it is not, we can ask whether we can restrict N in such a way that the resulting condition N* is, together with $C_1 \dots C_n$, sufficient for the application of that concept. I take it that that is what we do, at least implicitly, when we try to refine a necessary condition.

To take the case at hand: suppose we add to Mele's putative sufficient conditions a further condition which requires that the subject be making a relevant mistake about the self. We can then ask whether the resulting set of conditions will provide sufficient conditions for self-deception. If not, we can

ask whether there is a way of restricting the further condition so that they will. That is our way of refining the condition.¹¹

In fact we cannot do things quite so simply. Mele formulated his sufficient conditions in terms of an individual being self-deceived in forming a particular belief. I suggested, in contrast, that we should treat self-deception as occurring with respect to a subject matter. It turns out to be no easy matter to make the relevant translation. So let me instead suggest a new set of sufficient conditions, which draw on Mele's:

S is self-deceived about a subject matter α if

1. S's body of beliefs about α contains mistaken beliefs about the self.
2. S treats data relevant, or at least seemingly relevant, to the truth value of these mistaken beliefs in a motivationally biased way.
3. This biased treatment is a nondeviant cause of S's acquiring these mistaken beliefs.
4. The body of data available to S at the time provides greater warrant for rejecting these beliefs than for accepting them

Our aim then is to refine the first of these conditions, as a necessary condition, by seeing whether the four of them are jointly sufficient.¹²

Before we start, let us remind ourselves of what is meant by mistake about the self. It is mistake in the kind of belief that, were one to get it right in the right sort of way, would count as self-knowledge. To analyze this in turn we would need to say quite what self-knowledge amounts to; and that is no easy matter. We might try saying that it is knowledge about one's psyche: I remarked earlier that a mistake about one's shoe size would not normally count as a lack of it. But sometimes mistakes such as this might count: imagine that the person concerned is an anorexic, habitually over-estimating

¹¹Avoid an easy trap: having a necessary condition and a set of sufficient conditions does not give us necessary and sufficient conditions as these are ordinarily conceived. Being over 18 is a necessary condition on being eligible to vote in a European election; being over 18 and being a UK citizen together give us sufficient conditions. But they do not give us necessary and sufficient conditions as these are ordinarily understood; French citizens over 18, for instance, are also eligible to vote. When we talk of necessary and sufficient conditions we normally mean conditions that are individually necessary and jointly sufficient.

¹²I am not going to investigate whether conditions 2–4 are necessary. I doubt that they are. Indeed, if self-deception is a cluster concept, there will be no set of conditions that are individually necessary and jointly sufficient; indeed, even if it's fairly unified there is no reason to insist that there must be such a set. On the general issue of necessary conditions without sufficient, see Timothy Williamson *Knowledge and its Limits* (Oxford: Clarendon Press, 2000) pp. 31–3. On one reason for scepticism here: I doubt that the bias needs to be motivated, in the sense that it needs to be inspired by a desire on the part of the agent, either a desire to believe that proposition, or a desire that the proposition be true. Someone can be self-deceived in under-estimating their own competence without wanting either to believe that they are incompetent or wanting to be so. For some discussion of this issue see Martha Knight 'Cognitive and Motivational Bases of Self-deception' *Philosophical Psychology* 1 (1988) 179–88; and Alfred Mele 'Twisted Self-Deception', *Philosophical Psychology* 12 (1999) 117–37.

how big they are. So perhaps we should say more vaguely that it is knowledge of one's self which is somehow intimately connected with one's psyche. I do not know whether something along these lines can be made to work. I shall not try to define self-knowledge here. Rather, I mean the idea of mistake about the self to be dependent on our prior understanding of it. I hope that that is good enough. One thing though that can be said is that it is *de se* knowledge. I would not have self-knowledge in virtue of knowing all about the psychic workings of some individual who, unknown to me, happened to be me.¹³

With the first condition understood in this way, let us now ask whether the four conditions are sufficient for self-deception. We can start by returning to the case of Jean-Marie. In presenting the case I was assuming that Jean-Marie was not living up to his own standards when he formed his racist beliefs; that was where the mistake about himself, and hence the self-deception, came in. But we might imagine a racist who does fully live up to his own bizarre epistemic standards. Suppose he sincerely thinks that he should only believe what his friends tell him (and they happen to be racists). Moreover, suppose that he has no further reason for thinking this; it is his epistemic bedrock.

This person need not be self-deceived. Will he be making a mistake about himself? He will clearly not be making the same mistake as Jean-Marie; he will not think that he is living up to standards from which he is falling short. But in another respect he will be making a mistake about himself: he will believe that his own standards will lead him to true beliefs. So it looks as though he will meet the four conditions. Worse, this is just one instance of a very general problem. For in a similar respect it seems that just about every false belief will give rise to a false belief about oneself: it will give rise to the belief that what one believes is true.

Now in many cases we might argue that these beliefs will not be about the self in the sense that I am after; that is, they will not be beliefs such that, were the agent to get them right, they would count as self-knowledge. And perhaps we could say the same about the racist with the bizarre epistemic standards: perhaps we could say—though this is far from obvious—that knowledge about whether one's epistemic strategies will lead to the truth is not self-knowledge. But whether or not that is so, there will be cases that cannot be so easily handled. Thus suppose that, as a result of motivated bias in favour of the theories of graduates of my own university, I were to come to think that selves were nothing more than bundles of ideas and impressions. And suppose that I were to go on to apply this belief to myself: I convince myself that I am nothing more than a bundle of ideas and impressions. Let us assume that this is simply false: there is more to me than that, as I would have realized were it not for my bias. Then it seems plausible to say that I am lacking in self-knowledge. So I meet the first condition; and I meet the other three as well. Yet I am not self-deceived.¹⁴

¹³D. Lewis, 'Attitudes *De Dicto* and *De Se*' in *Philosophical Papers* Vol. I (New York: Oxford University Press, 1983). But note that since I want to contrast *de se* knowledge with *dicto* knowledge, I want to resist Lewis's device of converting the latter to the former.

¹⁴Thanks to Rae Langton for the example.

What is special about cases like these? The problem seems to be that the mistake arises, in the first instance, in a belief which is not about oneself. I form a belief about selves in general and then apply it to myself. Similarly the racist with his bizarre epistemic standards formed a general belief about which standards led to the truth, and then applied them to his own practice. And the same is true of the general move from having a false belief, to having a false belief about oneself, namely the belief that one has a true belief. If error about oneself is to give rise to self-deception, the error must begin at home.

We need some terminology. Distinguish errors about the self into two classes: those that result from the application to oneself of an erroneous belief that is not about oneself; and those that do not. Call the latter *fundamentally de se errors*. And when people make mistakes about themselves in ways that, had they got it right, would have been self-knowledge, call these *self-knowledge errors*. Then, putting together the considerations adduced over the last few paragraphs, we can reformulate the first, necessary, condition as follows:

1*. S's body of beliefs about α contains fundamentally *de se* self-knowledge errors.

That completes my refinement of the condition.

THE CENTRALITY OF MISTAKE ABOUT THE SELF

The sufficient conditions on self-deception that I have been working with are something of a medley. The first looks in the direction of the self as the subject matter of the deception; the rest look in the direction of the self as (unintentional) deceiver. I fear that that is as good as we will get. Our ordinary concept of self-deception is a medley. Why is it so? I am going to try to give an explanation; and in the process I shall try to make good my claim that the primary idea in self-deception is that of deception about the self. This will involve me in a fair bit of sweeping reconstructive intellectual history; a dubious activity at the best of times, and one for which I am badly under-qualified. But, with that warning in place, let me press on.

My thesis is this: given certain views about the self, the very idea of making a mistake about oneself will be problematic. It will be something that will demand an explanation. And the only explanation that is available is that the mistake must be the result of some intentional or at least culpable fault on the part of the subject: the self deceives itself. There should be an echo here of Descartes' views on mistake in general. And that is right; for Descartes is very much a part of the tradition of Christian thought that I want to discuss.

Let us start with the New Testament. In the Letter to the Galatians, Paul writes:

If a man thinketh himself to be something, when he is nothing, then he deceiveth himself (Gal. 6: 3–4)

Why does Paul say this? The claim might be analytic. He might hold that it just follows, as a matter of meaning, that when a man wrongly thinks himself to be something he deceives himself. That would be support for the claim that self-deception is simply mistake about the self. But I doubt that it is what is happening here. I suspect that what is going on is that Paul is offering an explanation of how such a mistake is possible. In this he starts a tradition of explanation that continues through the Augustine of the *Confessions*, and Pascal, and down to various Seventeenth and Eighteenth Century British thinkers such as Richard Baxter (*The Mischief of Self-Ignorance, and the Benefits of Self-Acquaintance* 1662) and John Mason (*Self-Knowledge: A treatise showing the nature and benefit of that important science, and the way to obtain it* 1745).¹⁵ Joseph Butler stood clearly in the tradition (two of his highly influential *Fifteen Sermons* (1726) are devoted to self-deception) as did some more secular thinkers like Adam Smith. Self-deception and self-ignorance remain absolutely central Christian concerns. A web search on the former term brings a huge tally of vaguely Freudian popular psychology sites; but they are outnumbered by the forces of electronic Christianity.

Clearly saying anything really worthwhile about this tradition would require more space and expertise than I have. I want just to make two observations. The first is that much of this literature takes it that the mind is available to us; insofar as it is hidden this is the result of our own doing. As Butler puts it:

If it were not for that partial and fond regard to ourselves, it would certainly be no great difficulty to know our own character, what passes within, the bent and bias of our mind.¹⁶

The explanation for this, I suggest, has largely to do with responsibility and blame; and this is my second point. It is a central concern of all of this literature that our lack of self-knowledge makes us sinful, for two reasons. Partly it is because as a result of not knowing our own motives we do not admit what we have done and beg God's forgiveness for it; and partly it is because, not realizing what we are doing, we go on doing it. But we are responsible for our sin. So if we sin as a result of our lack of self-knowledge, we must be responsible for that lack. It is this consideration, I suggest that leads the writers in this tradition to argue that mistake about the self must result from something over which we have control: from something like wilful

¹⁵For a discussion of self-deception in Augustine and Pascal, see van Fraassen, *op. cit.*; and on Augustine, P. Courcelle, *Connais-toi toi-même* (Paris: Etudes Augustiniennes, 1974) Ch. viii.

¹⁶Sermon x, 51-2. There are enormous differences within the tradition as to the explanation of why this information is available to us. Butler's position suggests a natural transparency of mind. In contrast, Augustine and Pascal stress the idea that God forces the information upon us:

[Y]ou thrust me before my own eyes so that I should discover my iniquity and hate it. I had known it, but deceived myself, refused to admit it, and pushed it out of my mind. Augustine, *Confessions* viii. vi (16).

deception of the self. ‘Know thyself’ is an injunction which we wilfully disobey.¹⁷

Quite how they explain this possibility varies from case to case. Sometimes it is explained by proposing a division within the person. Here is Mason:

Another considerable branch of self-acquaintance is, the knowledge of the true motives and secret springs of our actions ... It is not only very possible, but very common, for men to be ignorant of the chief inducements of their behaviour; and to imagine they act from one motive whilst they are apparently governed by another. If we examine our views, and look into our hearts narrowly, we shall find that they more frequently deceive us in this respect than we are aware of, by persuading us that we are governed by much better motives than we are... By thus disguising our own motives we may impose upon men, but at the same time impose upon ourselves; and whilst we are deceiving *others* our own hearts deceive *us*. And of all the impostures, *self-deception* is the most dangerous, because least suspected.¹⁸

At other times it is explained by positing more modest methods like selective attention and a biased starting point. Here again is Butler:

There is plainly, in the generality of mankind, an absence of doubt or distrust, in a very great measure, as to their moral character and behaviour; and likewise a disposition to take for granted, that all is right and well with them in these respects. The former is owing to their not reflecting, not exercising their judgment upon themselves; the latter to self-love.¹⁹

The details of the accounts need not concern us here. My point is simply that self-deception is brought in as an explanation of mistake about the self. That is why the two ideas get bound up in our ordinary notion of the former. I leave as an unanswered question whether those of us who are not moved by Christian concerns, and who do not think of the mind as fundamentally transparent, should be very interested in the concept that results. Perhaps we should be more interested in the simple idea of self-mistake.

¹⁷The Delphic precept has, of course, a long history before Christianity; although often it is interpreted as an injunction to know the nature of man, rather than to gain *de se* knowledge. Indeed, this former interpretation is often the one given to it in Christian thought. See Courcelle, *op. cit.*, and E. Gilson ‘Self-Knowledge and Christian Socratism’, Ch. ix of *The Spirit of Mediaeval Philosophy* (London: Sheed & Ward, 1936). Note too that there are other Christian traditions which put far less store on the possibility and value of self-knowledge. A more Calvinist approach has room to conclude that we can never know our own minds directly. Thus Daniel Dyke, *op cit.*, ‘I say that God only knoweth the heart *exactly* and *certainly*: Because Man and Angels may know it conjecturally, and by way of guessing’ (p. 399). The mind contains ‘a great mangle-mangle and confusion of thoughts, even as there is of dross and good metal in silver and gold, which is so confused together that the dross is not discernible’ (p. 402) Characteristically, what Dyke means by self-deception seems to be simply mistake about the self. For example: ‘We think through pride and ignorance that we are not so bad as in truth we are ... We think ourselves in good & happy estate before God, being indeed miserable’ (p. 49).

¹⁸Mason, Pt. 1 Ch xi

¹⁹Sermon x