

Bad Faith and the Unconscious

Jonathan Webber

Penultimate draft.

Please cite only the version published in *The International Encyclopaedia of Ethics*, ed. Hugh LaFollette, John Diegh, and Sarah Stroud (Wiley-Blackwell, 2012)

Self-deception is a complicated and puzzling phenomenon. How is it possible to actively hide something from yourself? Philosophers find this question interesting for two main reasons. One is that the answer promises to be informative about the mind more generally, or at least it seems that any theory of the nature and structure of the mind that cannot account for self-deception must be wrong. The other is that the ethical concern with how one should live and the more specific moral concern with right and wrong actions seem to require us to evaluate our own lives and those of others in an honest and unprejudiced way, a requirement we would be better equipped to fulfil if we had a clear idea of the ways in which self-deception about these things can occur. There have thus been major philosophical discussions of the nature of self-deception in both the anglophone and the European traditions of philosophy over the past century. However, these two traditions of thought have taken crucially different examples as their paradigms of self-deception.

Recent anglophone discussion has focused on the idea that one can deceive oneself into believing something. Self-deceptive belief seems paradoxical because of the features that distinguish deception from mere error. If you deceive me, then you bring me to believe something that you yourself do not believe. You might bring me to believe something that you know to be false, or bring me to believe something that you merely believe to be false, or bring me to believe something that you have no real opinion about whether it is true. When the deceiver and the deceived are the same person, it is puzzling how this condition can be met. How can you bring yourself to believe something that you do not believe? Can you inculcate in yourself a belief that is sufficiently robust to survive the continuing lack of evidence in its favour, or even the continuing presence of evidence against it?

Theories of self-deception in twentieth century European thought, on the other hand, have focused on the idea that one can hide one's own motivations from oneself. At the heart of Sigmund Freud's psychoanalysis is the idea that behaviour can be motivated by drives and impulses whose existence the agent cannot become aware of without the help of a psychoanalytic procedure. As the psychoanalyst comes close to uncovering these motivations, according to Freud, the person whose motivations they are tends to resist. Such motivations are not merely 'descriptively unconscious', in Freud's terminology; they are not merely motivations of which the agent is unaware. They are 'dynamically unconscious', motivations that the agent actively tries to keep out of conscious awareness (Freud 1923, chs. 1 and 2; see also Freud 1915, § II). This activity of resistance itself is psychologically puzzling. After all, Freud is describing people who have sought the help of a psychoanalyst to uncover the motivations underlying the behaviour that troubles them. Why do they then avoid awareness of those motivations? It seems that either their motivation for repressing these drives and impulses or their real motivation for seeking psychoanalytic help must also be hidden from them. This idea that one can spin a complicated web of internal deceit about one's motivations underwent various transformations during Freud's career. But it is fair to say that it remained architectonic of his own psychoanalytic theory and that of his followers, including most notably Melanie Klein.

The cardinal opposing view in twentieth century European thought is Jean-Paul Sartre's theory of bad faith. Sartre agrees with Freud that people have the capacity to conceal their own motivations from themselves, agrees that people generally do exercise this capacity, agrees that doing so can involve complicated webs of internal deceit, and agrees that people can hide their own motivations so successfully that they require the help of a psychoanalyst to uncover them (1943, pt. IV ch. 2 § I). But he does not agree that this is best explained by the Freudian picture of a dynamically unconscious activity of repression and resistance. He argues that such an account ultimately fails to explain the phenomena of self-deception (1943, pt. I ch. 2 § I). Neither does he agree with Freud's theory of the basic drives and impulses that underlie our behaviour and that we often repress. He considers Freud to have cut off the explanation of behaviour at an arbitrary point (1943, 656, 659-60; tr. 589, 592-3). These criticisms lead Sartre to offer his own account of bad faith in place of Freud's account of the repression of basic drives and impulses, and to argue that the theory of bad faith has much greater explanatory power.

This disagreement between Freud and Sartre concerns one's ability to prevent oneself from forming true beliefs about one's own motivations. This does not require the inculcation of false beliefs, since one might simply form no relevant beliefs at all. The central topic of the recent anglophone discussion of self-deception is therefore extraneous to the disagreement between Sartre and Freud. But it would be a mistake to infer from this that the two issues are entirely independent of one another. For any account of self-deceptive belief-formation needs to explain how one manages to keep one's motivation to deceive oneself hidden from oneself. After all, accidentally giving the wrong impression is not an act of deception; a deceiver must be motivated to deceive. But nobody will be deceived when they are aware of the would-be deceiver's motivation. In order to deceive oneself, therefore, it seems that one must be ignorant of one's motivation to deceive oneself. Of course, one can generally act on a motivation without forming a belief about that motivation. But self-deception seems to require

something stronger than this: it seems to require that one can hide the motivation from oneself, where this means that one prevents oneself from easily forming the belief that one has that motivation. Whereas one's ability to hide one's motivations does not require that one can deceive oneself into holding a particular belief, that is to say, the converse is not true. Self-deceptive belief-formation does seem to require that one can hide one's motivation from oneself.

For this reason, contemporary anglophone philosophical discussion of self-deceptive belief could do well to draw on the careful and detailed theories of self-deception about motivation developed by Freud and Sartre. Neither of these thinkers, however, ever presented a clear and distinct account of this aspect of their theory of mind. Freud's theory of repression and resistance was in continuous development throughout his long career. Sartre's critique and alternative develop across his career, though the basic structures remain those presented in *Being and Nothingness*. As a result, discussions of each of these thinkers, and of their disagreement on self-deception, present a variety of different pictures. In the case of Freud, this is largely due to some unclarity of theoretical and presentational detail on his part and the resulting difficulty of reading his works through the lens of more recent thought about the mind (see Gardner 2000). Unclarity on Sartre's part, on the other hand, is joined by a tendency among his readers to assume that *Being and Nothingness* is a sequential treatise, one that establishes views on distinct topics in each chapter by building on previous chapters, when the book should rather be read as progressively elaborating a single theory none of whose aspects or concepts have been fully explained until the end. In this vein, discussions of Sartre's response to Freud tend to focus on one relatively short passage early on in *Being and Nothingness* entitled 'Bad Faith and Lies' (1943, pt. I ch. 2 § I) while ignoring the much richer and more nuanced discussion towards the end entitled 'Existential Psychoanalysis' (1943, pt. IV ch. 2 § I). This discussion can be fully appreciated only in the light of much of the preceding argument, though it is also illuminated by the discussion of Freud to be found in Sartre's earlier book *Sketch for a Theory of the Emotions* (1939, § II).

As a result of this misunderstanding of the structure of Sartre's book, the opposition between the Freudian and Sartrean accounts of self-deception is often mischaracterised, with Sartre portrayed as balking at Freud's rejection of certain basic tenets of the Cartesian view of the mind. In his theorising about the dynamic unconscious, Freud has described purposive mental activity that occurs in a mechanistic fashion without any self-awareness. Sartre, according to this portrayal, wants to retain the identification of mentality with consciousness and to retain the dualistic denial that anything mechanistic or material could be part of the mind. Sartre's own descriptions of self-deceptive behaviour may be illuminating, it is argued, but ultimately the full range of self-deceptive strategies and the complexities of the human mind cannot be fully understood unless we abandon these Cartesian shibboleths as Freud has done (e.g. Soll 1981; Neu 1988). It is unfortunate that this reading of the early stages of *Being and Nothingness* exercises such influence over the understanding of the disagreement between Freud and Sartre on self-deception. For the sophisticated discussion of Freud towards the end of the book draws on intervening chapters that repudiate the idea that mentality is to be identified with consciousness as this is usually understood, and certainly as it was understood by Freud, and that argue that we cannot understand the human mind unless we see it as constituted by the living human body (see esp.: pt II ch. 1 §§ 1-5; pt. III ch. 2 § I). As we will see, the disagreement between Freud and Sartre is precisely the opposite of the way it is usually presented to be. Sartre is attracted to Freud's insights, but considers the Freudian account to retain an unacceptable vestige of the Cartesian view of the mind, a dualistic opposition between rational conscious thought and mechanistic drives and impulses. Sartre presents a more comprehensive rejection of Cartesian dualism than does Freud.

Freud's understanding of the mind rests on a distinction between two kinds of mental item (1923, chs. 1 and 2; 1915, §§ V and VI). One kind are sensitive to evidence of the nature of reality, are inferentially and evidentially related to one another, and are thereby available for beliefs to be formed about them, beliefs that can then be communicated. The other kind do not stand in any rational relations with one another or with anything else. They are drives and impulses which combine and compete according to

the laws of their 'cathectic energy'. Items of the second kind can motivate actions that aim to satisfy them. So the sex drive, for example, is manifested in sexual activity. This is unproblematic so long as the person accepts the drive and can reasonably engage in the behaviour that would satisfy it. Repression is the refusal to allow a drive to factor into one's behaviour. Freud sees this repression as a form of self-deception. Rather than acknowledging the presence of the troublesome drive or impulse, according to Freud, repression typically involves the denial that one possess it at all. The repressed drives are now 'dynamically unconscious'. But they do continue to be drives towards the relevant behaviour. Because the person will not allow actions aimed at satisfying them, according to Freud, these drives become manifested in actions that merely symbolise them. An illicit sexual desire, for example, might be manifested in some other illicit action that the person does not really desire in itself, such as petty theft of items the person does not even want (see e.g. Freud 1915, §§ IV and VII). This is why repression can lead to behaviour whose root causes cannot be uncovered easily without the aid of psychoanalysis.

The section of *Being and Nothingness* entitled 'Bad Faith and Lies' (1943, pt. I ch. 2 § I), is devoted to arguing that this Freudian distinction between two kinds of mental item does not have the resources to explain self-deception. For the activity of repression itself would seem to involve recognition of the drive as one that is to be repressed and the recognition of actions aimed at satisfying it, while the activity of resistance involves comparison of the psychoanalyst's suggestion about the repressed drive with the repressed drive itself. These are rational activities. Repression and resistance, therefore, cannot be the product of the interaction of non-rational drives. They consist in mental processing of the first kind, involving items that are sensitive to evidence, inferentially and evidentially interrelated, and available for beliefs to be formed about them and communicated. In which case, the activities of repression and resistance would be as consciously available as any other aspects of the mind. The problem is that self-deception seems to require that one hide the self-deceptive activity from oneself. So for repression and resistance to be forms of self-deception, one would have to be able to deceive

oneself about one's engagement in these activities. Nothing in Freud's account explains this possibility: Freud's account characterises the things about which we are self-deceived as beyond the reach of rational inference, but the activities of repression and resistance cannot themselves be beyond the reach of rational inference because they involve such inference. Moreover, it is difficult to see how Freud's theory could be augmented to accommodate this. It would not be enough to argue that repression and resistance can occur in the absence of some further condition required for conscious accessibility: one would have to argue that one can actively maintain the absence of that condition while preventing oneself from being aware of doing so, which just raises the problem of self-deception in another place.

Freud's departure from the Cartesian understanding of the mind consists in his denial that it consists solely in consciously accessible, informationally sensitive, inferentially available, linguistically articulable items. The drives and impulses that he discusses are genuinely parts of the mind, rather than merely aspects of the body, because they carry information that influences the operation of the rest of the mind (1915, § I; 1923, ch. I). Yet these operate, on Freud's theory, according to mechanical laws of causality rather than rational laws of thought. Freud has abandoned the Cartesian idea that mind has informational content which operates according to rational inference whereas everything else lacks such content and operates mechanically. But he has not done so by opposing the dualism of reason and non-rational causality. Neither has he opposed the dualism of informational and non-informational. He has simply denied that these two dualisms coincide. Cast in this light, it is clear just how much of the Cartesian picture is actually retained by Freud. It is also clear just where Sartre's criticism bites: since self-deception is an intelligent activity, it cannot be part of the nonrational aspect of the mind; since it is not immediately available for inference and articulation, it cannot be part of the rational aspect either.

Sartre's departure from Freud is to deny that behaviour can be traced back to drives and impulses that cannot be analysed, cannot be explained in terms of the individual's own choices or preferences, and are not responsive to the consciously accessible aspect of mind. These are 'pseudo-irreducible' items,

claims Sartre; it is simply arbitrary to accept them as basic (1943, 656, 659-60; tr. 589, 592-3). In denying that there are any such motivations, Sartre denies the dualism of reason and mechanism that Freud's account rests on. But he does not deny this dualism simply by excluding one side of it. He does not, that is to say, agree with Descartes that the mind should be identified with what Freud considered to be its inferential aspect. Neither does he attempt to show that the mind is entirely mechanical. His strategy is to present a new account of the mind, one that does not draw this dichotomy in the first place.

Sartre's theory of mind gives a central role to the idea of a project. Commentators on Sartre are often mistaken about this notion. A project need not be aimed at bringing about something that is not yet the case; staying alive can count as a project. Neither need a project be concerned directly with the sort of person one wants to become, though of course a corollary of any project one is that one becomes the person who pursued that project. Writing a book is a project whose direct aim is the writing of the book, although this brings with it that one becomes the person who wrote that book. One can engage in intelligent behaviour guided by a project, moreover, without any explicit awareness of doing so, as when one walks a familiar route while thinking about some difficult problem one needs to solve. Finally, a project can be pursued as a constitutive part of the pursuit of another project. Walking that familiar route might be the pursuit of the project of getting to the office where one works, which itself is a way of pursuing the project of keeping one's job, which is a way of ensuring that one stays alive and healthy (1943, 512; tr. 459). Central to the Sartrean idea of a project, moreover, is the idea of habit. Having actively pursued the project of staying alive for some time, many of the things one does to stay alive and healthy will have become so habitual that one might never really consider why one does them or whether it is worth continuing. One just feels the need to eat, the need to sleep, the need to get to work on time, and responds accordingly. But this does not mean that one's behaviour is not goal-directed or that one does not endorse the goal to which it is directed. It is just that one no longer thinks about it.

That one can pursue a project without thinking about it is only part of the Sartrean account of self-deception about one's motivations. For self-deception, as we have seen, requires the stronger point that one can hide this motivation from oneself in such a way that it becomes at least very difficult for one to see that it is one's motivation. This aspect of Sartre's theory of self-deception rests on his account of the way in which projects influence behaviour. Their role is to structure experience. The world as one experiences it, according to Sartre, is a world of invitations, demands, prohibitions, tools, and obstacles whose practical status is dependent on the projects that one pursues. This is what Sartre means by the term 'situation': one's experience reflects one's own goals and values (1943, 76-7; tr. 62-3). In responding to things as they seem, one serves one's own priorities. When one feels the need to eat, the need to sleep, the need to get to work on time, and responds accordingly, one is pursuing one's project of staying alive. It is not just one's awareness of one's surroundings and bodily states that is structured in this way, according to Sartre; all of one's experience is so structured, even one's consideration of one's own thoughts and motivations.

It is this aspect of the idea of a project that allows one to pursue projects that manage to conceal themselves. Sartre's central example in *Being and Nothingness* is the inferiority project. The goal of this is to present oneself to oneself and others as intrinsically inferior to other people. Sartre claims that pursuing this project involves adopting further projects that one cannot hope to achieve, or that one pursues in hopeless ways, so that when one fails to achieve these goals one can blame one's natural inferiority (1943, 550; tr. 493). This requires one to consider oneself to be genuinely attempting to achieve those other goals, since otherwise the failure to achieve them will not feel like genuine failure. The inferiority project must structure one's experience such that certain goals seem achievable when they are not, or seem achievable by means that will not in fact bring them about. It must also structure one's experience such that one seems to oneself to be genuinely pursuing those goals. One would then be unlikely to recognise that one is pursuing the project of proving oneself inferior; one would seem to

oneself to be genuinely pursuing other projects and failing at them. It would take a detached analysis of one's overall pattern of behaviour to identify the problem.

One central disagreement between Freud and Sartre, therefore, concerns the relation between the motivations that are hidden and the behaviour that they motivate. Freud sees this relation as one of symbolisation. Sartre, on the other hand, sees the behaviour as genuinely a way of pursuing the hidden goal (see Sartre 1939, § II). Whereas petty theft might symbolise some illicit sexual act, as Freud claims, it is not a way of pursuing that sexual goal. Trying to become a concert pianist without formal piano tuition, however, is a way of pursuing the goal of failing in one's projects and so showing oneself to be inferior to other people. This illustrates the deeper difference between the two accounts. On the Freudian view, self-deception about some motivation is an activity that is only contingently connected to that motivation. It is because one disapproves, usually because society disapproves, of that motivation that one hides it from oneself. On the Sartrean picture, however, it is in the very nature of the motivation itself that it must conceal itself if its goal is to be achieved. One will not succeed in persuading oneself of one's inferiority if one is clearly aware that this is one's goal.

Sartre thinks that the capacity for self-deception about one's motivations shows us something fundamental about the mind. It shows us that conscious experience, and consequent practical reasoning, is already structured by the goals that we are pursuing. This is how it is possible for those very goals to be beyond easy inferential reach. Sartre's view of the mind is a bold rejection of the more traditional view that practical reason involves rationally weighing considerations in order to formulate one's goals. One's prior goals, according to Sartre, determine which considerations will occur to one and how they will be weighted. Although challenging, this view of the mind is not entirely alien to current scientific psychology (e.g. Mischel and Shoda 1995). It remains an open question whether Sartre's view, or something like it, is the correct way to understand the mind. If it is, then it does seem to offer an account of how one can purposively hide one's own motivations from oneself. We have

seen that self-deception about belief seems to require self-deception about one's motivations. If the Sartrean view of the latter form of self-deception is correct, then perhaps the project-based view of the mind that requires it can also explain the capacity for self-deceptive belief. But that, too, remains an open question.

It is also a further question what the ethical implications are of our capacity for self-deception. Psychology alone cannot formulate ethical precepts. But it does constrain ethics in two ways. It constrains the moral psychology involved in normative ethical theories. Much of modern ethics, particularly utilitarian and Kantian accounts of morality, seems to assume that actions are generated by consciously accessible and rationally structured deliberation. If such an assumption cannot be incorporated into any theory of motivation capable of explaining our capacity for self-deception, then it seems that these theories will need to be revised or perhaps entirely discarded. Psychology also constrains the practical outcomes of moral theorising. There is no point in simply exhorting people to act on the right kinds of motivations if their own motivations are often opaque to them. Similarly, there is no point in critical self-reflection if one is likely to distort the evidence to fit one's favoured self-image. The ethical task of becoming a better person is partly a task of self-knowledge (see Cottingham 2010). If ethics is to be concerned with motivations at all, therefore, practical ethical prescriptions ought to include procedures for uncovering and fairly assessing those motivations as well as methods for changing them. Precisely how our capacity for self-deception constrains theories moral psychology and practical ethics, of course, depends on how that capacity is best understood.

Works Cited

- Cottingham, John. 2010. Integrity and Fragmentation. *Journal of Applied Philosophy* 27 (1): 2-14.
- Gardner, Sebastian. 2000. Psychoanalysis and the Personal / Subpersonal Distinction. *Philosophical Explorations* 3 (1): 96-119.
- Freud, Sigmund. 1915. Das Unbewusste. *Internationale Zeitschrift für Psychoanalyse* 3 (4): 189-203 and 3 (5): 257-269. Translated by James Strachey, Anna Freud, Alix Strachey, and Alan Tyson as 'The Unconscious'. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, edited by James Strachey and Anna Freud, volume 14. London: The Hogarth Press and the Institute of Psycho-Analysis, 1957.
- Freud, Sigmund. 1923. *Das Ich Und Das Es*. Leipzig: Internationaler Psycho-analytischer Verlag. Translated by James Strachey, Anna Freud, Alix Strachey, and Alan Tyson as *The Ego and The Id*. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, edited by James Strachey and Anna Freud, volume 19. London: The Hogarth Press and the Institute of Psycho-Analysis, 1961.
- Mischel, Walter and Yuichi Shoda. 1995. A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure. *Psychological Review* 102 (2): 246-268.
- Neu, Jerome. 1988. Divided Minds: Sartre's "Bad Faith" Critique of Freud. *The Review of Metaphysics* 42 (1): 79-101.
- Sartre, Jean-Paul. 1939. *Esquisse d'une Théorie des Emotions*. Paris: Hermann. Translated by Philip Mairet as *Sketch for a Theory of the Emotions*. Second edition. London: Routledge, 2002.
- Sartre, Jean-Paul. 1943. *L'Être et le néant: essai d'ontologie phénoménologique*. Paris: Gallimard. Translated by Hazel E. Barnes as *Being and Nothingness: An Essay in Phenomenological Ontology*. Second edition, edited by Arlette Elkaïm-Sartre. London: Routledge, 2003.
- Soll, Ivan. 1981. Sartre's Rejection of the Freudian Unconscious. In *The Philosophy of Jean-Paul Sartre*, edited by Paul Arthur Schilpp (La Salle, IL: Open Court, The Library of Living Philosophers).

Suggested Readings

Brown, Lee and Alan Hausman 1981. Mechanism, Intentionality, and the Unconscious: A Comparison of Sartre and Freud. In *The Philosophy of Jean-Paul Sartre*, ed. Paul Arthur Schilpp (La Salle, IL: Open Court, The Library of Living Philosophers).

Catalano, Joseph. 2010. *Reading Sartre*. Cambridge University Press.

Gardner, Sebastian. 1993. *Irrationality and the Philosophy of Psychoanalysis*. Cambridge University Press.

Webber, Jonathan. 2002. Motivated Aversion: Non-Thetic Awareness in Bad Faith. *Sartre Studies International* 8 (1): 45-57.

Webber, Jonathan. 2009. *The Existentialism of Jean-Paul Sartre*. London: Routledge.

Wollheim, Richard. *Freud*. London: Fontana, 1971.