

Research Article

A Dispositional Account of Self-Deception: A Critical Analysis of Sartre's Theory of Bad Faith

Guy Du Plessis¹

1. Utah State University, United States

This essay addresses the notion of self-deception as articulated by Sigmund Freud and Jean-Paul Sartre. More specifically, I will critically assess Sartre's notion of 'bad faith' (*mauvaise foi*) as a critique of Freud's depth-psychological account of self-deception. Sartre's main objection to Freud's account of self-deception rests on his argument that for self-deception to occur there needs to be a conscious awareness of the coexistence of mutually incompatible beliefs, and that Freud had obscured this fact by splitting the self and with a mixture of jargon. In conclusion, I suggest a speculative and tentative hypothesis for self-deception that suggests an alternative to Freud and Sartre's account of self-deception, one that views the self or self-system as having a disposition for self-deception when understood as having the capacity for fluid working self-states that are 'expandable' and 'contractable' that differ in self-complexity.

Philosophers are seldom famous in their own lifetime. Jean-Paul Sartre (1905–80) was an exception. In short, in the Liberation period after World War Two, he was a celebrity and public intellectual par excellence. For example, thousands attended his public lecture *Existentialism is a Humanism*, towards the end of 1945. In this lecture he gave the eager public a more accessible version of his Magnum Opus, *Being and Nothingness* (1943). In this essay, I will explore Sartre's notion of *mauvaise foi*, translated as bad faith, as articulated in *Being and Nothingness*.

Sartre's account of bad faith juxtaposes a critique of Sigmund Freud's depth-psychological account of self-deception – in the words of existential psychiatrist Medard Boss, he attempts to overthrow the “secret prejudices” of the Freudian *Weltanschauung*. Sartre's main critique of Freud's depth-psychological approach to self-deception can be summed up as: Freud is charged with splitting the

subject into meta-psychological parts (Conscious, Preconscious, Unconscious, or Id, Ego, Superego) and provides a misguided mechanistic explanation of how there can be a “liar” and a “lied to” duality within a single consciousness, which simply transfers the problem where it remains unsolved, thus consisting in a pseudo-explanation (which today might be called a ‘homuncular fallacy’). Sartre’s articulation of bad faith can be understood to be a phenomenon distinctive of the *for-itself*.

Sartre’s (1943, 570) theory of mind and his own methodology, existential psychoanalysis, attempts to remain faithful to Edmund Husserl’s phenomenology, simply put, that all psychic acts are “coextensive with consciousness.” Yet, Sartre, rejection of Freud’s explanation is more influenced by his dismissal of psychic determinism than it being an inheritance of Brentano’s descriptive psychology (as was the case for Edmund Husserl).

In this essay, I will critically assess Sartre’s notion of bad faith, by contextualizing it with Freud’s articulation of self-deception. In conclusion, I provide a dispositional account for self-deception that is not beholden to Freud nor Sartre’s theory of mind, but views the self as having a disposition for self-deception when understood as having the capacity for fluid working self-states that are “expandable” and “contractable” that differ in self-complexity.

Freud’s Depth-Psychological Account of Self-Deception

In philosophical literature, and according to the traditional model, self-deception requires an individual to hold contradictory beliefs, and the individual must intentionally believe something which s/he knows to be false. The traditional model has raised two paradoxes, the so-called static paradox (how can an individual hold contradictory beliefs simultaneously?) and the dynamic or strategic paradox (how can an individual deceive herself without her intentions being rendered ineffective?) (Mele, 1987, 2001).

Yet, for those that ascribe to a depth-psychological perspective, these paradoxes are not problematic. According to depth-psychological perspectives, individuals can hide their own motivations from themselves (Lockie, 2003). Central to Sigmund Freud’s (1915/1957, 1923/1961) theory of mind, and his therapeutic methodology of psycho-analysis, is the premise that we can be motivated by unconscious drives or impulses of which we are not aware – which push for satisfaction, even at the expense of our conscious beliefs and wishes. The aim of psychoanalysis is to bring awareness to these “hidden” drives.² According to Freud (1915), these concealed motivations are not merely ‘descriptively

unconscious,' instead they are 'dynamically unconscious,' which highlights that the individual is actively (through the use of defence mechanisms) trying to keep motivations out of awareness.

In Freud's (1915, 1923) model of the mind, the self is viewed as comprised of two or more distinct and warring systems, each with its own goals. On the one hand, we have our conscious mind where thoughts, beliefs, desires and aspirations are accessible and can be conceptually communicated. On the other hand, there is the unconscious mind comprised of drives and impulses which can compete with each other according to the laws of their 'cathectic energy.' The unconscious contains socially unacceptable ideas, desires, memories and motives, that are associated with conflict, emotional pain and anxiety. The unconscious is not accessible to our awareness and is composed of non-conceptual and symbolic elements that cannot be communicated through language.

It must be noted that this notion predates Freud, for instance, *The Philosophy of the Unconscious* by Von Hartmann was published in 1869. And prior to that, Arthur Schopenhauer (1819/1969), in *The World as Will and Representation*, argued for a conceptual and aconceptual divide of the mind. He was critical of Immanuel Kant's (1781/1998) theory of mind which claim that all cognition is conceptual. He argued that Kant did not separate perceptual knowledge from abstract knowledge.

Schopenhauer (1819-44/1969) asserted:

The intellect remains so much excluded from the real resolutions and secret decisions of its own will that sometimes it can only get to know them, like those of a stranger, by spying out and taking unawares; and it must surprise the will in the act of expressing itself, in order merely to discover its real intentions. (II, 209)

In another passage, Schopenhauer's (1819-44/1969, II, 135) vivid imagery resonates with Freud's famous metaphor of the mind being akin to an iceberg.

Let us compare our consciousness to a sheet of water of some depth. Then the distinctly conscious ideas are merely the surface; on the other hand, the mass of the water is the indistinct, the feelings, the after-sensation of perceptions and intuitions and what is experienced in general, mingled with the disposition of our own will that is the kernel of our inner nature. (II, 135)

Similarly, Nietzsche (1881/1982, 76) proposed: "All our so-called consciousness is a more or less fantastic commentary on an unknown, perhaps unknowable, but felt text." And again in the *Gay*

Science (1882/1974, 298-299):

Man, like every living being, thinks continually without knowing it; the thinking that rises to consciousness is only the smallest part of all this – the most superficial and worst part – for only this conscious thinking takes the form of words, which is to say signs of communication, and this fact uncovers the origin of consciousness.

Freud's presentation of the tripartite structure of the mind (id, the ego, and the superego) in *The Ego and the Id* (1923), and his psychic determinism and construct of the id (*das Es*), is echoed in Schopenhauer's concept of the will, "the inner- most essence, the kernel of every particular thing and also of the whole" (1819-44/1969, I, 110), but devoid of Schopenhauer's metaphysical cosmic dimension.

Nietzsche's (1872/1993) dichotomy of the Dionysian and the Apollonian in his *Birth of Tragedy* (deeply influenced by Schopenhauer) share similarities to Freud's id and ego. For example, Nietzsche used the German pronoun *das Es* (translated as the id) to denote the unconscious/instinctual, the personal pronoun *das Ich* (translated as ego) to represent the conscious, part of the mind, and *das Selbst* (the self) to signify the personality.

Freud's method of psychoanalysis is based on the premise that the schism between our conscious and unconscious minds can become problematic when we do not accept a drive and do not engage in behaviour that would satisfy it. For Freud (1915) repression is a defence mechanism we employ to prevent a drive to factor in our behaviour, and he views repression as a form of self-deception. Instead of acknowledging a drive, repression involves the denial that one possesses the drive that is actively being repressed. According to Freud (1915), when a drive has been repressed it exists as 'dynamically unconscious.' Even though a drive has been repressed it can still motivate behaviour by pushing for satisfaction. Because a person will not engage in behaviour to satisfy a drive, according to Freud, a drive becomes manifested in actions that merely symbolize it. For example, an illicit sexual desire can result in a person engaging in shoplifting items that they do not need. The aim of psychoanalysis is to find the root causes of neurotic behaviour that are inexplicable unless we discover the drive that motivates the behaviour.

In opposition to Freud's view of self-deception (and his antecedents, Schopenhauer and Nietzsche), Sartre (1958) proposes that his theory of bad faith has superior explanatory power to account for self-deception, in contrast to one that proposes a dynamic unconscious as being central to an account of

self-deception. In the next section of the essay, I briefly discuss Sartre's theory of mind which lays the foundation for his account of bad faith.

Sartre's Theory of Mind

Sartre's main departure from Freud is that he denies the dualism of reason and mechanism that informs Freud's model. Sartre proposes a model of the mind that is not based on this dualism and psychic determinism. It must be noted that Sartre is in agreement with many of Freud's foundational premises. He agrees that psychic life remains inevitably "opaque" and often impenetrable to us, as well as that the philosophical understanding of human reality requires a method for investigating the meaning of psychic facts. Sartre (1958, 354) argues:

I do not cease to hear sounds; they are simply lost in the undifferentiated totality which serves as the background for my reading. Correlatively my body does not cease to be indicated by the world as the total point of view on mundane totality, but it is the world as ground which indicates it.

But Sartre disagrees that the natural scientific method and its notion of causality are helpful in this regard. According to Sartre, the *for-itself* remains free regardless of external and social constraints. He does not adhere to a strict psychic determinism but does not deny the influence of the past – but he contends that past events are determined in relation to one's present choice, and understood as the consequence of the power invested in this free choice. As he (1958, 503) puts it:

Since the force of compulsion in my past is borrowed from my free, reflecting choice and from the very power which this choice has given itself, it is impossible to determine a priori the compelling power of a past.

Conversely, determinist explanations that construe one's present as a mere consequence of the past proceed from a kind of self-delusion that operates by concealing one's free project, and thus contributes to the obliteration of responsibility. Sartre hence seeks to redefine the scope of psychoanalysis: rather than a proper explanation of human behaviour that relies on the identification of the laws of its causation, psychoanalysis consists in understanding the meaning of our conducts in light of one's project of existence and free choice. One might wonder, then, why we need any such psychoanalysis, if the existential project that constitutes its object is freely chosen by the subject. Sartre (1958, 570) addresses this objection in *Being and Nothingness*, claiming that

if the fundamental project is fully experienced by the subject and hence wholly conscious, that certainly does not mean that it must by the same token be known by him; quite the contrary.

Discussing self-deception or bad faith Sartre (1958, 49) proposes that

the one to whom the lie is told and the one who lies are one and the same person, which means that I must know in my capacity as deceiver the truth which is hidden from me in my capacity as the one deceived. Better yet I must know the truth very exactly in order to conceal it more carefully - and this not at two different moments...but in the unitary structure of a single project.

Sartre's theory of mind, as indicated above, puts an emphasis on the notion of the "project." A project is something a person is committed to and remains committed to. A project can also be habitual, and one can engage in a project without much thought, yet it remains goal-directed. The idea that one can pursue a project without thinking is one aspect of Sartre's theory of mind that explains motivation in relation to self-deception. According to Sartre, our experience is structured by the pursuit of projects. It is this characteristic of a project that allows us to pursue projects that conceal themselves from us. The type of project that is central in *Being and Nothingness* is what he calls the "inferiority project."

According to Sartre, pursuing an inferior project is pursuing something that is either beyond our abilities, or the manner we go about it is doomed to failure. When we then fail at these projects, we blame our natural inferiority. The inferiority project requires that we genuinely believe we are pursuing these goals, otherwise they would not feel like authentic failures. Webber (2013, 10) states that

"[t]he 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 recognize 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. (Webber, 2013, 10)

The preceding discussion highlights one of the central disagreements between Freud and Sartre, regarding the nature of the relationship between motivations that are concealed and the behaviour

that they influence. For Freud, self-deception is only contingently related to motivation – it is because one disapproves of one’s motivation that one conceals it from oneself. For Sartre, on the other hand, “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” (Webber, 2013, 11). Sartre (1958) argues that our capacity for self-deception is a failing that reveals something about the nature of our minds and human nature in general. It reveals that conscious experience and practical reasoning are given structure by our projects, and our prior projects influence our practical reasoning.

In the next section, I will briefly discuss Sartre’s account of bad faith as an alternate explanation and critique of Freud’s account of self-deception.

Bad Faith

In Sartre’s book, *Being and Nothingness* (1958), a section entitled ‘Bad Faith and Lies’ he argues that Freud does not provide a sufficient justification for self-deception by making a distinction between conscious and unconscious mental processes. Sartre claims that in the act of repression, there is awareness of the drive that is being repressed as well as an awareness of the actions that aim to satisfy it – and simply put, these are both rational activities. According to Sartre the Freudian ‘censor’ must first register the drive or impulse before preventing it from becoming conscious. Sartre (1958, 52-53) argues that

[i]t 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? ... [a]ll knowing is consciousness of knowing.

Consequently, for Sartre, the act of repression is not due to the dynamics of nonrational drives. They are available for conceptual understanding and can be communicated. The problem is that for it to count as self-deception one needs to hide the self-deceptive activity from oneself.

Freud’s departure from the Cartesian understanding of the mind is that he does not view the mind as only consisting of consciously accessible and linguistically available items. The drives he proposes are part of the mind, but these operate according to mechanical laws rather than laws of rational thought.

Freud rejected the Cartesian and Kantian idea that the mind has only informational content that operates according to rational inference (Webber, 2013). Freud's (1915) conception of the mind draws a clear dualism between the conceptual and non-conceptual parts of the mind. Jonathan Webber (2013, 8) states that according to the Freudian model of the mind,

the dualism of the conscious and unconscious does not coincide with the dualism of the rational and mechanical: all rationally structured items in the mind are consciously accessible; some purely mechanical items are consciously accessible, but others are dynamically unconscious. 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 purely mechanical aspect of the mind; since it is not immediately available for inference and articulation, it cannot be part of the rational aspect either.

Sartre (1943, 86) writes in his chapter on "Bad Faith" in *Being and Nothingness* that a human being "can take negative attitudes towards himself." He states that "self-negation" is one of these negative attitudes, and he chooses for his phenomenological analysis a form of self-negation that is "essential to human reality" namely bad faith or "*mauvaise foi*." Sartre (1958, 87) distinguishes bad faith from lying as "a lie to oneself" only if "we distinguish the lie to oneself from lying in general."

Sartre (1958, 55) argues for a phenomenological account of self-deception, for this reason, he proposes that '[i]f we wish to get out of this difficulty, we should examine more closely the patterns of bad faith and attempt a description of them.' In *Being and Nothingness* Sartre (Ibid) provides such a phenomenological account of self-deception or bad faith in an example of a woman "who has consented to go out with a particular man for the first time." This is an example of self-deception because the woman is aware of the man's sexual interest in her, and the potential consequences of that, but because of her ambivalence, she pretends that nothing is being asked of her. How does she accomplish this? She does this by restricting her "behaviour to what is in the present [immanent]; she does not wish to read in the phrases which he addresses to her anything other [transcendent] than their explicit meaning" (Ibid). Although she tries to disarm the 'transcendent aspect,' that is which is beyond the immanent, she does not want to deny it as "she would find no charm in a respect which would be only respect" (Ibid). Thus, she behaves in a way that can maintain these contradictory wishes. As Sartre explains, "[t]his time then she refuses to apprehend the desire for what it is; she does not even give it a name; she recognises it only to the extent that it transcends itself toward

admiration, esteem, respect” (Ibid). Sartre highlights, using his terminology, that she strips that man’s conduct of all “transcendence,” as well as also stripping his desire of all immanence. She thus rejects the implied sexual implications (transcendence) and sees the man’s behaviour as only a lofty “concern” (imminence) for her. Therefore, she can enjoy the excitement of the moment while denying the sexual implications and the potential choices it may have. According to Sartre “this woman is in bad faith” (Ibid).

Why is this woman in bad faith? For Sartre, she is in bad faith or self-deceived because she conceals something from herself at the same instant, she brings it to conscious awareness (Webber, 2013).

Critique of Bad Faith

I will now provide a brief critique of Sartre’s notion of bad faith and argue that modern advances in research and academic psychology are more congruent with Freud’s view of how hidden motivations can influence our behaviour, than with Sartre’s notion of the radical freedom of the *for-itself*.

There is an overwhelming amount of data that supports the notion that thoughts, feelings, and motives can be part of unconscious processes (Westen, 1998b). Sartre’s theory of bad faith is based on the proposition that we need conscious access to our motivations and thoughts at the same time as we are denying them. Only by consciously holding two contradictory thoughts are we capable of self-deception. I am skeptical of this view, as I will highlight below that much of our thoughts and motivations can occur outside of awareness, so we can hold two conflicting views and motivations at the same time - one in awareness and the other outside of awareness.

Partly due to advances in functional imaging, we now have access to the neurological bases of instinctual drives and emotions, and evidence for their role in mental processes (Etkin et al., 2004) and findings support Freud’s view that mental activity is influenced by phylogenetically old emotion and motivation systems (Panksepp, 1998). Research on unconscious affect provides strong support for Freud’s central thesis, which has guided psychoanalytic practice for the last century, namely that people can think things of which they are not aware and act on feelings that they are not aware of (e.g., see Westen, 1998a, 1998b). Thus, we can have motivations and thoughts outside of awareness that are contradictory to those held in awareness at the same time.

Sartre’s theory of mind is more congruent with theories of rational agency that view humans as agents with practical reasoning systems, using logic to decide which actions to perform, guided by their beliefs and worldviews (Wooldridge, 2000). From this perspective, humans are understood as

“intentional systems” (Dennett, 1987). But as highlighted above there is overwhelming evidence that our mental processes and behaviour are also significantly influenced by non-logical and emotional processes. For example, Antonio Damasio (1994) argues that rationality cannot be separated from emotions, which are “an integral component of the machinery of reason” (p. xii). Emotions can negatively affect our rational thinking, but their absence can be equally adverse. Damasio (2003, 6) highlights the centrality of emotions in our being-in-the-world, saying that “feelings are the expression of human flourishing or human distress, as they occur in mind and body.” He proposes that rational deliberation activates “gut feelings” that guide us in the process of reflection. The somatic marker “forces our attention on the negative outcome to which a given action may lead, and functions as an automated alarm which says: Beware of danger ahead if you choose the option which leads to this outcome” (Damasio, 1994, 173). So, feelings may have their basis in body representations, but we do not have conscious access to the neuronal processes that underlie bodily homeostasis and emotion states (Craig, 2002, 2009).

In conclusion, I do not believe Sartre provides a convincing critique of Freud’s account of self-deception, and his theory of mind is not congruent with recent advances in psychological research and academic psychology, nor with clinical practice – as most psychotherapists would agree, except those that adhere to a ridged cognitive and behavioral view, that it requires more than mere rational deliberation to change deeply entranced behaviour.

A Dispositional Account of Self-Deception

I will now provide a brief, speculative and tentative hypothesis for self-deception that suggests an alternative to Freud and Sartre’s account of self-deception. Instead, it proposes that the self or self-system has a *disposition*³ for self-deception, depending on context.⁴ In this approach, the self or self-system is understood as having the capacity for fluid working self-states that differ in self-complexity. Simply put, an individual is more prone to self-deception when in a contracted and limited self-state of lower self-complexity – in this self-state an individual has limited access to inner and outer data (and past experiences) and may make discussions that contradict motivations or plans that are reasonable (or has been decided) in more expanded self-states of increased self-complexity. For example, giving into a craving for a piece of cake or cigarette even though one has made a conscious decision to stop. The craving creates a limited and contracted self-state (an innate function or our reward system to satisfy basic human needs), where giving in to the craving seems

logical (and not contradictory) within the context of that self-state, even though it is incompatible or illogical with other more expanded self-states (contradictory with other self-states).

By drawing on self-affirmation theory (Steele, 1988), working self perspectives (Markus & Wurf, 1987), self-complexity theory (Linville, 1987) and expandable selves theory (Walton, Paunesku & Dweck, 2012), my hypothesis is based on the premise that self-deception can be the result of a 'limiting' and 'contracted' self-state and mode of being-in-the-world where the working self has limited self-complexity, and has fewer resources with which to function effectively in challenging situations, which can result in self-deception. Given the size and complexity of their self-concepts, people can think about only limited aspects of their identity at any moment. The aspects of identity about which people are aware at a particular time have been called the phenomenal self, spontaneous self-concept, and working self-concept (Markus & Wurf, 1987).

This hypothesis is consistent with recent theorizing about the representation and functioning of the self. For instance, McConnell's (2011, 3) multiple self-aspects framework (MSF) portrays the self as "a collection of multiple, context-dependent selves." At the heart of these theories is the contextual nature of the contents of the self, the idea that the self is not unitary or constant but changes in reliable ways as situations evoke different context-dependent selves. The type of self-state that an individual embodies has significant influence over their behaviour and how they deal with the situation at hand. These self-states can also differ in complexity. This view is supported by the self-complexity theory (Linville, 1985). Self-complexity theory has emphasized how the structure of a more global and stable self can differ for different people. The complexity of this structure in turn can affect people's ability to cope with threats or failures and can ultimately affect their overall well-being (Linville, 1985).

Combining insights from the working self and self-complexity theory Walton, Paunesku & Dweck (2012) argue for an 'expandable selves theory' and propose that the working self, depending on the situation, can vary in its breadth, complexity, and structure (vary in degree of self-complexity) - that it is expandable and contractable. Different situations may bring to the fore either a wide range of aspects or only isolated aspects of the self. As a result, the working self in a situation may be relatively broad or relatively narrow. Consistent with self-complexity theory, they suggest that, in general, when the working self is narrow it may be more vulnerable and have fewer resources with which to function effectively in challenging situations.

Consequently, I propose that what can account for the phenomenon of self-deception or bad faith, a common attribute of behavioral disorders like substance use disorder, is that an individual may have a belief of p in a certain self-state and a belief of $\sim p$ in another self-state, where each can lead to different discussions or behaviors.

This perspective could possibly explain the common observation and enigma of why individuals suffering from substance use disorder can display various degrees of control over their use of substances - depending on the context. A dispositional perspective highlights that certain context-dependent situations (for example, high-risk situations) can elicit contracted and limited self-states of lower self-complexity, and in these self-states using substances is a rational response, even though it is in contradiction to the intentions held when the individual is in more expanded self-states.

These self-states can 'flip-flop' or alternate rapidly, and what may appear to be an individual holding p and $\sim p$ at the same time, is in fact the result of rapidly alternating self-states, which are not problematic or paradoxical at all. Thus, from this dispositional perspective the so-called static paradox and the dynamic paradox may indeed be false dilemmas - and thus no need to invoke a dynamic unconscious nor acts of bad faith to explain self-deception.

Conclusion

In this essay, I addressed the notion of self-deception as articulated by Sigmund Freud and Jean-Paul Sartre. More specifically, I critically assessed Sartre's theory of bad faith as a critique of Freud's depth-psychological account of self-deception. Sartre's main objection to Freud's account of self-deception rests on his argument that for self-deception to occur there needs to be a conscious awareness of the coexistence of mutually incompatible beliefs, and that Freud had obscured this fact by splitting the self and with a mixture of jargon. I then presented a speculative and tentative hypothesis for self-deception that suggests an alternative to Freud and Sartre's accounts, one that views the self or self-system as having a disposition for self-deception when understood as having the capacity for fluid working self-states that are 'expandable' and 'contractable' that differ in self-complexity.

Footnotes

¹ Contact author at guy.duplessis@usu.edu or guy@guyduplessis.com

² Freud (1915, 1923) provides an account of self-deception in his discussion of illusion, repression, and delusion. He uses the words *Selbstbetrug* and *Selbsttäuschung* interchangeably when referring to self-deception. *Selbstbetrug* is usually translated as ‘self-deceit’, or ‘deception’, while *Selbsttäuschung* is usually translated as ‘self-deceit’ or ‘delusion.’

³ I apply Barry Smith’s definition of disposition. See *Building Ontologies with Basic Formal Ontology* (Arp, Smith & Spear, 2015).

⁴ See *Developing General Models and Theories of Addiction* (West, Christmas, Hastings & Michie, 2019) for a dispositional account of addiction.

References

- Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with Basic Formal Ontology*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/9780262527811.001.0001>.
- Boss, M. (1983). *The existential foundations of medicine and psychology*. Jason Aronson.
- Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews. Neuroscience*, 3 (8), 655–666.
- Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews. Neuroscience*, 10, 59–70.
- Damasio, A. (1994). *Descartes’ error: Emotion, reason, and the human brain*. Quill.
- Damasio, A. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Harcourt.
- Dennett, D. C. (1987). *The intentional stance*. MIT Press.
- Etkin, A., Klemenhagen, K. C., Dudman, J. T., Rogan, M. T., Hen, R., Kandel, E. R., et al. (2004). Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. *Neuron*, 44, 1043–1055.
- Freud, S. (1915/1957). ‘Das Unbewusste’, *Internationale Zeitschrift für Psychoanalyse*, vol. 3, no. 4, pp. 189–203 and vol. 3, no. 5, pp. 257–69. 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, vol. 14, ed. James Strachey and Anna Freud. London: Hogarth Press and the Institute of Psycho-Analysis.

- Freud, S. (1923/1961). Das Ich Und Das Es. Leipzig: Internationaler Psycho-analytischer. 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*, vol. 19, ed. James Strachey and Anna Freud. London: Hogarth Press and the Institute of Psycho-Analysis, 1961.
- Heidegger, M. (1957/1962). *Sein und Zeit*, Tübingen: M. Niemeyer. Translated as *Being and Time*, John Macquarrie and Edward Robinson (trans), SCM Press.
- Kant, Immanuel (1781 and 1787/1998), *Critique of Pure Reason*, ed. Paul Guyer and Allen Wood. Cambridge University Press.
- Linville, P. W. (1985). Self-complexity and Affective extremity: Don't put all your eggs in one cognitive basket. *Social Cognition*, 3, 94–120.
- Lockie, R. (2003). Depth psychology and self-deception. *Philosophical Psychology*, 16(1), 127–148.
- Maldonato, M., & Valerio, P. (2018). Artificial entities or moral agents? How AI is changing human evolution. In *Multidisciplinary approaches to neural computing*. Springer, 379–388.
- Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, 38, 299–337.
- Mele, A. (2001). *Self-deception unmasked*, Princeton: Princeton University Press.
- Mele, A. (2000). Self-deception and emotion, *Consciousness and Emotion*, (1), 115–139.
- McConnell, A. R. (2011). The multiple self-aspects framework: Self-concept representation and its implications. *Personality and Social Psychology Review*, 15, 3–27.
- Nietzsche, F. (1974). *The gay science* (W. Kaufmann, Trans.). New York, NY: Vintage Books. (Original work published 1882)
- Nietzsche, F. (1982). *Daybreak: Thoughts on the prejudices of morality* (R. J. Hollingdale, Trans.). Cambridge, UK: Cambridge University Press. (Original work published 1881)
- Nietzsche, F. (1990). *Beyond good and evil* (R. J. Hollingdale, Trans.). London, UK: Penguin Books. (Original work published 1886)
- Nietzsche, F. (1993). *The birth of tragedy out of the spirit of music* (S. Whiteside, Trans.). London, UK: Penguin Books. (Original work published 1872)
- Nietzsche, F. (1994). *On the genealogy of morality: A polemic* (K. Ansell-Pearson, Ed.; C. Diethe, Trans.). Cambridge, UK: Cambridge University Press. (Original work published 1887)
- Panksepp, J. (1998). *Affective Neuroscience*, Oxford University Press.

- Wells, A. (2005). Detached mindfulness in cognitive therapy: A metacognitive analysis and ten techniques. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, 23, 337–355.
- Westen, D. (1998a). The scientific legacy of Sigmund Freud: Toward a psychodynamically informed psychological science. *Psychological Bulletin*, 124, 333–371.
- Westen, D. (1998b). Implicit cognition, affect, and motivation: The end of a century-long debate. In: *Empirical Studies of Unconscious Processes*, ed. R. Bornstein & J. Masling.
- Wooldridge, M. (2000). *Reasoning about rational agents*. MIT Press.
- Sackeim, H. (1988). Self-deception: a synthesis. In J. Lockhard & D. Paulus (Eds) *Self-deception: An adaptive mechanism?* Englewood Cliffs, Prentice Hall
- Sartre, J-P. (1958, 2018). *L'être et le néant: Essai d'ontologie phénoménologique*, Paris: Gallimard. Translated as *Being and Nothingness: An Essay on Phenomenological Ontology*, Hazel E. Barnes (trans.), Routledge.
- Sartre, J-P. 1946a [2007], *L'existentialisme est un humanisme*, (Collection Pensées), Paris: Nagel. Translated as *Existentialism is a Humanism*, John Kulka (ed.), Carol Macomber (trans.), Yale University Press, 2007.
- Schopenhauer, A. (1969). *The world as will and representation* (2 Vols.) (E. F. J. Payne, Trans.). New York, NY: Dover Publications. (Original volumes published 1819 and 1844 respectively)
- Schopenhauer, A. (1974). *Parerga and paralipomena: Short philosophical essays* (2 Vols.) (E. F. J. Payne, Trans.). Oxford, UK: The Clarendon Press. (Original work published 1851)
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). Academic Press.
- Von Hartmann, E. (1931). *Philosophy of the unconscious* (W. C. Coupland, Trans.). London, UK: Kegan Paul, Trench, Trubner & Co. (Original work published 1869)
- West, R., Christmas, S., Hastings, J., & Michie, S. (2019). *Developing general models and theories of addiction*. In *The Routledge handbook of the philosophy and science of addiction*. Eds., Hanna Pickard and Serge H. Ahmed, 160–172. New York, NY: Routledge

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.