Routledge
Taylor & Francis Group

Depth psychology and self-deception

ROBERT LOCKIE

ABSTRACT *This paper argues that self-deception cannot be explained without employing a depth-psychological ("psychodynamic") notion of the unconscious, and therefore that mainstream academic psychology must make space for such approaches. The paper begins by explicating the notion of a dynamic unconscious. Then a brief account is given of the "paradoxes" of self-deception. It is shown that a depth-psychological self of parts and subceptive agency removes any such paradoxes. Next, several competing accounts of self-deception are considered: an attentional account, a constructivist account, and a neo-Sartrean account. Such accounts are shown to face a general dilemma: either they are able only to explain unmotivated errors of self-perception—in which case they are inadequate for their intended purpose—or they are able to explain motivated self-deception, but do so only by being instantiation mechanisms for depth-psychological processes. The major challenge to this argument comes from the claim that self-deception has a "logic" different to other-deception—the position of Alfred Mele. In an extended discussion it is shown that any such account is explanatorily adequate only for some cases of self-deception—not by any means all. Concluding remarks leave open to further empirical work the scope and importance of depth-psychological approaches.*

> There is an innocence in lying which is the sign of good faith in a cause.
> (Nietzsche, 1886/1978, §180)

1. The notion of "depth psychology" under consideration

This paper defends the view that mainstream academic and clinical psychology should make space for psychoanalytic approaches, by arguing that such approaches are needed to explain the phenomenon of self-deception. More will be said about self-deception shortly. In this section, what is meant by a "psychoanalytic approach" is qualified and explained.

Whether in attack or defense, one should distinguish what it is to be a psychoanalytic approach as such, from the specifics *of* any one such psychoanalytic approach—for there are a great number of psychoanalytic theories, differing widely between themselves. A consideration of these differences shows that an emphasis on sexuality and a psychosexual notion of the unconscious is not essential to psycho-

Robert Lockie, Psychology Department, University of Luton, Luton, Bedfordshire LU1 3JU, UK, email: bob.lockie@luton.ac.uk

analysis; nor is the id/ego/superego; nor is the Oedipus Complex; and so on down a long list. What, then, is essential to psychoanalysis as such?

Essential to psychoanalysis is the notion of a *dynamic unconscious*; any position which employs such a notion is known as a *depth psychology*. The notion of a dynamic unconscious is very much more than that notion of the unconscious which just means "not conscious"—a notion employed in almost every area of modern psychology and cognitive science after Helmholtz and his concept of "unconscious inference." Distinguishing a specifically *dynamic* unconscious is not easy, and I do not propose to give an analytic definition of it; but in explication, it can be seen to feature something like the following:

- The person is made up of *parts* of some kind.
- The parts have their own motivational interests (desires, bio-psychological drives, socio-psychological norms and ideals).
- The activities and motivations of these parts are not necessarily known to, or shared by, the other parts, or the "person as a whole."
- The sense in which these parts may have or lack knowledge of the other parts, or the person as a whole is an *active* sense. So, these parts have (partially successful) means of selectively concealing, revealing, and deceiving the other parts, and in other ways competing for psychological resources, in accordance with their (the parts') sub-interests. This internal psychical "ecology" is an *active system* of interacting parts (involving processes of competition, cooperation, symbiosis, etc.); it is in this sense that it is "dynamic."
- The activities of these parts may often be revealed only obliquely, by skilled inference (psychological "detective work") to uncover the meaning revealed in slips, dreams, bungled actions, selective amnesia, neurotic symptoms, character traits, etc. (typically by using clinical interview techniques—free association, hypnosis, and other techniques allied to the clinical experience).

Many attacks on this notion are, even now, a hangover from the positivist/behaviorist era, in that they would attack the Helmholtzian notion of the unconscious no less than the Freudian. This battle is long since won, and I am not concerned to re-fight it here. A case in point is the homuncular objection to psychoanalysis—an objection long since overturned on behalf of modern (e.g. cognitive) psychology by such figures as Fodor (1968) and Dennett (1978); yet routinely flourished at undergraduates by mainstream (e.g. cognitive) psychologists to rubbish the notion of there being the "parts" of which depth psychology treats [1].

A further distinction is needed to qualify the conception of depth psychology being defended here. For Freud, psychoanalysis was three things: first, a type of treatment; second, a method of investigation; third, a body of psychological theory (albeit ostensibly derived from the method of investigation and underpinning the treatment). This paper is concerned only with psychoanalysis as the third of these. In particular, I am not here defending the psychoanalytic emphasis on clinical methods, whether of discovery or (especially) confirmation. Though there might be much, at least of the former, worth defending, *I am only defending the existence of a*

dynamic unconscious as such. At a stroke, the numerous methodological criticisms of psychoanalysis then become entirely irrelevant to this argument.

Consistently among psychologists, methodological (e.g. Popperian) criticisms are taken to have application against *any* depth-psychological theory—are taken to have established that the "very idea" of a depth psychology is and must be untestable [2]. However, unless these arguments are augmented by some of the long-over-turned positivist or behaviorist presuppositions already dismissed for ruling out even a Helmholtzian unconscious, they will apply only to psychoanalysis as *method*—sense two, above. The "very idea" of a dynamic unconscious is not intrinsically, method-ologically, anything. In physics, one may advance a theory of action-at-a-distance that is intrinsically testable (Newton) or not, and likewise for psychology and the very idea of a dynamic unconscious [3].

In the light, then, of these caveats, here are three questions concerning this notion of the dynamic unconscious:

(1) Do you use some version of this notion to help make sense of yourself and others?

I think your answer will be "yes." I think there will probably be no reader of this journal who does not. I doubt there will be any academic or clinical psychologist for whom, outside of their published professional work, the answer would be "no."

(2) Does academic psychology as presently conceived make space for the notion of a dynamic unconscious?

The answer is "no." Mainstream academic and clinical psychology is not engaged in any project to uncover the unconscious meaning behind slips, dreams, selective amnesia, neurotic symptoms or behavior and mental life generally. So, such psychologists study dreams, e.g. physiologically (Crick & Mitchison, 1983, 1986), and perhaps—at best—in terms of the Helmholtzian (information-processing) unconscious (Evans, 1984). They study slips and bungled actions likewise (Norman, 1981; Reason, 1992), and equally for all the clinical and everyday phenomena of which psychoanalysis treats. They situate no *meaning* at all in such phenomena. Within mainstream psychology, depth psychology is alternately scorned and ignored.

(3) If (1) and (2) don't match up, who is right and who is wrong?

I think that mainstream psychology is wrong, and I think a consideration of one of the most everyday of phenomena makes clear that it is wrong.

Before we get to the commonplace, note that there are a number of more esoteric phenomena which appear exceedingly difficult to account for without the notion of a dynamic unconscious. Among numbers of others these include the "hidden observer" phenomenon in hypnosis (Hilgard, 1977), dissociative identity disorder ("multiple personality"), secondary gain from illness, perceptual defense and related experimental phenomena (Dixon, 1981; Kline, 1984), the flashbacks and dreams of post-traumatic stress disorder sufferers, and, in general, the mysteri-ous case of hysteria—now so rare, in Freud's day so common (Shorter, 1997). All of these are plausibly *dissociative* [4] phenomena: a splitting of one part of the self

from another. Psychoanalysis is uniquely well suited to give accounts of such phenomena; but there is a far more ordinary phenomenon than these that compels us to acknowledge the existence of a dynamic unconscious. This is *self-deception*.

2. Accounts of self-deception

How can the self deceive the self? Well, what is *deception* as such—say, of one agent, Smith, by another, Jones? Jones knows that *p*. Jones has an interest in seeing that Smith either comes or continues not to believe that *p*; or even to believe that *not-p*. Jones has the power to bring about that Smith comes or continues to not believe that *p* or believe that *not-p*. Jones successfully acts in the exercise of this power, for the purpose of achieving his interests.

Make (Smith and Jones) both Smith, or both Jones, and paradox beckons. How can Smith believe that *p*, yet not believe that *p* or believe that *not-p* (the "static paradox"; Mele, 1987)? How can Smith, believing that *p*, *act* in pursuit of his interests, to *undermine* his, Smith's, believing that *p* (the "dynamic paradox"; Mele, 1987)?

For us, after Freud, such paradoxes are not a problem. One part of Jones' mind knows that *p* (that Jones sexually desires men). Perhaps (most obviously) because that part actually does desire men—it's the *desiring part* of Jones' mind. A second part of the mind has an interest in seeing that Jones does not believe that he sexually desires men. Perhaps (most obviously) because a third part of Jones' mind—the part that contains Jones' social influences—will raise merry hell if it's forced to come face to face with this desire. So the second part of the mind uses a variety of devices on the first and third parts of the mind, to prevent the one part of the mind from fully realizing the desires of the other part of the mind. It *defends* the conscience from confronting these desires. And the devices used to do this are called *defense mechanisms*. This is one quasi-Freudian explanation of one imagined case of self-deception; there are many other depth-psychological explanations possible, and many other cases to explain.

How, though, might one account for the phenomenon of self-deception if one eschews the notion of a dynamic unconscious? This question permits of a number of responses which are by no means mutually exclusive. At one level, responses like (1)–(3):

(1) Self-deception is a phenomenon of *attention*: non-attending to some evidence, excessive redirection of attention to other evidence.
(2) Self-deception is a *constructive* phenomenon: either a phenomenon of social construction or individual construction (representation).
(3) Self-deception is a case of *bad faith* or a "disavowal of engagement."

Taken somewhat differently, this question elicits responses like (4)–(6):

(4) Real self-deception doesn't exist. A case like that described above is then an example of a person deceiving others, not himself; and likewise, in different ways, for our other cases of self-deception.

(5) Self-deception exists but is not to be understood on the model of other-decep-tion. If understood in a different way, there is no reason to postulate a depth psychology.

(6) Although real self-deception does exist, and to that extent, the dynamic unconscious exists, it's much less prevalent than the psychoanalysts would have us believe. This phenomenon, hence the dynamic unconscious, doesn't suffuse our entire mental lives.

Positions (1)–(3) represent rather more specific, contentful, accounts, whereas (4)–(6) represent more structural, or "logical," responses. Response (5) is the major philosophical challenge, but material from considering the specifics of positions (1)–(3) is necessary to appraise the viability of this stance. There are specific accounts other than (1)–(3) available, but there are reasons, to be given below, why they will probably not represent challenges distinct from these.

3. Attentional accounts of self-deception

The first position to consider is that self-deception is a phenomenon of attention: one attends to data conducive to the false conclusion, *not-p*, and does not attend to data conducive to *p*. In this way one deceives oneself into believing *not-p*. How, though, might the details go here?

We have a good idea of what a theory of attention looks like as such: modern cognitive psychology began with the development of such theories (Broadbent, 1958; Cherry, 1953). The details of such accounts vary rather significantly—early versus late bottleneck theories, capacity theories, etc.—but the central ideas are well established, and not, at the level of abstraction we are concerned with, importantly distinct. To understand what all such theories have in common, two issues should be kept in mind: *why* there is attention and *how* it is directed. The reason for the existence of attention—the "why"—that is embraced by all existing psychological theories is that we have limited cognitive resources. We can devote such resources to processing exhaustively (i.e. semantically) only some information—say, as on certain theories, just the one channel of communication (one conversation, per-haps). The question which then arises is which channel of communication, of all those we are presented with, we should focus on. The answer is obvious: the most *important* one. The problem then is equally obvious: unless we exhaustively process all channels, we won't know what this *is*.

Plainly then, this "why" answer—cognitive resource limitation—severely con-strains which explanations are permissible for *how* we direct attention. For this reason, in the early theories of attention, the component of the mind with the executive power to direct attention was absolutely precluded (on pain of the behaviorists' regress criticisms of homuncularism) from having the power to (seman-tically) *attend to* the unprocessed channels. Only very limited, non-semantic process-ing could occur prior to the bottleneck. This processing obeyed simple protocols: redirect attention if you detect a loud bang, or someone screams, etc. At best, some dedicated processing resources were posited to detect the subject's own name and

automatically redirect attention—the famous "cocktail party effect" (Cherry, 1953). In later theories limitations on pre-attentional processing were liberalized very slightly (e.g. Kahneman, 1973), but because of the constraint of the "why" question, nothing even vaguely resembling intelligent agency could answer this "how" question: executive control over attention simply could not involve anything seriously like *understanding*, *judging*, or *interpreting* the unattended data.

With this in mind, try to apply an attentional account in a case of self-deception. Joe's attention is serially focused on instances of his heterosexuality: he retrieves memories of his marriage ceremony, of others in his military unit thinking of him as macho, of examples of him exchanging heterosexual pornographic magazines with them, and the like. Whenever memories of sexual acts with same-sex strangers in public lavatories occur, attention is rapidly redirected away, to examples of his heterosexuality. When memories/fantasies of sexual arousal by same-sex persons becomes too pressing, masturbation, or other fantasy arousal, or even full sex, can occur with attention devoted to it and utterly *dissociated* from attention that will later be devoted to his "straight life." Two strands to Joe's existence emerge— Straight Joe and Gay Joe. These are, as near as is possible, utterly separated: he *attends* to one, and actively *ignores* the other. The other perhaps emerges in erotic dreams, or a sordid, encapsulated series of sexual releases—whether in fantasy or reality. Just enough attentional resources are devoted to these latter as are needed to effect them, remove the pressing need, then return to the other, dominant, identity.

Ask, then, of such an account, the two questions applicable to any theory of attention: what *directs* this attention (the "how") and what *motivates* it (the "why")? What directs it is some part of the mind with the ability to first *detect*, and then *interpret*, data as being evidence for homosexuality and heterosexuality alike: that is, to *see both sides* (in order to focus attention on evidence for only one side). And what motivates it is the desire, for whatever reason in turn, to have an identity (a self-identity) as straight. But the former does not look at all like the low level processing that occurs in any extant, cognitive, theory of attention—say, prior to the ordinary attentional bottleneck. And the motivation for this filter is not at all the motivation of cognitive economy, but a reflexive, semantic motivation (for whichever reason): the motivation to understand oneself only in a certain way.

Now such an account would not, despite appearances, thereby have to be vulnerable to neo-behavioristic regress arguments—in a way that the Helmholtzian accounts are not—simply because the motivation for this account is not at all the motivation of cognitive resource economy. This account would have it that attention is paid to two sources of evidence (gay and straight) by one part of the mind, which part then filters and screens evidence to the other part(s)—which, as explanation is reasonable enough, provided the evidence turns out to be there for it. The problem is that this "attentional" account appears now either just to be one (plausible) implementation hypothesis for the notion of a depth psychology, or little more than metaphor (we talk of *attention* here, just as above, I talked of "*see*ing both sides"; the latter is only analogically *vision*, the former only analogically *attention*). As a genuine implementation hypothesis for a depth psychology it will deserve scientific evalu-

ation with all the rest. As a theory of the phenomenon of self-deception which nevertheless is not such a depth psychology, it looks thus far, to be a nonstarter.

4. Constructivist accounts of self-deception

Consider then, the second option identified above—that self-deception may be accounted for with a psychology of social or individual constructivism. Constructivism derives ultimately from Kant, and the idea that things in themselves (*noumena*) are unknowable, what is given to us are appearances (*phenomena*). These appearances are not passively registered, bottom–up; they must be actively organized, categorized, and interpreted top–down before they can yield knowledge. Knowledge is *constructed* out of the appearances by the active, organizing principles of the mind. For Kant, the organizing principles were pre-set categories; it is with the development of German idealism in the 19th century that we see the idea that they are mutable, and can be socially determined.

These constructivist views were "psychologized"—made into empirical science—by the early 20th century: in different ways by the Gestalt psychologists, Piaget, and especially Bartlett (a social as well as individual constructivist). Their influence (especially Bartlett's, sans the social) is strongly recapitulated in the work of the "first wave" of American cognitive psychologists: Neisser, Bransford, and so forth. Whether in perception, memory, problem-solving, or any of the other traditional topics of cognitive psychology, we don't build up our picture of the world associatively, as the behaviorists had insisted; rather, we actively *construct* (e.g. *inferentially*) our perceptions, memories, reasoning strategies and the like, through the use of *schemata*—active, top–down, cognitive organizing principles.

An individual version of constructivism was endorsed by Sackeim (1983), who saw it as offering the resources to provide an alternative account of self-deception. Conventional depth-psychological views of self-deception are *dissociative* in this respect: they deal with a self of "parts" (in the ways detailed above), there being two or more "streams" of processing, as Sackeim put it. So, one "stream" desires sex with other men, another desires a heterosexual self-image, a third mediates these two, and so on.

If, however, all cognition is *constructive* (Sackeim cites Neisser, 1967), why do we need two or more "streams" of processing? Sackeim suggests a *single stream* account. The same data may be construed one way or another. An example of this, from a current constructivist theory, is the cognitive theory of depression (Beck, 1976, 1991) which makes sense of two persons, one depressed, one not, each presented with the same data. These data—say, the same social occasion attended by each—may be organized according to "happy" schemata (rose-tinted spectacles) or "depressive" schemata (gray-tinted spectacles). The "happy" schemata will involve the incoming stimuli being assimilated to a certain set of existing memories, thought processes, inferences and the like, while the "depressive" schemata will organize this material very differently. Likewise, a gay man may have sex with a man who is in radical denial of his own sexuality: their *construction* of the same situation

will be radically different; it is in this difference in construction that the possibility of self-deception will consist.

This constructivism as such may become a *social* constructivism (Gergen, 1985). As an example, consider the evolving recognition ("construction") of what was called, at different times, shell-shock, neurasthenia, battle fatigue, PTSD, etc.—specifically, the ability of soldiers to save their lives by being withdrawn from the line on having their behavior thus categorized. The soldier is saved from being regarded as a coward, whether by others or himself, by the presence of a new label for his behavior and underlying state of mind; but it is in the *construction* of that behavior and state of mind that the possibility of self-deception may exist. Likewise, for our sexuality example, and the case of those societies which have far more blurred boundaries than our own with regard to sexuality. In this regard, Kinsey speaks of (1940s) Italian society in which a man who had active but not passive intercourse with another man was not regarded in his *society* as a homosexual, and thus did not have to *construe himself* as such. Would constructivism of either variety give us a non-depth-psychological account of self-deception—a "single stream," as opposed to a "dual processing" account, as Sackeim puts it?

Sackeim's account actually urges two things, one of which we may concede. He is at pains to point out that if all processing is constructive, one tenet of a classical Freudian model need not be the case. This is the Freudian idea that one "stream" of processing is veridical, neutral, answers to the Reality Principle, sees the world as it really is. He concedes that there are cases which seem to be instances of this (he cites Hilgard's "hidden observer" hypnosis work); nevertheless, on a picture of things in which *all* processing is constructive, there is no particular reason to think that there is *always* a neutral representation of reality located somewhere in the cognitive (or socio-cognitive) system—say, that I *really am* a homosexual, or I *really am* just frightened to face the enemy guns. This point tells against the specifics of Freud's own notion of ego-processing, at least in non-psychotic subjects, and is well taken. But Sackeim runs it together with another point, which is that we may replace the depth-psychological notion of dual (or plural) processing—the mind of "parts," as I put it above—with just a *single stream* of processing. He contends that constructivism provides us with the resources to see that even a single stream of processing may be biased or colored by differing top–down organizing principles—differing schemata, say—and it is in this that the possibility of self-deception may consist.

This second point is not well taken. What is required for this to be an account of self-deception? It is not necessary that one stream processes accurately, granted; but it is necessary for one stream of processing to act in a motivated way to subvert another stream of processing. If I just see myself as straight, even if not accurately processing, I am not *self-deceiving*. If I live in a society which (radically) has no "label" for homosexuality, then, again, I am not self-deceiving. Sackeim's first point suggests that our earlier interpretation of self-deception—whereby one part of the mind *knows that p*—must be re-formulated, say, to holding that one part of the mind must *strongly believe that p*, or something similar. But for there to be *self-deception*, as opposed to merely inaccurate perception, there must then be motivated and effective subversion of that belief. And this still requires the person of "parts." Even if this

subversion of belief occurs not by repression, or one of the classical defense mechanisms, but by this new (and highly-suggestive) means of top–down, constructive, schematic, cognitive, and conceptual re-organization, still, there must be an agency at work with the *motive* and *executive power* to instigate such re-conceptualization—and instigate such in the teeth of a believed-to-be-accurate representation to the contrary. We have, then, a suggestive research program for depth psychology provided for us by modern cognitive and social psychology—one which, in light of the considerable clinical successes of cognitive therapy, we should surely want to explore further—but still no account of self-deception without such a depth psychology.

5. Neo-Sartrean accounts of self-deception

What of the third position, that self-deception is a case of bad faith (Sartre), or a disavowal of engagement (Fingarette)? Sartre's own positive account is not really in a form that permits of clear evaluation as a psychological explanation, for reasons to be given shortly. Fingarette's neo-Sartrean account will therefore serve for us instead. However, before we get to this positive account, Sartre's famous negative criticism of any psychodynamic account will need a brief response. Sartre sought to demonstrate that a putatively depth-psychological explanation of self-deception (Freud's) would actually have to be committed at some level to the same *conscious* coexistence of mutually incompatible beliefs as a non-depth-psychological account, and had merely obscured this fact with a mixture of jargon, and a splitting of the self into parts. But his argument is patently question-begging. Consider his objection to Freud's theory, that for repression to take place, the Freudian censor must first register the information that one part of the mind really has impulse p and that another part of the mind repudiates this (*not-p*), before executing action to prevent the impulse p from becoming conscious. In his attempted refutation of this position, Sartre's question-begging elision from these impulses being *registered* (in the Freudian/Helmholtzian sense) to their having to be *conscious* is indicated here by the relevant words being italicized.

> It is not sufficient that [the censor] *discern* the condemned drives; it must also *apprehend* them as to be repressed, which implies in it at the very least an *awareness* of its activity. In a word, how could the censor *discern* the impulses needing to be repressed without being *conscious of discerning* them? … *All knowing is consciousness of knowing.* (Sartre, 1969, pp. 52–53)

What Freud shares with a non-dynamic notion of the unconscious is that mental processing does not generally require "awareness," much less consciousness. This (obviously true) framework assumption of *all* modern psychology—dynamic and academic—is the answer to Sartre's otherwise rhetorical question [5].

It is worth noting of Sartre that both his positive account and his animadversions against the psychodynamic account answer to constraints which are very different than those of this paper and the psychological literature generally. This paper is concerned with what might, on its own terms, be an adequate psychological

explanation of self-deception. The Sartrean project is strongly constrained by trying to satisfy the Kantian and Cartesian desideratum of showing how we may have sufficient *freedom* for moral and epistemic responsibility; this is its primary aim, and, as a result his own talk of "bad faith," comes to be couched in a form that does not readily permit of appraisal as a prospective scientific explanation [6]. For this reason, we will instead evaluate Fingarette's neo-Sartrean account as a candidate positive alternative to a depth-psychological explanation of self-deception.

Fingarette's (1969) position is that self-deception is a *failure of engagement*: one *does x, y*, or *z* but there is a refusal to avow identity with the self that does *x, y*, or *z*—a *disavowal*. This is a philosophical description of something possessed of important phenomenological resonance; but the psychologist (even one of a theoretical bent) will ask what sorts of processes might implement such a failure of engagement. Fingarette suggests one possibility: a *refusal to spell out*. So, Smith has sex with men, fails to engage with an identity as a homosexual, and the way this disavowal takes place is that Smith refuses to spell out how having sex with men, fantasizing about sex with men, etc., makes him a homosexual [7]. What, though, is necessary for this to be more than a *failure* to see something (say, because of a tumor or stroke), but rather a *refusal* to spell out: a genuine act of *self-deception*? To be self-deceived, rather than merely self-non-perceived, *I* must refuse to spell out my identity, the meaning of my engagements in the world. That is, this "failure" is more than just a failure, it is a motivated act, a *refusal* ("I don't want to see that I am a homosexual, I don't want to spell this out to myself!"). This means that it must have been spelt out already how spelling out would lead to the forbidden discovery. As Sartre repeatedly points out (he thinks, for some bizarre reason, *contra* Freud), some part of the mind must possess the *knowledge* that *p*, and the *motive* to conceal that *p*, and the *ability* to conceal that *p*: here by the means of obstructing a "spelling out" from taking place by another, second, part of the mind—a spelling out, which, of course, must have taken place already, though in a way that is concealed from that second part of the mind or the "mind as a whole." Just as for the other explanations considered already, the "failure to spell out" is both a phenomenologically resonant idea, and also highly suggestive if considered as a prospective (here) psychological device for implementing self-deception. But it is a depth-psychological mechanism, albeit a new one, to be added to the list of defense mechanisms we have encountered thus far [8].

6. A general criticism of any non-depth-psychological account of self-deception

We now have sufficient material on the table for a general pattern to have emerged, and with it, a general criticism of the non-depth-psychological approaches considered. Each of these approaches is vulnerable to a kind of dilemma argument. Each takes something—some stage or type of processing σ—that is necessary for the knowledge that *p*, and situates its theory of self-deception in the withdrawal or otherwise thwarting of this processing. So, for the case where σ is the stage of processing concerned with direction of attention: the knowledge that *p* requires

attention to a body of evidence. We undermine self-knowledge by not adequately directing attention to that evidence. Or, σ may be the stage of processing concerned with "spelling out": we undermine self-knowledge by preventing the stage of processing concerned with spelling out that p. Or, σ may be the processing involved in top–down schematic construction: one prevents knowledge that p by not deploying a p-apt schema. Other, related, possibilities here are avoidance of inference, disregarding an implication, failing to gather evidence or failing to appropriately weight evidence that has been gathered, failing to integrate the material into one's acknowledged self-identity, and numbers of others besides.

Whatever may be the stage or type of processing, σ, the withdrawal of which prevents the knowledge that p, is this withdrawal a *motivated, purposive act, in pursuit of an interest by the agent*? If it isn't, we have self-misperception, but not self-deception—and these accounts are explanatorily inadequate for their intended purpose. A stroke or tumor victim unable to attend to p is not deceiving himself about p; and likewise for non-organic, psychogenic (but unmotivated) causation: as long as this failure of processing is just that, a failure, rather than (Fingarette) a *refusal*, we may have a (Helmholtzian) unconscious account of self-misperception, but not a (dynamic) unconscious account of self-deception.

What, then, if these accounts take the withdrawal of processing stage σ, to be *motivated, purposive*? In each such case, a part of the mind, σ, the output of which is necessary for knowledge that p, is obstructed because it is *known* that σ-processing, if not obstructed, will lead to knowledge that p. But then, it is necessary for σ-processing *already to have taken place*, since, by hypothesis, σ-processing is *necessary* for just that knowledge (that p) which had to be possessed for the subversion of σ-processing to be a *motivated act, in pursuit of the agent's interests*. And that just means that σ-processing, though it must take place even to be obstructed, must do so with its output diverted away from a further stage of processing (or "part of the mind")—one normally constitutive of a certain type of awareness—by an executive controller [9] which appropriates its output to itself alone, and for a purpose of its own. Consider, however, what this now involves: a person of parts, these parts with their own motivational interests, whose activities and motivations are not necessarily known to, or shared by, the other parts, because these parts have means of selectively concealing, revealing, and deceiving the other parts, etc. Any such account *just is* a depth-psychological explanation. And this conclusion is not obviously vulnerable to *any* new, exciting, details about *what σ*, and σ-processing, is: σ may be attention, spelling out, avoidance of inference, etc.; or σ may be a processing stage which implements any of the classic Freudian defense mechanisms (to the extent these be distinct); or σ may be another, to-be-discovered mechanism besides.

7. The "different logic" approach to self-deception

However, this conclusion needs considerable qualification and defense in the light of the more structural responses—especially position (5) above, the subject of this section. A way into this position is to point out that our conclusion here should really be conditional: *if* self-deception exists then a depth psychology is required.

Position (4) is then the simplest, and least serious, of these more structural positions: that the phenomenon of self-deception—or, perhaps, "real" self-deception—does not exist.

There may be something to position (4)—perhaps some of the cases we would want to say are self-deception are in part deception of others. Besides, after G.H. Mead, one may see the self as developing from others, and thus this distinction itself as not clear-cut. However, unless merely a terminological variation on position (5), it seems absurd to suppose that *every* putative case of self-deception can be dismissed thus; and if only many such cases are, this turns into position (6), the subject of the final section.

Position (5) is the most serious challenge: that while self-deception as construed above on the model of other-deception indeed does not exist, this is simply the wrong way to characterize matters. Remove from the characterization of self-deception those features distinctive of other-deception, the need for a depth-psychological explanation vanishes. What are these "other-deceptive" features? Mele, the major proponent of position (5), labels them as *static* and *dynamic* according to what type of "paradox" they lead to. His own conception of these other-deceptive assumptions is contentious; I would instead favor something like the following:

- Static assumption: in self-deception, as in other-deception, something is known (believed, represented) which strongly *conflicts with* something else believed or represented—as, say, where p and *not-p* are represented in the same person, at the same time. For many (e.g. Mele, 1987, 1997, 2001), the static assumption is met only if psychological conflict is as severe as this: at the level of contradictory beliefs [10].
- *Dynamic assumption*: in self-deception, as in other-deception, the false belief "not-p" must be brought about through a motivated and directed *act*, which is the exercise of *agency*, for the *purpose* of achieving some *interest* of that agent. (Mele, in what I argue is an unhappy turn of phrase, repeatedly glosses the dynamic assumption as being that self-deception is "intentional.")

Clearly there is no "static" or "dynamic" paradox for so long as there are two separate persons, Smith and Jones—one knowing p, leading the other to believe *not-p*. When the static conflict is within only the one person, Smith, at a time, though, it requires us to take at least one of these beliefs/representations to be *unconscious*. When the dynamic conflict is within only the one person, it further requires us to invoke the notion of a *dynamic* unconscious. So, if, for the special case of *self-deception*, one could abandon the static and dynamic assumptions, there would cease to be a paradox requiring a depth psychology for its solution.

At the level of psychological detail, how one accounts for deception without these assumptions is accomplished in somewhat different ways, for the different σs, as I have put it above. Certainly, Mele considers attention, and spelling out, among a number of other processes. I will work through his account mostly as it applies to attention, since this explains clearly his general strategy, as well as arguably offering his strongest case.

Consider his example of Beth, a child mourning her recently deceased father,

self-deceived in thinking that p: he loved her best of all his children. Although self-deceived, Mele denies that Beth possesses the co-occurrent contradictory belief that $not\text{-}p$ (his strong take on the static assumption). He further denies that she directs any processes through unconscious agency to achieve that deception (the "dynamic assumption"). The deception is effected by Beth's attention being directed to memory evidence favoring her desired conclusion that p, and away from memory evidence favoring that $not\text{-}p$. But how can this process be *directed* if not by a *director*, a locus of *agency* (the "dynamic assumption")? And how can some memory evidence be unattended to, if not by seeing that it favors $not\text{-}p$, the conclusion to be avoided (the "static assumption")?

These questions are answered by Mele with an account based on operant learning theory (though he does not himself use this term). Memories of episodes when Beth's father played with her alone are intrinsically rewarding ("comforting"); memories of him playing with her brothers at least as much as her, are neutral; and of him playing with her brothers to the exclusion of her, intrinsically aversive. So, by rewards, extinction, and punishment, which are intrinsic to (what she takes to) the material, Beth's attention to memory evidence is *conditioned*. When, subsequently, she tries to make an honest assessment of her father's relative affections, she can do so only on the basis of memory evidence which has been thus conditioned. This evidence, on which she must base her judgment, is now weighted at retrieval to favor the false conclusion that p: "Daddy loved me best." This is a single stream account. There is no controller-as-agent directing attention away from $not\text{-}p$, and towards p; rather, there is the direction of attention without agency, through rewards and punishment intrinsic to (that subject's processing of) the material.

It is a major achievement of Mele's to have shown, with cases like that of Beth, how self-deception is conceptually possible, and probably sometimes actual, without the static or dynamic assumption being satisfied, and thus without necessarily involving a depth psychology. He devotes substantial space to addressing opponents whom he anticipates will object *a priori* that cases like Beth's cannot be *real* self-deception without these assumptions being met. I am not among those he needs to convince here, but his arguments are from Ockham's razor (better, in this context: Lloyd-Morgan's Canon) that a theory must explain the data, and if it can do so most *simply*, it is to be preferred. For the case of Beth, it is reasoned that since a neo-behaviorist learning theory of attention direction is simpler than a depth-psychological theory of attention direction, we have no reason to look beyond the former. This defense indicates a point of potential vulnerability also, for Ockham's is a double-edged razor: there may prove to be other cases where the *simplest* account of the data indeed requires us to posit a depth-psychological apparatus.

Consider then, his case of Sam (Mele, 1997, 2001). Sam's wife, Sally, is unfaithful to him ($not\text{-}p$). Sam is presented with a body of evidence supportive of this conclusion: Sally's "working late" several nights a week, she's been seen in a theatre and a local bar in the company of one "Mr. Jones," etc. *Ex hypothesi*, this evidence normatively should convince Sam, and would convince Sam were he not personally involved. But he is, and it does not; he is self-deceived into believing his wife is faithful (p). Various specific processes (σs) are suggested as possibilities for this

self-deception, among them two that we have considered: that he does not ade-quately *attend* to the evidence for *not-p*, nor *spell out* its meaning—that *not-p*. As for the case of Beth, the idea is that self-deception does not require the static or dynamic assumption to be met. We saw how that could indeed be the case for Beth. How would such an account go here? Mele (1997 p. 95), anticipating a line of attack, says:

> Evidence that one's spouse is having an affair [*not-p*] … may be threatening even if one lacks the belief that that evidence is stronger than one's contrary evidence [*p*].

But why is this evidence "threatening"? What it is for something to be *evidence* at all is that it contributes warrant towards some conclusion—normatively or descrip-tively, or here, presumably, both. The data as such are not intrinsically appetitive or aversive—there's nothing intrinsically aversive about Sam's wife working late (rather, perhaps, the reverse). And there's nothing intrinsically aversive about his wife going to the theatre or a bar with a man—say, with a theatrical gay friend, or with her brother. The data only are aversive in as much as their *meaning* has at some level been attended to, and spelt out: that these events are *evidence for adultery*. This line of reasoning may, minimally, apply to Beth's case also, but at a much lower level. Note the contrast here. Beth's memories of playing with daddy are intrinsically rewarding, hence attention-reinforcing, independent of the conclusion, *p*, they will eventually contribute towards (albeit that the conclusion "daddy loved me best," once established by these memories, might feed back to redouble the rewarding nature of the subsequent retrieval of them). Her memories of relative neglect, when he played instead with her brothers, are intrinsically less rewarding/aversive, hence attention-reducing; and this substantially independently of the conclusion, *not-p*, they would otherwise contribute towards (albeit that there might be a sense in which the jealousy she feels on retrieving memories of daddy playing with her brothers is bound up with the proposition *not-p*: "daddy doesn't love me best"). However, in the case of Sam, the whole of the appetitive/aversive powers of the evidence emerges *after* processing of what it means. It is conatively neutral but for that processing having taken place.

Sometimes "evidence," memories, a line of thought, etc., can be treated (for our purposes) as just a "stimulus," intrinsically aversive or appetitive. But sometimes one has to go down the line of thought before it can be aversive or appetitive—the aversiveness is consequent upon the ratiocination; it is intrinsic to, or is internally connected with, a set of inferences having taken place. And sometimes matters will lie somewhat in between these two possibilities. All conation may require *some* processing, but some requires more than others. Sam's case requires inference to the effect *that* this evidence urges the conclusion *not-p*, in order to motivate aversively. Certainly in its positive aspect, Beth's happy memories of daddy playing with her do not require the conclusion "daddy loved me best of all" to be pleasant. The latter case clearly permits of a more bottom–up explanation than the former.

Here we connect with several important themes from the history of psychology; though the specifics of the attentional account just considered need to be situated

within the general aim of Mele's project before these connections can be made. Mele emphasizes the *opacity* of his "intentions" [11]—as when one intends *to kick the ball*, but not *to break the window* which follows from this act. Mele wishes to show generally, that the mental processing which underlies self-deception may be motivated ("intentional"), but not motivated by the desire ("intention") to avoid the belief that *p*—only by the desire to avoid something else, which turns out to be extensionally (materially, or by causal consequence) equivalent to this. Intentional contexts are *intensional* contexts, that is, they are truth-substitutionally opaque. So, Beth wants *to comfort herself* (with happy memories). In so doing she brings about the false belief *that daddy loved me best*, but this latter was not her intention. Sam is held to intend to *avoid evidence*, not to intend to avoid the conclusion ("my wife is unfaithful") which this evidence warrants—though this latter (belief avoidance) is brought about by the former (evidence avoidance). This point about opacity is intended by Mele to undermine both the static and dynamic assumptions. If the agent doesn't believe both *p* and *not-p*, but only *p* and something which brings about, or is extensionally equivalent to *not-p*, then there is no more a (static) paradox than if Lois Lane believes she has kissed Superman and not kissed Clarke Kent; nor yet (dynamic paradox) if she *intends* the former and not the latter. If, in this way, neither the static nor dynamic assumptions are satisfied, the depth-psychological apparatus needed to avoid the static and dynamic paradoxes becomes otiose.

Although this point about the intensionality of intentionality is sound and important in the abstract, applying it to reassign cases formerly considered to satisfy the static and dynamic assumptions carries clear danger. One can always find *some* extensional re-description. The lesson of history is that we should treat a general, theoretical, pressure to re-describe in non-intentional terms with great caution. In the history of psychology, traditions have differed over whether, generally, they sought after more distal or more proximal stimuli as springs to action. The reductionist, bottom–up tendency is to search for ever more proximal stimuli—the rat conditioned at the level of bodily movements, or the cat wanting to claw the lever. A rat may be seen as motivated to *follow the scent of food* without being seen as motivated to *escape from the maze*—though doing the former will lead to the latter. It may be seen as an *error*, even (a "stimulus error") to invoke an intentional idiom and describe the rat as seeking after "escape," or to *plan his way out of the maze*. But one may also be motivated top–down and not bottom–up (Thorndike's cats wanting *escape*, not to claw the lever, Tolman's rats having a *goal-directed* mental map of their intentional object).

There is a reductionist sense of "simple" in which, however contrived, a proximal explanation of a phenomenon (say, Skinnerian behaviorism, or Titchenerian structuralism) will always be simpler than its more distal competitors. But in an *explanatory* sense it may not—which is why Tolman could justifiably claim to be the true Ockhamist in holding that his rats' maze-running behavior could only be adequately explained by postulating goal-directed (albeit Helmholtzian) unconscious representations. How proximal/distal a spring to action is will vary according to the case in point. There are highly unusual cases where the cat seeks only to claw the catch as end in itself, but most commonly the cat seeks *escape*, with catch-claw-

ing better described as a *means* to that end. Mele tells us that one may be "intentionally ignoring unpleasant evidence … without intentionally deceiving [one-self]" (1997, p. 130). Agreed—as his case of Beth indicates: if the evidence is intrinsically aversive, that is, aversive independent of the conclusion it warrants, this is possible. However, in most cases, one does not seek to *ignore evidence* as an end in itself (any more than, conversely, one might seek to "be deceived" as end in itself). One avoids evidence *because* of what it *means* ("my wife is unfaithful!") with the condition of *being deceived* as the necessary concomitant of the pursuit of that end—avoiding that conclusion.

Sometimes we need to account for behavior only by postulating unmotivated self-misperception (what Mele calls "cold bias"); and sometimes by postulating motivated, but non-depth-psychological bias (perhaps Beth). Sometimes, however, Ockham's razor requires that we postulate strong self-deception on the model of interpersonal deception.

When, then, would we be justified, on explanatory grounds, in going for a Mele-style account, and when would we be justified, on just those same grounds, in insisting on a psychodynamic account? This is an important and complex question; a number of factors come into play. Although there is no strict criterion, thought experiments can make clear that there are principled reasons to be brought to bear here. Consider the following case: At an early age, Michael sees his mother lose a terrible, drawn-out battle against bowel cancer. It does a lot to him, psychologically. He avoids doctors, hates hospitals, will sternly silence his family when conversation turns to matters medical over the dinner table, etc. Should such topics arise on the TV or radio he will switch over, and will self-censor his newspaper and magazine reading. Should he begin to present symptoms requiring medical consultation—say, an itchy, raised, irregular mole—he may not take the necessary steps until it is too late, until he is dying, even. However, for this case, he needn't necessarily have drawn the conclusion at some level ("I've got a suspected malignant melanoma!") for the unconscious dread of this diagnosis to prevent him from obtaining early diagnosis. It is possible rather that the direction of cause is from his standing directive against all matters medical, preventing him from knowing the much publicized self-examination procedures conducive of early diagnosis, or preventing him from consulting a doctor about anything—even what he believes to be a minor matter. Explaining his behavior may only require us to postulate a motivated bias against all matters medical, which achieves for him the regrettable ignorance of his own condition.

But now imagine that his mother's death had been from a malignant melanoma, or that Michael, in later life, had shown symptoms of bowel cancer. It is not nearly so plausible that the avoidance behavior is to be accounted for without postulating that he drew the specific inference, then fled from it. The prior, psychological *competence* needed to draw the forbidden conclusion is provably present, and we must explain the performance error.

Note that such competence is of course present in *degrees*. Signal Detection Theorists [12] talk of "discrimination" (d'). This refers to the ability of the "receiver" to detect the difference between signal and noise versus noise only—

whether the signal is a set of medical symptoms, a pure tone in an audiometer, a body of evidence for adultery, etc. Discrimination (d') is a function of signal strength and receiver sensitivity. Whether the receiver *actually* claims a signal is present depends on another factor also—their *bias* (β). Where discrimination is high—a big noise in the earphones of a man with excellent hearing, a textbook case of advanced neoplasm presented to a consultant oncologist, one's spouse *in flagrante delicto*—a failure to register a "hit" requires massive bias. These are unlikely to be cases capable of explanation by self-deception, the level of discriminability means that the signal's presence was almost certainly registered in consciousness: the bias needed for a "miss" response is then most likely downright lying, or at best perhaps hysteria [13]. When discriminability is very low, again, self-deception is not obviously warranted—the data (for adultery, cancer, etc.) isn't there to provoke even an unconscious representation that might call for concealment. Responding is explained by standing response biases, but without detection (hence representation) of the stimulus. But in between these cases lie cases like Sam's, or perhaps the latter version of Michael, which indeed require us to posit that the receiver registered the presence of the stimulus below the level of consciousness, then thwarted some processing step needed, and normally present, to bring that signal to consciousness.

This talk of error in the face of a proven psychological competence, re-emphasizes that willful ignorance cannot be defined out of classification as problematical, "paradoxical" self-deception, merely in virtue of it stopping the agent just short of the conclusion, *not-p*, which would generate *formal* contradiction with *p*. In self-deception, certain kinds of profound psychological conflict (the "paradoxes," however construed) generate the need for a depth psychology. Talk of strict contradiction, "at the level of intentions," etc., merely represents various (questionably relevant) attempts to capture a very precise version of such conflict for use in philosophical analysis; but such talk can too easily obscure the fact that there isn't a *logical* puzzle of self-deception, there's a *psychological* one [14]. If one "solves" the paradoxes of self-deception by requiring the person to internalize and then manage powerful psychological conflict—albeit conflict which stops fractionally short of some proprietary definition of static or dynamic paradox—then still, the depth-psychological apparatus needed to deal with such conflict is by no means otiose. We are doing theoretical psychology, not philosophy of language here. We need to explain how and why the exercise of an obvious, transparent, normally-present capacity to see that *this* strength of evidence is adequate for *this* conclusion, nevertheless becomes thwarted. We need to explain how and why one undermines the equally obvious capacity to see that *this* locution is extensionally equivalent to, and only barely, notionally, distinct from, this other, forbidden locution—and so on for our other cases. What, after all, are these (thwarted) capacities but themselves precisely the kind of psychological processes (σ) whose motivated obstruction gave our argument its point of departure [15]?

Mele establishes, with cases like Beth, that some instances of genuine self-deception may be capable of being accounted for without postulating a depth psychology. That is a major accomplishment. It establishes, for one thing, that a depth psychology is not *conceptually* necessary to account for self-deception. But other

cases, like Sam's, do require us to invoke a psychodynamic explanation. Recall, however, that position (5) must deny that there are *any* cases satisfying the static and dynamic paradoxes—were there merely many which did not, this would be a discussion of (6). It is this position then which we turn to.

8. Conclusion: on the importance of depth psychology

The final response option should be read in the light of the point already made: that these positions are not mutually exclusive; and then in light of the fourth response option—that many apparent cases of self-deception might really be other-deception (or self-misperception). The adherent of (6) accepts that certain phenomena exist which require a psychodynamic account; but holds that, on examination, these will prove to be sufficiently rare as to relegate depth psychology to being of minor importance [16].

Here though, note that the strength—the "depth"—of the depth-psychological system we must postulate will vary with the power of the phenomena we are called upon to explain; and that, in this paper, the need for some kind of depth-psychological *explanans* has only been argued for through considering the *explanandum* of self-deception. As was pointed out above, there are a range of dissociative clinical and experimental phenomena greatly more startling than most cases of self-deception. If even the latter requires some kind of depth psychology, then, *a fortiori*, so will the former. Still, this objection will be that all the phenomena calling for depth-psychological explanation put together will still not be sufficient to make the neglect of depth psychology by academic and clinical psychology a major oversight or omission.

What this has to do with, in essence, is the issue of *psychic determinism*. Despite the emphasis of many commentators, *biological* determinism is largely a red herring in seeking to understand mature Freudian theory—being neither necessary nor sufficient for Freud's approach, much less for other psychoanalytic accounts. It is not sufficient, in that numbers of decidedly non-*depth*-psychological accounts of phenomena that Freudian theory competes to seek to explain are nevertheless biologically determinist in nature. It is not necessary, in that in the practice of providing psychoanalytic explanations, biological determinism plays no significant, functioning, explanatory role. However, for Freud and most other depth psychologists, *psychic* determinism is a given. An excellent example here is from the *Introductory Lectures*, where he considers objections to his depth-psychological account of forgetting and the "parapraxes" (e.g. slips of the tongue—"Freudian slips"). Cannot such forgetting, and such slips, occur for purely physiological ("somatic") reasons, he asks? He considers numbers of such reasons, reasons located in (Helmholtzian) unconscious neuropsychological or cognitive processes [17].

> The influence on the production of slips of the tongue by physiological dispositions must be recognized at once … [b]ut how little they explain. … These somatic factors only serve … to facilitate and favor the particular mental [*depth-psychological*] mechanism of slips of the

tongue … Suppose that one dark night I went to a lonely spot and was there attacked by a rough who took away my watch and purse … I laid my complaint at the police station with the words: "loneliness and darkness have just robbed me of my valuables." The police officer might then say to me "in view of what you say you seem to be unjustifiably adopting an extreme mechanistic view. It would be better to represent the facts in this way: 'under the shield of darkness and favored by loneliness, an unknown thief robbed you of your valuables'" … Such psycho-physiological factors as excitement, absent-mindedness and disturbances of attention will clearly help us very little towards an explanation. … The question is rather what it is that has been brought about here by the excitement, that particular distracting of attention. And again, we must recognize the importance of the influence of sounds, the similarity of words and the familiar associations aroused by words. These facilitate slips of the tongue by pointing to the paths they can take. But if I have a path open to me, does that fact automatically decide that I shall take it? I need a *motive* in addition before I resolve in favor of it and furthermore a *force to propel me along the path*. (Freud, 1917/1978, pp. 72–73)

For Freud, whether in forgetting, slips of the tongue, dreaming, a failure to pay a bill, or any and all of the whole rich course of one's life, *psychic* determinism applies: the search for a *motive* not merely a cause (though also a cause)—the motive that is the *force which propels me along the path*. The objection we are entertaining now is simply the rejection of this psychic determinism. Not the view that there isn't *any* depth-psychological determination of action—whether in forgetting, slips, dreaming, self-deception, or neurotic symptom formation—merely that there is too little to make depth psychology a very central area of psychology as such.

The first thing to concede is that Freud's total psychic determinism is false, and indeed, wildly false. The century and a half of academic psychology since Helmholtz has involved the search for causes for many of the quotidian and clinical phenomena of which depth psychology treats; and much of this search for causation sans motivation has been fairly successful. To that extent the Helmholtzian and Freudian unconscious are in competition: the success of explanations couched in terms of the former is apt to remove much of the point to a search for explanations couched in terms of the latter. (Consider, as example, the relative successes of modern memory research in the teeth of Freud's extreme view that, saving organic damage, we always need a motive to completely explain any case of forgetting.) Nevertheless, the two kinds of unconscious explanation are not in any necessary competition: reasons can also be causes, as Freud certainly emphasized, *psyche* shades into *soma*; and the biological or cognitive unconscious can be the implementation mechanism for something which at another level of explanation is *meaningful, symbolic, motivated*. Since nearly everyone will accept this for the case of an implementation mechanism for conscious meanings, symbols, motives, there is nothing fundamentally baffling about the idea that (dynamic) unconscious processes should similarly comport with (Helmholtzian) unconscious processes. It is just that *how much* of our mental lives

needs depth-psychological explanation is now very much at issue: a good deal less than Freud maintained, but a great deal more than the nothing which mainstream academic and clinical psychology has maintained for more than a century of Freudian influence in other disciplines and everyday life, including the everyday lives of mainstream academic and clinical psychologists. But then, there is dissociation found in more areas of all our lives than perhaps we would wish to investigate.

Acknowledgements

Notes

[1] Not only by psychologists: see Grünbaum (2001, p. 131) and Gergen (1985).

[2] This comes from Popper, who contrasts psychoanalysis with Marxism, which latter he held to have been originally testable, hence scientific, though subsequently rendered "immune" from such tests by the activities of Marxists after it had been refuted. Psychoanalysis, rather, was "immune to start with and remained so" (1983, p. 128).

[3] If a psychoanalytic theory were developed which satisfied modern academic psychology's methodological standards, a predictive judgement of mine is that the details of this or that existing psychoanalytic theory would be unlikely to survive, except as syndromes of clinical interest. Kline (1981, 1984), for example, sought to establish the existence of the *Anal personality* by using factor analysis. Such work might establish the syndrome, without vindicating the entire Freudian psychosexual stage theory of development, which gave it its logic of discovery. This point would likewise apply to the theories which gave rise to talk of superiority/inferiority complexes, universal bisexuality, the collective unconscious, and many other suggestive clinical syndromes.

[4] This term used descriptively, not invoked as *explanans*, which later is a committed, controversial, position (Hilgard, 1977).

[5] The line of reasoning Sartre demonstrates here is by no means one that is restricted to himself, nor yet to continental philosophy (compare with Gergen, 1985, p. 232).

[6] I embrace, actually, this Kantian and Cartesian tendency (called, in the analytic tradition, epistemic internalism), but disagree that it would be jeopardized by a psychology of the unconscious mind.

[7] Compare with Sartre (1969, p. 63).

[8] As Sackeim puts it: "in strictly logical terms Sartre's resolution is actually quite Freudian. It would seem to make little difference whether we speak of unconscious ideas or nonthetic knowledge. We are still stuck with the problem of individuals believing or not believing at the same time." Quite (Sackeim, 1988, p. 152). Fingarette (1969, p. 137) is more qualified than Sackeim in accepting this point.

[9] Talk of a "controller" implies nothing more substantial than reference to the presence of an act or acts of executive control exercised by some cognitive part, at a time, in the explanation of any given case of self-deception—in other words, it does not imply (or exclude) the presence of a single, general, Freudian Censor. Much of the time, Freud's own explanations don't imply the existence of the Freudian Censor either. Overemphasis on such a single, univocal, Censor anyway sits ill with the hugely *dynamic* aspect of psychodynamic theory.

[10] The presence of contradictory beliefs is both too strong and too weak to be *criterial* for satisfaction of the static assumption. Mere contradiction, somewhere in the totality of a large, and largely inert or unconnected belief set, is too weak. I give reasons below as to why it is too strong.

[11] I find Mele's gloss on the dynamic assumption in terms of "intentions" most unsatisfactory—and

thus put this term in scare quotes, precisely because of the issue concerning unconscious "intentions" flagged above (with regard to Sartre, and the danger of begging questions). When read in terms of the philosophers' "intentionality" (aboutness-as-such) it is a harmless locution, however, and Mele's arguments only require us to read it thus.

[12] At the level of detail, Mele's current theory of choice for explaining self-deception appears to be a type of signal detection theory, called the PEDMIN analysis—though to my knowledge neither he nor the progenitors of this theory in social psychology refer to it as such. Considerations of space prevent me from dealing with this further.

[13] Although either of these cases (and especially the latter) may plausibly require a depth-psychological explanation, they do not do so as evidence for the presence of self-deception.

[14] Consider Mele's suggestion that "Sam's aim may simply be to put off for a while the painful process of reflecting on a painful prospect," as opposed to, "*trying to protect his belief in Sally's fidelity* and trying to deceive himself into holding onto this belief" (Mele 2001, p. 59). That the former is not intensionally equivalent to the latter is well taken; my point is that the latter is (normally) *psychologically prior*. Mele has indicated (personal communication) that he will take issue here.

[15] Thus, it is worth asking whether extensional re-description, rather than representing a neutral, schematic way of grouping specific alternatives to a depth psychology, should instead be taken merely to be itself one, specific, implementation hypothesis for precisely just such a depth psychology. Considered in this way, its similarity to the "constructivist" stratagem is striking: Joe is not *a homosexual*, say; but Joe *has sex with men*.

[16] A view which can be treated as a variant on this, develops from any position which holds that the term "self-deception" does not pick out some one, single, phenomenon needing to be explained. Any such position can be assimilated to (6) as follows: if my arguments in response to position (5) hold up, there are *some* cases subsumed by the term "self-deception," the explanation of which requires us to postulate a depth psychology. Position (6), then, just becomes the view that *most* of the heterogeneous phenomena covered by this term don't require explanation by depth psychology.

[17] An example of currently influential work along these lines is the information-processing approach to the "psychopathology of everyday life" of Reason (1992) and Norman (1981).

References

BECK, A.T. (1976). *Cognitive therapy and the emotional disorder*. New York: International Universities Press.

BECK, A.T. (1991). Cognitive therapy. *American Psychologist, 46*, 368–375.

BROADBENT, D.E. (1958). *Perception and communication*. Oxford: Pergamon.

CHERRY, E.C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America, 25*, 975–979.

CRICK, F. & MITCHISON, G. (1983). The function of dream sleep. *Nature, 304*, 111–114.

CRICK, F. & MITCHISON, G. (1986). REM sleep and neural nets. *Journal of Mind and Behavior, 7*, 229–250.

DENNETT, D. (1978). Artificial intelligence as philosophy and as psychology. In *Brainstorms*. Brighton: Harvester Press.

DIXON, N.F. (1981). *Preconscious processing*. Chichester: Wiley.

EVANS, C. (1984). *Landscapes of the night: how and why we dream*. New York: Viking Press.

FINGARETTE, H. (1969). *Self-deception*. London: Routledge & Keegan Paul.

FODOR, J. (1968). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy, 65*, 627–640.

FREUD, S. (1917/1978). *Introductory lectures on psychoanalysis*, J. STRACHEY (Trans.). Harmondsworth, Middlesex: Penguin.

GERGEN, K. (1985). The ethnopsychology of self-deception. In M. MARTIN (Ed.) *Self-deception and self-understanding*. Lawrence, KS: University Press of Kansas.

GRÜNBAUM, A. (2001). Does Freudian theory resolve "the paradoxes of irrationality"? *Philosophy and Phenomenological Research*, *62*, 1.

HILGARD, E. (1977). *Divided consciousness: multiple controls in human thought and action*. New York: Wiley.

KAHNEMAN, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

KLINE, P. (1981). *Fact and fantasy in Freudian theory*. London: Methuen.

KLINE, P. (1984). *Psychology and Freudian theory*. London: Methuen.

MELE, A. (1987). Recent work on self-deception. *American Philosophical Quarterly*, *24*, 1–17.

MELE, A. (1997). Real self-deception. *Behavioral and Brain Sciences*, *20*, 91–136.

MELE, A. (2001). *Self-deception unmasked*. Princeton, NJ: Princeton University Press.

NEISSER, U. (1967). *Cognitive psychology*. New York: Appleton Century Crofts.

NIETZSCHE, F. (1886/1978). *Beyond good and evil*, R.J. HOLLINGDALE (Trans.). Harmondsworth, Middlesex: Penguin.

NORMAN, D.A. (1981). Categorization of action slips. *Psychological Review*, *88*, 1–15.

POPPER, K. (1983). The problem of demarcation. In D. MILLER (Ed.) *A pocket Popper*. London: Fontana.

REASON, J.T. (1992). Cognitive underspecification: its variety and consequences. In B.J. BAARS (Ed.) *Experimental slips and human error: exploring the architecture of volition*. New York: Plenum Press.

SACKEIM, H. (1983). Self-deception, self-esteem, and depression: the adaptive value of lying to oneself. In J. MASLING (Ed.) *Empirical studies of psychoanalytic theories*, *Vol. 1*. Hillsdale, NJ: Analytic Press.

SACKEIM, H. (1988). Self-deception: a synthesis. In J. LOCKARD & D. PAULUS (Eds) *Self-deception: an adaptive mechanism?* Englewood Cliffs, NJ: Prentice Hall

SARTRE, J.P. (1969). *Being and nothingness*, H. BARNES (Trans.). London: Methuen.

SHORTER, E. (1997). *A history of psychiatry: from the era of the asylum to the age of Prozac*. New York: Wiley.