Received: 7 October 2024 | Revised: 30 December 2025 | Accepted: 23 February 2026

DOI: 10.1111/ejop.70081

ORIGINAL ARTICLE

European Journal of Philosophy WILEY

# 'A Perpetually Disintegrating Synthesis': Sartre on Bad Faith, Good Faith, and the Projects of Selfhood

**Mark A. Wrathall** | **Wanda von Knobelsdorff**

University of Oxford, Oxford, UK

**Correspondence**
Mark A. Wrathall, University of Oxford, Oxford, UK.
Email: mark.wrathall@philosophy.ox.ac.uk

**Funding information**
Leverhulme Trust, Grant/Award Number: MRF-2022-032

## Abstract

An oft-overlooked aspect of Sartre's concept of selfhood is his rejection of good faith and sincerity as normative ideals. We argue that Sartre's paradoxical treatment of good faith – claiming both that it is a manifestation of bad faith and the antithesis of it – holds a key to understanding Sartre's account of selfhood. Contrary to other critics who have discussed good faith, we contend that Sartre sees no normative distinction between bad faith and good faith, and that he is right to do so. We begin with an analysis of Sartre's account of the reflexive structure of selfhood and the negation involved in self-consciousness. We then provide an in-depth examination of the 'circuit of ipseity' – a concept that has been largely overlooked, but which is crucial to understanding reflexivity as 'an immediate and non-cognitive relationship of self to self.' This underwrites our account of the way that both bad faith and good faith are pathological forms of selfhood. We conclude with a discussion of Sartre's claim that selfhood necessarily involves a 'disintegrated structure'. Sincerity and good faith are bad-faith efforts to overcome disintegration, while authentic selfhood involves assuming the disintegration and acting with an understanding that these contradictions are an integral part of being human.

## 1 | PREFACE

Sartre distinguishes between the Self understood as the 'Ego,' and selfhood or 'ipseity.' The Ego – the self-conception that Sartre rejects – is an *object* for consciousness. When we reflect on ourselves, the Ego appears as 'a transcendent object realizing the permanent synthesis of the psychical' (Sartre 1966: 54–5/28).[1] But this Ego, Sartre insists, 'is only the sign of a personality;' it is not what 'confers personal existence on a being' (Sartre 1976: 140/160). Ipseity or authentic selfhood, by contrast, is a *project* we undertake of relating ourselves to ourselves: 'I become conscious (of) myself as one of my free possibilities and I project myself toward myself in order to actualize this ipseity' (Sartre 1976: 327/390).

One perplexing, and often undertheorized, feature of Sartre's account of selfhood is his rejection of good faith or sincerity as an ideal of selfhood. In this paper, we argue that Sartre's rejection of good faith and sincerity as normative ideals provides an important key to understanding his account of selfhood. For Sartre, I am not authentically being a self when I try simply to be who I am, when I try to promise only what I intend, when I try to say only what I really think without dissimulation, or when I try to be true to my feelings. Rather to have authentic selfhood, to be a self in the preeminent sense, is constantly to be grappling with an ambiguous world where my actions lack determinate meaning, and where, consequently, good faith and sincerity are mere pretense.

## 2 | A PUZZLE ABOUT BAD FAITH

One of the marks of selfhood is a certain capacity for reflexivity. This could take the form of the ability to recognize the specific relevance that events have to oneself as opposed to others. It might involve an understanding of oneself as distinct from others. Perhaps reflexivity is expressed when one takes a stand on who one is, or re-invents oneself. It might be found in conscious acts in which one makes oneself the object of one's thoughts, or in which one monitors one's own actions.

Sartre insists, however, that at its most fundamental level, the reflexivity that makes us into selves takes the form of a minimal presence to oneself that he refers to as the 'for-itself'. This minimally reflexive state of self-consciousness, Sartre argues, is 'an immediate and non-cognitive relationship of self to self' (Sartre 1976: 19/11). The basic form of self-consciousness is non-cognitive because it involves no thoughts directed at the self (although it will, of course, involve thoughts directed at other things). We are self-conscious, in other words, even when the self is not an object of thought. Sartre illustrates this with the example of being absorbed in the act of counting cigarettes in a cigarette case. '[M]y impression,' Sartre observes, 'is that of disclosing an objective property of this group of cigarettes: *they are twelve*' (Sartre 1976: 19/11). 'And yet, at the moment when these cigarettes disclose themselves to me as "twelve,"' Sartre continues, 'I am non-thetically conscious of my adding activity' (Sartre 1976: 19/11). In describing himself as *non-thetically* conscious, Sartre means that the activity of counting is not an object of consciousness – as long as he is focussed non-reflectively on the cigarettes, the activity of counting is not placed before his consciousness as its intentional object. And yet, he has some sort of consciousness of himself: 'Indeed, if I am questioned, if someone asks me "What are you doing?" I will reply immediately "I am counting."' Sartre takes this as indicating that there is a non-reflective self-consciousness that 'makes reflexion possible' (Sartre 1976: 19/11–12).

In order to explore the structure of this minimal reflexivity, Sartre takes up the phenomenon of bad faith (*mauvaise foi*) in Part I of *Being and Nothingness* – perhaps the most famous section of the book. In bad faith or self-deception, unlike in a straightforward lie, I relate to myself without knowing how I am relating to myself: I conceal the truth from myself without knowing that I am concealing anything. Self-deception thus offers a case study in non-cognitive reflexivity. By understanding bad faith, I can come to understand how it is possible to be a self (and thus, to relate to myself) without having a reflective awareness of myself.

Now, bad faith is not simply bad in the way any lie is bad – that is, because it falls short of the truth and deceives someone. Sartre introduces bad faith precisely in terms of a pathology of selfhood. Bad faith is 'a negation of self'

(*négation de soi*) (Sartre 1976: 82/88). When I act in bad faith I undermine my own status as a self because bad faith renders me unable to take responsibility for myself. One might think that this self-negation consists precisely in bad faith's lack of reflective awareness of itself. I could not deceive myself, after all, if I knew what I was up to. And so one might argue that *good faith* or *sincerity* are healthy forms of selfhood (and, arguably, better phenomena on which to base a theory of selfhood). But, paradoxically, when Sartre *does* address good faith, he insists that 'it is a matter of indifference whether to be in good or bad faith, because bad faith takes hold of good faith and slides into its project even at its origin' (Sartre 1976: 106 n. 1/117 n. 27). A 'project' in Sartre's jargon is a purposive activity. So the claim here is that at the very moment I set out to be in good faith – to make a promise I fully intend to keep, to tell the truth just as I see it – I am already in bad faith. Along similar lines, Sartre tells us that sincerity is 'precisely a phenomenon of bad faith' (Sartre 1976: 98/108, translation modified), and 'sincerity's essential structure does not differ from that of bad faith' (Sartre 1976: 100/110).

Even though Sartre's account of bad faith has possibly attracted more discussion in the secondary literature than any other aspect of his work, few commentators have paid significant attention to these paradoxical claims about good faith and sincerity. The existing accounts can be summarised as falling into one of three approaches. i) Several commentators argue that Sartre *endorses* good faith as a kind of corrective to the failings of bad faith (e.g. Santoni 1995; Catalano 1980; Morris 2008; Gordon 1999; Webber 2009; McNulty 2025). Joseph Catalano, for example, argues that 'bad faith is an attempt to flee from our freedom, whereas good faith is an attempt to face our freedom' (Catalano 1980: 89). Similarly, Ronald E. Santoni suggests that 'good faith may be viewed as an attitude that confronts and affirms, rather than flees from, the freedom and responsibility to which (for Sartre) we have been abandoned' (Santoni 1995: 110). We will discuss their readings in more detail in section 5.[2] ii) A different reading is given by Zheng (2005) who argues that while Sartre sees no significant normative difference between good and bad faith, good faith is nonetheless marginally better than bad faith.[3] iii) Finally, a couple of commentators hold that for Sartre good faith is no better than bad faith (Manser, 1987; Bell, 1989).[4] In this paper, we develop in more detail Manser's and Bell's insight that, insofar as selfhood is at stake, Sartre sees no normative difference between bad faith and good faith and that he is right not to do so. Both fail to achieve authentic selfhood. What is more, a failure to recognize that good faith is no better than bad faith obscures Sartre's account of selfhood. When we resolve the puzzle of good faith, and can explain how it is that good faith is simultaneously the antithesis of and a phenomenon of bad faith, we will be in a position to understand something important about Sartre's account of selfhood – namely, that to be a self necessarily involves a 'disintegrated structure' (Sartre 1976: 671/806). Sincerity and good faith are bad-faith efforts to overcome the disintegration. Authentic selfhood, by contrast, involves assuming the disintegration, and acting with an understanding that these contradictions are an integral part of what it means to be human.

To understand how Sartre thinks of good faith, we need to clarify several of the concepts Sartre uses to frame his discussion of selfhood. First and foremost amongst these is Sartre's understanding of the negation involved in the very structure of consciousness; we take this up in section 3. This leads us to look more carefully at Sartre's description of the minimal reflexivity involved in selfhood. Sartre describes this kind of reflexivity in terms of a 'reflection-reflecting' (*reflet-reflétant*) structure – a kind of reflection which is to be sharply distinguished both from a consciousness that is turned back on itself (Sartre calls this *réflexion*, reflexion-with-an-x), and a consciousness engaged in reflective or deliberative (*réfléchi*) thought. We discuss reflection-reflecting, reflexion-with-an-x, and reflectivity in section 4. That finally will allow us to explore in section 5 the contrast between good faith, sincerity, and bad faith. We will conclude in section 6 with a discussion of Sartre's account of selfhood as being in a state of perpetual disintegration.

## 3 | SARTRE'S ACCOUNT OF CONSCIOUSNESS AS A 'UNITY OF BEING AND NOT BEING'

We call 'Sartre's paradox' his claim that:

*Sartre's Paradox*: A human being (a) is what it is not, and (b) is not what it is (see Sartre 1976: 93/101).

This paradoxical structure of human existence is the 'condition of possibility' of bad faith (Sartre 1976: 109/121), because bad faith involves consciousness turning its negation 'against itself' (Sartre 1976: 82/88). It is also the condition of the possibility of sincerity and, by implication, good faith. Sartre explains:

> we encounter, in the depths of sincerity, an incessant play of mirrors and reflections, a perpetual passage from being that is what it is into being that is not what it is and – in the other direction – from being that is not what it is to being that is what it is. And what is the goal of bad faith? To make it the case that I am what I am, in the mode of 'not being what one is,' or that I am not what I am, in the mode of 'being what one is.' Here again we encounter the same play of mirrors. Indeed, in order for the intention to be sincere to exist, it is necessary that – from the outset and at the same time – I am [what I am] and I am not what I am.
>
> (Sartre 1976: 101/111)

In the next section, we will come back to the idea of a 'play of mirrors and reflections.' But before we can say more about that, we need to untangle this paradox of human existence.

Let's start by laying out one of Sartre's examples to give us a phenomenological basis for thinking about the paradox. Sartre famously describes the case of a café waiter whose

> movements are animated and intent, a bit too precise, a bit too quick; he approaches the customers with a bit too much animation; he leans forward a bit too attentively, his voice and his eyes expressing an interest in the customer's order that is a bit too solicitous…. His behavior throughout strikes us as an act. He concentrates on his successive movements as if they were mechanisms, each one of them governing the others; his facial expression and even his voice seem to be mechanical; he adopts the pitiless nimbleness and rapidity of things.
>
> (Sartre 1976: 94/102–3)

In short, he is trying to be a waiter in the way that a table is a table – that is, in Sartrean jargon, to be a waiter in the mode of the in-itself. But human beings are not brute things. They choose their form of existence and maintain it through a free choice. '[B]y the very fact of maintaining this role in existence,' Sartre writes, the waiter 'transcend[s] it from all sides' (Sartre 1976: 95/104). Thus, Sartre concludes in a restatement of the paradox, 'I am a waiter in the mode of being what I am not' (Sartre, 1976: 95/104).

What are we to make of this? Philosophers are used to treating the verb 'to be' as multiply ambiguous, and distinguishing between the 'is' of predication, the 'is' of existence, the 'is' of identity, the 'is' of class-membership or class-inclusion, and so on. So one approach to making sense of the paradox might be to see it as trading in the distinction between predication and identity. When Sartre says, for instance, that 'I am the waiter in the mode of being what I am not,' one might read this as saying 'I am (predication) a waiter but I am not (identity) a waiter.'[5] But this would suggest that Sartre thinks that there is a simple, straightforward way in which predicates like 'is a waiter' are true of me, even if identity claims are not. This is something Sartre denies. For instance, even when it is true to say 'I am (predication) sad,' Sartre argues that there is also a sense in which 'I am not (predication) sad': 'the being of my sadness eludes me, through and in my very act of adopting it' (Sartre 1976: 96/105).

A second approach, proposed by McCulloch, is to distinguish between two different verbs, both of which are confusingly denoted in ordinary language as 'to be.'[6] One verb means 'to be in the mode or manner In-Itself' – which McCulloch glosses as having 'a thinglike, choiceless existence' (McCulloch 1994, 57, 58). The other means 'to be in the mode or manner For-itself' – which McCulloch glosses as having 'extreme freedom or transcendence' (McCulloch 1994, 57, 56). But it is not clear how this helps with Sartre's paradox because, as Morris points out,

'human reality is itself "ambiguously" both transcendence and facticity' (Morris 1997, 472) – both free and thinglike. What is the sense of 'to be' that means both free being and thinglike being? It cannot be either of McCulloch's two alternatives, and McCulloch's proposal offers us no further possibility.[7]

We propose that Sartre's paradox is best understood by disambiguating between two different *predicative* uses of the verb 'to be' – following Sartre's lead, we call the first 'external' or 'objective' predication, and the second 'internal' or 'determinative' predication.

In *external* or *objective* predication, to say that '*S* is *p*' is to say that *p* is an occurrent property of *S*, or that *p* is actually and fully instantiated in *S* at the present moment. To say that '*S* is not *p*' is to say that *p* is not an occurrent property of *S*, and that *p* is not actually and fully instantiated in *S* at the present moment. Sartre's examples of external affirmation are claims like 'The table is round' or 'The wall-hanging is blue' (Sartre 1976: 90/97). Examples of external negation are claims like 'The vase is not white' or 'The inkwell is not on the table' (Sartre 1976: 211/249).

In *internal* or *determinative* predication, to say that '*S* is *p*' is to say that what it is to be *S* is determined or defined by *p*, or by *S*'s relationship to *p*. And to say that '*S* is not *p*' is to say that not-being-*p* determines or defines *S*. Sartre illustrates 'internal negation' through examples like the assertion 'I am not rich' or 'I am not beautiful.' When denying that one is beautiful in the internal or determinative sense, one is not merely saying that beauty is a property that one happens to lack. Sartre explains:

> Uttered with a certain melancholy, these phrases mean not only that I am withholding from myself some particular quality but that the refusal itself comes to influence, in its internal structure, the positive being to whom it has been refused.... What I mean is that 'not being beautiful' is a particular negative property of my being, which characterizes me from inside, and also that – *qua* negativity – 'not being beautiful' is a real quality of myself, and this negative quality explains my melancholy, for example, just as effectively as my lack of worldly success. By 'internal negation' we mean a relation between two beings, where the one that we deny in relation to the other qualifies the other, at the heart of its essence, precisely by its absence. In this case the negation becomes an essential connection of being, since at least one of the beings on which it bears is such that it indicates the other, and bears the other in its heart as an absence.
>
> (Sartre 1976: 211/249–50)

One important upshot of this distinction is that I can be defined by properties that I lack in the objective sense. And it is possible to stand in an external or non-determinative relation to properties that I possess in the objective sense.

If we return now to Sartre's paradox, we can see that unraveling it turns on the thought that objective predication does not entail internal predication, and vice versa. So when Sartre says of the for-itself that 'it is what it is not,' the first 'is' is the 'is' of internal predication; the second 'is' is the 'is' of objective predication. We could rephrase the claim in the following way:

> '*S* is what it is not' = *S* has a significance (partially) determined by some property *p*, and *p* is not objectively instantiated in *S* at present.

For example, we are determined by our future aims and ambitions, even when (or especially when) we do not instantiate the property to which we aspire – think of the way that being a student is determined by the property of being a graduate, a property that students do not yet instantiate. As Sartre puts it, 'I am already what I will be (otherwise I would have no interest in being any particular way), *I am the one who I will be in the mode of not being he*' (Sartre 1976: 67/70).

The second half of Sartre's paradox likewise turns on the distinction between objective and internal predication. Here again, the first 'is' is the 'is' of internal predication; the second 'is' is the 'is' of objective predication. So we can rephrase the claim as follows:

'*S* is not what it is' = *S*'s significance is not determined by *p*, even though *S* objectively instantiates *p*.

McNulty has shown why Sartre holds that this paradox is an essential feature of human existence. Given the intentional structure of consciousness itself, consciousness is always defined by an object that it is not, and thus 'the *possibility* of bad faith is present with the advent of the for-itself' (McNulty 2025: 162; emphasis supplied). In what follows, we look at the 'down-stream' consequences of this basic structure of the pre-reflective cogito.

Sartre draws an ontological distinction between entities in terms of the kinds of predication they can sustain. He argues that ordinary objects like tables, inkwells, and rocks have 'the mode of being of the in-itself' because such objects do not admit of a gap between their determinative properties and their objective properties. The in-itself, as Sartre explains, has 'the mode of "being what one is"' (Sartre 1976: 101/111) or 'the mode of not-being-what-one-is-not' (Sartre 1976: 106/117). With the in-itself, there is no internal negation. A conscious being can assert objective negations of the in-itself, but such negations are not properly determinative of the in-itself; they are accidental rather than essential features of the in-itself. Sartre calls them 'external' negations, because they posit a relationship for an outside observer or witness, but this relationship 'is not constituted by its terms themselves' (Sartre 1976: 270/320). The fountain pen is in no way defined by not being a table; in itself, the in-itself simply is not what it is not.[8]

By contrast, conscious beings like us admit of internal negation; they have 'the mode of being what they are not and not being what they are' (see Sartre 1976: 217/256). Something that has this mode of being is never fully or exclusively determined by the properties it instantiates; it is to a significant degree defined by properties that are not occurrent, present, or objectively instantiated: 'The for-itself is being which, in its being, is not what it is and is what it is not' (Sartre 1995: 439/211). The for-itself's power of nihilation is the power to refuse determination by the properties it instantiates, and the power to determine itself through uninstantiated properties.

Following Heidegger's lead, Sartre calls all that can be objectively predicated of someone their '*facticity*.' Our facticity is comprised of features that we encounter as simply given, as what we must be[9]: 'This given manifests itself in several ways, although within the absolute unity of a single illumination. It is *my place*, *my body*, *my past*, *my position*' (Sartre 1976: 534/638).

A being that can bear an internal relationship to properties that it does not objectively instantiate enjoys '*transcendence*' – it is determinable by a relationship to properties it does not occurrently possess. At the same time, a transcendent being 'chooses itself as *this* surpassing *here* of the given' (Sartre 1976: 553/662, emphasis supplied), meaning that it is defined by the specific ways that it surpasses its occurrent properties. Of course, this means that, in a certain sense, the transcendent being cannot be defined *without* reference to its occurrent factical properties. The capacity for transcendence is a product of consciousness's ability to negate – to refuse to identify with its factical features, and to define itself in terms of what it is not. Thus, a conscious being always 'exists *at a distance from itself*' (Sartre 1976: 114/128) – it has determinative properties that are uninstantiated; it has instantiated properties that are non-determinative but serve to delimit the bounds of its transcendence.

## 4 | REFLEXION AND THE 'REFLECTION-REFLECTING' STRUCTURE

Having clarified the mode of being of consciousness, we now need to explore the type of minimal reflexivity which is constitutive of Sartrean selfhood.

Sartre is committed to two further fundamental theses about consciousness. The first thesis is Sartre's appropriation of what we can call

*The Husserlian intentionality thesis*: 'All consciousness is consciousness *of* something.'[10]

That is, all consciousness is directed at or aims at an object (and 'object', here, is understood in a very broad sense to include whatever we can perceive, think about, manipulate, and so on). Sartre calls the specific form of consciousness that is directed at an intentional object 'positing' or 'thetical.' Insofar as consciousness is intentional, then, it can also be said to be 'positional' or 'thetic.' Thus, Sartre reformulates the Husserlian Intentionality Thesis in this way:

> *The Sartrean intentionality thesis*: 'All consciousness is positional in that it transcends itself to reach an object, and it is exhausted by just this act of positing' (Sartre 1976: 18/10).

The second thesis to which Sartre is committed is:

> *The minimal reflexivity thesis*: 'All positional consciousness of an object is at the same time a non-positional consciousness of itself' (Sartre 1976: 19/11).

At its most basic level, then, our relationship to ourselves is not positional. The minimal reflexivity that is constitutive of selfhood is not the kind of reflexivity in which consciousness itself is the intentional object of consciousness. That means, on Sartre's account, I can be in a minimally reflexive state without turning my attention to my self, my ego, my state of consciousness, or indeed any conscious act.

In discovering the form of reflexivity that is essential to selfhood, Sartre argues that we can rule out *introspection* – i.e., states or acts of consciousness that make the agent herself, her 'I' or ego, into an object of consciousness. Sartre dubs introspective acts '*reflexive*'(—with-an-x, *réflexif*), and he calls a process or activity of examining one's mental states and acts '*reflexion*'(—with-an-x, *réfléxion*). In introspection, we posit states or acts of consciousness as the objects of an intentional act. But clearly I can be conscious without introspecting: 'a consciousness has no need of a reflecting consciousness in order to be conscious of itself. It merely does not posit itself to itself as its own object' (Sartre 1966: 29/11).

A closely related form of reflexivity involves states or acts of consciousness that make the agent's *actions* (rather than the agent herself) into an object of consciousness. We can call such states and acts '*reflective*' or '*deliberate*.' A reflective mental act is one in which I monitor myself and my progress as I do something – like watching my fingers type on a keyboard. In a 'reflective state,' Sartre explains, 'I am watching myself act, in the same sense that we say of someone that he is listening to himself talk' (Sartre 1966: 42/20, translation modified).

But deliberate self-monitoring is no more necessary to consciousness than introspection is. That is, most of the time the aim or target of my consciousness is not me – not the ego – but the equipment and people with which I concern myself. And most of the time I am not monitoring my own actions: I am absorbed in responding fluidly to the matters of concern. Sartre illustrates with examples of fluid, skillful action:

> When I run after a tram, when I look at the time, when I become absorbed in the contemplation of a portrait, there is no I. There is a consciousness of the *tram-needing-to-be-caught*, etc., and a non-positional consciousness of consciousness. In fact, I am then plunged into the world of objects, it is they which constitute the unity of my consciousnesses, which present themselves with values, attractive and repulsive values, but as for *me*, I have disappeared, I have annihilated myself. There is no place for me at this level, and this is not the result of some chance, some momentary failure of attention: it stems from the very structure of consciousness.
>
> (Sartre 1966: 32/13)

The phenomenology of ordinary, everyday, fluid action thus shows that our most fundamental state of consciousness is both *prereflexive* – it is not directed at the ego – and *unreflective* – there is no self-monitoring going on. Fluid action is 'a way of losing myself in the world, of allowing things to soak me up – like ink by a piece of blotting

paper – so that an equipment-structure, oriented toward an end, stands out synthetically against the ground of the world' (Sartre 1976: 298/355–6).

Because I can consciously engage with the world without monitoring myself or making myself into an object, Sartre regards our active, fluid responsiveness to the affordances of the world as the most fundamental or basic form of consciousness. It is '"first order" or "*unreflective*" consciousness' (Sartre 1966: 24/16; translation modified). Introspection and reflective self-monitoring are second-order operations of consciousness, conscious acts directed at consciousness.

But this presents something of a paradox. If first-order consciousness is unreflective and prereflexive, in what way can it have even a minimal reflexivity? What is it to have a minimally reflexive relationship to myself even when I am fully absorbed in dealing with the world around me?

According to Sartre, the most fundamental or 'first order' form of reflexivity – the form at play when we are caught up in unreflective activity – consists in a mirror-play of meanings between us and the world. This mirror-play creates what, following Sartre, we call '*The Reflection-Reflecting Structure*' (Sartre 1976: 112/125):

a. the world shows up as a *reflection* (*reflet*) of the agent's aims and intentions;
b. the affordances of the agent's environment inspire her to pursue certain projects; thus her intentions are *reflecting* (*reflétant*) her situation.

As Sartre puts it with respect to (a), my projects 'bring the world into being' (Sartre 1976: 373/447). He does not mean that I make objects exist, but rather that my projects give them the order and significance through which they solicit me to act. But that means that, insofar as the agent sees and responds to the way the world shows up, the world *reflects* her back to herself: 'the world sends back to us, through its very articulation, the image of what we are' (Sartre 1976: 507/606).

Of course, when we are fluidly responding to the world we do not have a sense that we are playing a role in the way things show up and solicit us to respond. We got a taste of this already in Sartre's description of running after the streetcar; when absorbed in this activity, he noted, objects 'present themselves with values, attractive and repulsive values,' albeit values that reflect back to me my concerns (Sartre 1966: 32/13). Sartre explains:

> It is just as if we lived in a world where objects, apart from their qualities of heat, odour, shape, etc., had those of repulsive, attractive, charming, useful, etc., etc., and *as if these qualities were forces that performed certain actions on us* (Sartre 1966: 41-42/19, emphasis supplied).

So, even though we are not typically aware of it, it is nevertheless the case that the world as it appears shows me the character of my current active engagement with the world. Simply in acting, the world reflects me back to myself through what I am drawn to do.

The reflection-reflecting relationship between the agent and the world does not run in just one direction, and the agent-in-flow is not fully in control of her intentions. So with respect to (b), minimal reflexivity is found in the way that the field of affordances within which I find myself guides the way I move my body and inspires in me intentions that I did not have before. Thus, my actions are always *reflecting* the world.

Sartre notes that there is

> a radical '*Unselbstständigkeit*' [lack of independence] in the two terms, 'reflected and reflecting' … i.e., an inability to posit themselves separately, such that the duality remained perpetually evanescent, and such that each term, by positing itself for the other, became the other.
>
> (Sartre 1976: 187/219)

So through the minimal reflexivity of the '*reflection-reflecting*' (*reflet-reflétant*) relationship, the agent and her world stand in a state of radical interdependence – the meaning of each is 'constantly referred' to the other (see Sartre 1976: 189/221). This 'quasi-duality' of agent and world makes up 'the primary internal structure' of the self (Sartre 1976: 187/220). And in second-order forms of reflexivity like introspection and self-monitoring, it is the reflection-reflecting structure itself – this evanescent, reversible agent-world relationship – that is reflected-on (Sartre 1976: 188/220). But (and this is an important qualification) ordinary reflexion 'can never decipher this worldly image' (Sartre 1976: 237/281). As we will discuss in the following section, our reflexive acts typically distort the true character of the interdependence between the agent and her world.

Sartre dubs this most minimal form of reflexivity 'the circuit of ipseity.' 'Ipseity' is Sartre's term for referring to selfhood (see Sartre 1976: 52/52). The reflexivity involved in selfhood, Sartre argues, takes the form of a circular mirror play between our intentional being toward the world, and the world's solicitations to action. 'To establish the circuit of ipseity' is to 'insert[] … the world between the for-itself and its being' (Sartre 1976: 645/776), and 'it is on the world's basis that human-reality becomes acquainted with what it is' (Sartre 1976: 234/277–8). In other words, the order of significance that I find in the world 'is the image, projected into the in-itself, of my possibilities, i.e., the image of what I am' (Sartre 1976: 237/281). Imagine Sartre reading a book in a Paris café. All around him are the clatter of dishes, the conversations at nearby tables, the movement of the waiters, the flow of traffic outside, the smell of coffee. Because his project is to be a philosopher and writer, all of that withdraws to allow the book to stand out and solicit him to read it: 'I do not lose sight of the colors or the movements surrounding me, or cease to hear the sounds – it is simply that they become lost within the undifferentiated totality that serves as the background for my reading' (Sartre 1976: 374/447). The fact that the book stands out as the salient feature of the world, that it calls for a response, shows Sartre something about who he is.

With this, we can return to the distinction Sartre wants to draw between internal and objective determination. My objective properties cannot, in the first instance, define the self that I am because I only discover their significance when I see how the world gets organized in response to them, or how they influence my response to the affordances of the world. I discover what it means to be tall, for instance, because I have to duck to get through doorways. I discover that I am thirsty because I am drawn to drink from the water glass. In other words, my facticity is always disclosed as significant through the world and in terms of my projects: 'what we have called freedom's "facticity" is the given that it *has to be*, and which it lights up with its project' (Sartre 1976: 534/638).

To be a self, then, is to

> (A) *experience the world in the light of my own projects, and my projects in the light of my world.*

That means that a self recognizes a 'negation' or lack in this actual particular world – namely, the fact that a possibility which the self desires has not been actualized – and feels itself solicited by the world to correct that lack. So, in addition, being a self involves:

> (B) *an activity of setting out to actualize my projects and thereby to give me, as an individual, a concrete presence in the world.*

That is to say, selfhood aims at expressing itself in the world. To be a self is to 'surpass' the world 'toward a new concrete "state" of the same world' (Sartre 1976: 644/775; translation modified). Sartre reiterates this in the *Notebooks for an Ethics*, arguing that our purpose is 'unveiling the maximum of being by being one's self as much as possible (not as Ego but in terms of ipseity)' (Sartre 1983: 503/487, translation modified). 'This unveiling,' Sartre concludes, 'occurs as surpassing toward…' – i.e., as an active transformation of the situation one inherits (Sartre 1983: 503/487, translation modified).

Selfhood or ipseity, then, does not require a deliberate or reflexive form of reflexivity. Selfhood is not a state of mind, but an on-going purposive activity in which I assert myself as an individual in the world. But with this, we

return to the puzzle about bad faith and good faith with which we started. If selfhood does not require reflexivity, the problem with bad faith cannot be that it lacks reflective awareness of itself. In what sense, then, is bad faith a 'negation of self', a pathological condition of selfhood?

To answer this question, we have to recognize one further necessary condition of authentic selfhood. Sartre argues that to be a self also involves

> (C) *being subject to a normative constraint, namely an obligation to take responsibility for oneself*.

Given Sartre's paradox, taking responsibility for who one is cannot mean accepting one's occurrent character traits and projects as fixed or necessary. Instead, Sartre draws a distinction between taking responsibility as 'acceptance' and taking responsibility as 'assumption':

> Not to *accept* what happens to you. That's too much and not enough. To *assume* it (when you've understood that nothing can happen to you except by your own hand), in other words to adopt it as one's own, exactly *as if* one had given it oneself by decree, and, accepting that responsibility, to make it an opportunity for new advances, *as if* that were why one had given it oneself.
>
> (Sartre 1995: 296/95)

Acceptance is 'not enough' because it falls short of genuine responsibility by failing to appreciate our role in disclosing the significance of the world. And acceptance is 'too much' because it treats characteristics as *fixed* when in fact they are open to modification, and it treats the meaningful situation as *objective* (i.e., as independent of us) when in fact we play a role in disclosing it. Acceptance thus leads to a resignation that surrenders our responsibility to change the world.

To *assume*, by contrast, 'means to adopt as one's own, to claim responsibility' for one's actions and one's circumstances, even while understanding that we are never in a position to fully determine who we are and what we do (see Sartre 1995: 143/113). Too much of who we are is a product of contingencies – facts about where we were born, how we were raised, what predispositions and traits we have, and so on. Too much is also subject to the legitimate interpretation others impose on our actions. But precisely because so much of who we are is beyond our control, the only hope we have of justifying our existence depends on our 'assuming' or 'taking over' who we are, and using that as the basis for questioning and surpassing our default condition.

In saying that responsible persons act 'as if' they had decreed the situation they find themselves in, Sartre insists that the 'as if' 'is not a lie' (Sartre 1995: 296/95). Sartre believes that there is a real sense in which 'all that does happen to [a person] can do so only by [her] own hand and within [her] responsibility.... One is totally responsible for one's life' (Sartre 1995: 296–7/95). Our discussion of the circuit of ipseity helps to explain this view. The situations in which we find ourselves, as well as the things we do and suffer, are unveiled only through our way of projecting ourselves into the world. The contingent features of the world, the facticity of things, present 'coefficients of adversity' rather than absolute determinations of the meanings we can unveil through our projects and actions:

> the coefficient of adversity of things cannot be an argument against our freedom, because it is through us, which is to say by means of an end that we have posited beforehand, that this coefficient of adversity arises. This rock, which manifests a profound resistance if I want to move it, will on the contrary become a valuable aid if I want to climb up it in order to contemplate the landscape. In itself – if it is even possible to consider how it might be in itself – it is neutral, which is to say that, before it can show itself to be an adversary or a help, it awaits the illumination of an end.... It is our freedom therefore which constitutes the limits it will thereafter encounter.
>
> (Sartre 1976: 527/629–30)

As we have seen, we are not ordinarily thetically aware that we are playing a role in the disclosure of the limits of our world. The soldier does not literally determine by decree, for instance, that the situation he confronts will be a terrifying one. The language of 'as if' also acknowledges this – we should not treat our role in disclosing the world as though it were a reflective act when, of course, it is not. But whereas acceptance is an enervating response, proceeding as if we are responsible is invigorating; it encourages us to seek out the possibilities for questioning and surpassing the status quo.

Sartre, then, argues that to be a self is (a) to experience the world in the light of my projects and to experience my projects in the terms provided by my world; (b) to work toward actualizing my projects, thus giving me a concrete presence in the world; and (c) to have an obligation to take responsibility for the situation in which I find myself: I take responsibility by surpassing my givens – that is, by transforming their significance through the projects I undertake.

## 5 | BAD FAITH AND GOOD FAITH

With Sartre's account of selfhood in view, we can see more clearly that what is bad about bad faith is precisely that it distorts my understanding of the true nature of my interaction with the world, and it does so in such a way as to conceal from me my obligation to take responsibility for myself. 'The goal of bad faith,' Sartre explains, 'is to put oneself out of reach; it is an act of flight' (Sartre 1976: 100/111).

As Jonathan Webber (2009: 74, 96) points out, Sartre uses 'bad faith' in a variety of ways. In its broadest sense, 'bad faith' describes a flight from responsibility abetted by a failure to coordinate facticity and transcendence. Sartre explains that

> what unites these various aspects of bad faith … is a certain art of forming contradictory concepts, i.e., concepts in which an idea and the negation of that idea are united. The underlying concept generated in this way makes use of the twofold property of human beings, of being a *facticity* and a *transcendence*. These two aspects of human-reality are, in truth -– and ought to be – capable of being validly coordinated. But bad faith does not want to coordinate them, or to resolve them by means of a synthesis. From its point of view, it is a matter of affirming their identity, even while preserving their differences. Facticity must be affirmed as *being* transcendence and transcendence as *being* facticity, in a way that allows us, at the moment we apprehend one of them, to find ourselves suddenly faced with the other.
>
> (Sartre 1976: 91/98–9)

So in its broad sense, 'bad faith' characterizes any project that fails to coordinate or synthesize facticity (the fixed elements of our being-in-the-world, which are beyond our capacity to change at will) with transcendence or freedom (our constant activity of changing and surpassing the current situation while creating and discovering new possibilities).

In addition to this generic conception of bad faith, Sartre also identifies a number of different species of bad faith. It is this variety of types of bad faith that allows Sartre to hold both that

a. Good faith is a type of bad faith (see, e.g., Sartre 1976: 106 n.27 / 117 n.27);

and that

b. Good faith is opposed to bad faith (see, e.g., Sartre 1976: 105/116–7).

In order to reconcile (a) and (b), we need only recognize that good faith is a species of bad faith, but one that differs in an important way from other types of bad faith. To explain this, we will begin by surveying the different possible forms that bad faith can take.

If we take Sartre's observation that human being essentially involves a twofold property of facticity and transcendence, and combine that with his observation that these properties can be either affirmed or denied, we can distinguish four different ways of trying to 'coordinate' our facticity with our transcendence (see chart 1).

One possibility, of course, is the 'valid' coordination of facticity and transcendence that Sartre calls 'authenticity.' A valid coordination would mean that we acknowledge as fixed what in our situation is fixed, but also recognize what is capable of being surpassed. We'll look briefly at authenticity in section 6. Setting authenticity aside, then, that leaves us with three basic varieties of bad faith.

First, one might define oneself in terms of one's factical features while denying one's capacity for transcendence. In this form of bad faith (BF1), as Sartre puts it, 'transcendence is affirmed as *being* facticity' (Sartre 1976: 91/99). As a result, one comes to see one's actions as the causal consequence of one's psychic characteristics and traits. For instance, suppose Jones breaks up with his partner via WhatsApp – an action conventionally regarded to be a cowardly way of ending a relationship. Instead of recognizing that he could have acted differently, he justifies himself by saying 'I am a coward. This is just who I am.' Jones is right to identify his action as cowardly; he is not free to redefine its social significance at will. And so he correctly recognizes the factical character of his situation. But instead of acknowledging that his action was a free choice, he seeks to evade responsibility by defining it as the product of a fixed character trait – his *cowardice*. In this way he denies his transcendence – his capacity to act in a way that is necessitated neither by the factical features of the situation, nor by his occurrent psychological properties.

The second form of bad faith (BF2) is the inverse of the first: here, one defines oneself by one's capacity for transcendence while failing to recognize appropriately the factical constraints that inform the meaning of one's action. As Sartre puts it, 'facticity is affirmed as *being* transcendence' (Sartre 1976: 91/99). While our facticity is something that we *can* transcend, this does not mean that our factical situation is something we can redefine at will: 'surpassing it is not to negate it … but to invent something on its basis' (Sartre 1983: 80/74–5). So to transcend our

**CHART 1**

| | Affirming transcendence as being transcendence (i.e., appropriately recognizing one's ability to surpass the constraints of facticity) | Affirming transcendence as being facticity (i.e., inappropriately treating the features of one's consciousness as if they are fixed and determinate factical properties) |
|---|---|---|
| Affirming facticity as being facticity (i.e., appropriately recognizing the constraints that fixed factical properties impose on the significance of our actions) | Authenticity: A valid coordination of facticity and transcendence | Bad Faith 1: An agent recognizes genuine factical features of the situation, but erroneously takes those factical features as evidence of fixed character traits and determining psychological structures – thereby denying her transcendence and responsibility. |
| Affirming facticity as transcendence (i.e., inappropriately treating the fixed factical features of an action as if they can be freely redefined) | Bad Faith 2: An agent correctly recognizes that her actions are not wholly determined and defined by factical properties. But she erroneously acts as if factical constraints can be freely dismissed and redefined by her capacity for transcendence. | Bad Faith 3: An agent treats an aspect of her transcendence as a fixed and determining factical property, thus attributing fixed traits to herself that she does not instantiate. At the same time, she denies to her actual factical traits their true significance. |

facticity validly is to come to terms with the constraints it imposes on us, to recognize it as establishing 'coefficients of adversity' that shape all of our undertakings: 'The project is surpassing something given toward an end. But this end, which is a change in the current state of affairs, must be constituted in and through the current state of affairs' (Sartre 1983: 145/137). 'Bad faith,' however, sometimes takes the form of imagining 'a purely fictive surpassing which refuses any compromises' (Sartre 1983: 141/133) – fictive, because it fails to recognize that transcendence must work with the givens of facticity. It indulges in the self-deception that it can freely dismiss factical constraints. Consider Mathieu, the protagonist of Sartre's novel *The Age of Reason*: Mathieu is obsessed with his own freedom, wanting to avoid all permanent attachments and commitments. Then he discovers that the woman he has been seeing for years, Marcelle, is pregnant. He desperately tries to organise an abortion, as he feels his freedom threatened by the prospect of parenthood and marriage. The other characters, however, recognise that Mathieu is in bad faith. When Mathieu claims that all he wants is 'to retain [his] freedom', his brother Jacques points out that what Mathieu really wants is 'a comfortable life' and 'an appearance of liberty,' enjoying the advantages of marriage, while 'avoid [ing] its inconveniences' (Sartre 1945: 114/137). If I engage in this type of bad faith, Sartre explains, 'precisely through my transcendence, I escape from everything that I am' (Sartre 1976: 92/100). If I deceive myself into acting as if I am at liberty to redefine the meaning of my actions in any way that I want, I lose my grip on the factical features that give positive content to my existence. True freedom does not mean that Mathieu can be his transcendence, without the constraints of facticity. True freedom, as Jacques explains in Sartre's voice, 'consist[s] in frankly confronting situations into which one had deliberately entered, and accepting all one's responsibilities' (Sartre 1945: 115/138).

Finally, the third form of bad faith (BF3) combines the errors of both of the first two forms. It treats as *factical* certain features of one's self-understanding that are transcendent (for instance, one's aspirational aims). And it simultaneously denies the significance of certain occurrent factical properties of the situation (that is, it misunderstands its transcendence as the power to redefine factical features at will). Suppose, for instance, that Brown is a hiring coordinator who aspires to diversify the personnel in her workplace. She treats this aspiration as an occurrent feature of her psyche: she takes herself to *be* in actuality a progressive champion of diversity and equality in the workplace. At the same time, whenever there is a job opening at Brown's company, she ends up hiring the white candidate (even when equally- or better-qualified minority candidates are available). If you ask Brown about this, she tells you, and actually believes herself, that she really wanted to hire a minority candidate. But she gives you a list of reasons why the white candidate was better qualified for the job. Like Jones, she regards the meaning of her action to be fully determined by a factical property – only in Brown's case it is one that her actions do *not* in fact instantiate. And like Mathieu, she is in denial of the factical meaning of her actions.

On Sartre's analysis, each of these forms of bad faith results in a failure 'validly to coordinate' 'the twofold property of human beings, of being a facticity and a transcendence' (Sartre 1976: 91/99). Each in a different way either affirms facticity as being transcendence, or affirms transcendence as being facticity, or both. In other words, they each give rise to a distinctive form of bad faith (as summarized in Chart 1).

What, then, shall we say about *good* faith?[11] If we focus only on the generic concept of bad faith as a flight from being responsible for myself, then it is perfectly natural to suppose that 'good faith' must be the embrace of self-responsibility. Perhaps it is no surprise, then, that several commentators who discuss good faith take just this line. Santoni, as we saw, interprets good faith as 'an attitude that confronts and affirms, rather than flees from, the freedom and responsibility to which (for Sartre) we have been abandoned' (Santoni 1995: 110). While he distinguishes good faith from authenticity, Santoni nonetheless reads good faith positively, as 'the basic attitude of accepting ourselves – without regret, remorse, despair, or excuse – as anguished freedom and of taking responsibility for our choices, projects, attitudes' (Santoni, 1995: 87).[12] Catalano holds a similar view. While he accepts that 'as ideals, both good and bad faiths must fail, since man can never be one with any of his ideals', he goes on to claim that 'Sartre repeatedly states that bad faith is an attempt to flee from our freedom, whereas good faith is an attempt to face our freedom' (Catalano 1980: 89). Thus, while Catalano thinks that good faith is doomed to slide into bad faith, he sees a real normative difference between the two. But neither Catalano nor Santoni can offer any textual support

for these claims. In fact, we cannot find Sartre stating even once, let alone 'repeatedly,' that good faith is an attempt to face our freedom. To the contrary, Sartre repeatedly describes good faith as itself a form of flight (see, e.g., Sartre 1976: 105/117). It is simply not the case that Sartre draws the contrast between bad faith and good faith in terms of the contrast between flight from and facing up to our freedom. Instead, the contrast is drawn in terms of two different kinds of flight: 'Good faith wants to flee from "not believing what one believes" into being; bad faith flees from being into "not believing what one believes"' (Sartre 1976: 105/116–7).

A more promising line of interpretation involves reading good faith as an epistemic attitude that Sartre endorses. For instance Morris, as well as Santoni and Catalano, argues that a key difference between bad faith and good faith lies in the way we treat evidence. While in bad faith we set 'the standards of evidence low,' a person living in good faith is committed 'to examin[ing] the evidence critically' (Morris 2008: 87). Unlike the previous interpretation of good faith as an acceptance of freedom, this reading at least has some textual basis – Sartre does indeed draw a contrast between bad faith and good faith in epistemic terms: 'bad faith,' Sartre explains, 'does not retain the norms and criteria of truth that are accepted by the critical thinking of good faith' (Sartre 1976: 103/114). While recognizing that Sartre categorizes good faith as a form of bad faith, Webber nevertheless sees this passage as a kind of endorsement of good faith. According to Webber, '[g]ood faith is holding with unwarranted certitude a belief that is nonetheless warranted by the evidence' (Webber 2009: 94). Thus, while its degree of confidence is unwarranted, Webber argues, 'good faith does at least base the content of the belief on the preponderance of evidence, whereas bad faith takes any evidence to be sufficient warrant for belief' (Webber 2009: 94). Gordon (1999: 56) too sees this passage as proof that Sartre endorses at least a kind of good faith. He distinguishes between 'ignorant good faith' and 'authentic good faith.' The former involves 'believing what one believes, in short, belief without awareness of the belief as a belief,' whereas the latter 'involves making the choice to oneself to treat evidence in appropriate ways conducive to the admission and self-admission of the situation of truth' (Gordon 1999: 56).

But these interpretations turn on assuming that Sartre *endorses* good faith's 'norms and criteria of truth' – that Sartre believes an attitude of good faith can actually secure the truth about us. But can it?

For starters, it is important to note that Sartre discusses good faith's 'norms and criteria of truth' in order to contrast two different ways of responding to 'evidence.' 'Evidence' is a term of art that Sartre borrows from Husserl, and which Sartre defines as 'the possession of an object in intuition' (Sartre 1976: 103/113). For Husserl, truth consists in the *fulfillment* of an intention. *Evidence* is (a) the presentation in intuition of a state of affairs that (b) motivates me to accept an intention as 'fulfilled' or true (Husserl 1984, 'Sixth Investigation', esp. §39). The phenomenon of evidence thus includes two components – the agent's 'meaning-intention,' and the 'state of affairs' that is intuited as fulfilling that intention. But Husserl also insists that the *experience* of a congruence between the intention and the fulfillment is not the same as the *recognition* of truth – this requires a further 'act of objectifying interpretation' (Husserl 1984, §39).

It is in this gap between an experience of evidence and a recognition of truth that a certain kind of bad faith operates. Through this kind of bad faith, Sartre argues, 'a particular type of evidence appears: non-persuasive evidence. Bad faith grasps the evidence, but is resigned in advance to not being fulfilled by this evidence' (Sartre 1976: 103/114, translation modified). The kind of bad faith Sartre has in mind here is one which exploits Sartre's paradox by treating 'objects' as having the type of being that belongs properly only to consciousness – that is, it treats 'the type of being possessed by objects' as a being that 'is what it is not, and is not what it is' (Sartre 1976: 103/114; see our discussion of Sartre's paradox above). This allows one to resist the inclination to judge to be true the intuitive presentation of things. Bad faith thus treats the factical givens as if they can be dismissed or freely reinterpreted. The norm or criterion of truth for this approach to evidence is: I must accept to be true only what I decide to affirm: 'faith is a decision, and … after each intuition, it is necessary to determine what is and to will it' (Sartre 1976: 103/114, translation modified). Taking evidence to be non-persuasive in this way is a defining characteristic of the kinds of bad faith we designated as 'BF2' and 'BF3'. Both these forms of bad faith treat factical constraints as something that can be freely dismissed or reinterpreted – we can decide not to be bound by the evidence. In addition, BF3 employs the power to decide what *is* in order to treat its transcendent projects as

objective, factical features: 'I resolve to be poorly convinced in order to convince myself that I am what I am not' (Sartre 1976: 104/115).

Sartre contrasts this approach to evidence with the approach taken by 'good faith' and 'sincerity'. Where BF2 and BF3 treat objective properties as being what they are not and not being what they are, good faith and sincerity treat consciousness (transcendence) as being what it is and not being what it is not. But this is just the inverse form of the error that BF2 and BF3 make. While in BF2 or BF3 I am resigned in advance to not being persuaded by evidence, in good faith 'I give in to my impulses to trust [the evidence], I decide to believe it and to stand by my decision, and, finally, I behave as if I were certain of it – all in the synthetic unity of a single attitude' (Sartre 1976: 104/115). But this good faith attitude, which moves immediately from evidence to conviction, is also a distortion of an authentic relationship to evidence. It tries to cover over the gap we observed earlier between the experience of fulfillment and the intentional positing of the fulfillment as the truth. But, Sartre observes, we cannot help but be non-thetically conscious of this gap: 'one's non-thetic consciousness (of) believing is destructive of belief…. [T]he law of the prereflexive cogito dictates that the being of believing must be a consciousness of believing. Thus, belief is a being that puts itself in question in its own being, and that can be actualized only in its destruction' (Sartre 1976: 104/116).

So, according to good faith's criterion of truth, we should accept that what an action truly *is*, is simply determined by its objective (factical) properties. Likewise, according to the norms of sincerity, a person truly and simply is what she is and is not what she is not. But, Sartre notes, 'the ideal of good faith (believing what you believe) is, like that of sincerity (being what you are), an ideal of being-in-itself' (Sartre 1976: 105/116) – hence, not an appropriate ideal to fix the norms and criteria of truth as applicable to conscious beings. Good faith and sincerity alike aim at taking evidence at face value, while ignoring the role of human transcendence in unveiling of truth.

The upshot is that good faith is an attitude that fails to coordinate our facticity and transcendence (thus, an attitude that belongs to the broader genus of bad faith attitudes). Indeed, good faith belongs to the class we have identified as BF1. Good faith first identifies some actual immediate objective property of an action; it then takes that property as fully definitive, and thereby treats the action as something in-itself – something that just is what it is. It is characteristic of good faith and sincerity that they attempt to overcome the tension between simple faith (e.g. simply being convinced that Pierre is my friend) and the uncertainty that accompanies belief once it is recognised as merely a belief. Good faith and sincerity deny that belief is always also characterised by some measure of uncertainty. They 'turn[] into bad faith because [they are] going to neglect that quiet voice that says: "I really only believe it"' (Sartre 1983: 492/475).

It is clear, then, that Sartre does not endorse good faith in opposition to bad faith in general. Instead, he criticizes good faith as a particular species of bad faith. Like BF2 and BF3, good faith (i.e., BF1) fails when it comes to effecting a valid coordination of facticity and transcendence.

This is clear from the fact that good faith tries to uphold as an ideal values like: we should say only what we really think; we should promise only what we really intend to do; we should attribute motivations and intentions to ourselves and others only when they can be proven apodictically. But simply in accepting those values, Sartre argues, we buy into a distorted picture of human action. Each of those good-faith norms assumes that what I believe, what I intend, what motivates me to act, is fully fixed by occurrent features of me and the world. The norms of good faith thus presuppose that my actions are definable entirely in terms of objective psychic states that exist in the mode of the in-itself.

For instance, Sartre invites us to imagine under what conditions 'I would be "in good faith" in declaring that I am not a coward' (Sartre 1976: 101/112). To assert this in good faith, you would have to be not a coward 'in the straightforward mode of not-being-what-one-is-not' (Sartre 1976: 101/112). But the very fact that you are called upon to assert it in the first place means that there is room for the possibility that you are in fact a coward. In your good faith declaration that you are not a coward, then, you are trying to get others 'to hang an objective label on you that you will then internalize in the form of an element of your psyche or as an in-itself-for-itself' (Sartre 1983: 491/474). What is more, one could only objectively establish oneself as *not* being a coward by pointing

to occurrent facts about actions one has performed, or mental affects one has experienced in the face of danger. But, in fact, there is no objective definition of what one has to do or not-do, what mental states one must have or lack, in order to be a coward. 'It is false to see some given quality here,' Sartre notes, 'since it has to be continually modified' (Sartre 1983: 490/474). The upshot is that when Sartre points out that good faith gives in immediately to the impulse to trust, he is not *endorsing* good faith; he is *criticizing* it (see Sartre 1976: 104–5/115–6).

If we return now to the epistemic reading of good faith, we can see that good faith is really not any better off as an epistemic attitude than bad faith. Interpreters like Webber suggest that good faith is better than bad faith insofar as 'good faith does at least base the content of the belief on the preponderance of evidence, whereas bad faith takes *any* evidence to be sufficient warrant for belief' (Webber 2009: 95). And it is true that good faith is grounded in evidence in a way that other forms of bad faith are not. But this is a little like thinking that it is more epistemically sound to base my decisions on my horoscope than on crystal ball gazing, because the horoscope is based on a consideration of astrological facts. In other words, the way that good faith uses evidence to construct a fictional account of the agent's character is just as much of a distortion of what is really going on as the other forms of bad faith. This is why Sartre insists that 'it is a matter of indifference whether to be in good or bad faith, because bad faith takes hold of good faith and slides into its project even at its origin' (Sartre 1976: 106/117).

## 6 | SELFHOOD AS A PERPETUALLY DISINTEGRATING SYNTHESIS

In recognizing that good faith is in no sense endorsed by Sartre – in recognizing, that is, that good faith is just a particular species of bad faith – we are directed forcefully to a core aspect of Sartre's account of selfhood: to be human is necessarily to be in a state of perpetual disintegration. 'The very nature of our being,' Sartre argues, involves 'an internal disintegration at the heart of being' (Sartre 1976: 105/117). 'Bad faith's most basic act is to flee what we are' (Sartre 1976: 105/117). This flight takes two forms: 'Good faith [BF1] seeks to flee from the internal disintegration of my being, in the direction of the in-itself that it has to be, and is not. Bad faith [BF2 & BF3] seeks to flee from the in-itself into the internal disintegration of my being' (Sartre 1976: 105/117). As long as one mistakenly believes that Sartre endorses good faith, this claim appears to amount to a repudiation of disintegration. But by recognizing that good faith is a form of bad faith – a form of flight from responsibility – we can see that Sartre is making a very different point: namely, that rather than overcoming it, we need to find an appropriate way of embracing the disintegration at the heart of our being. The very next sentence makes this clear. The problem with BF2 and BF3 is not just that they flee from their facticity into disintegration, but that they flee in a way that 'denies this very disintegration' (Sartre 1976: 105/117).

To explain this, we need to emphasize that Sartrean disintegration (*désagrégation*) does not mean the disappearance of an object as it breaks up into fragments or particles. Rather, it is a dis-*integration* – a resistance to integration, a tendency within the elements that make up a whole to pull apart, to lose their sync. Things are prone to disintegration when their aspects or parts necessarily belong together, but resist being determined simultaneously or brought into view all at once. A familiar example of a disintegrating phenomenon is the relationship between figure and ground in perception. Sartre explains: 'there are two figures and the condition of one of them appearing is the disintegration of the other… [W]e can constitute one object as the *figure* by repelling some other object to the point at which it becomes the *ground* and *vice versa*' (Sartre 1976: 55/55).

The 'circuit of ipseity' is just such a disintegrating phenomenon. As we saw in Section 4, Sartre argues that my selfhood is found in the constant reciprocal interaction between objective features of my situation and my ability to surpass my situation, between our intentional projects and the factical setting of the world. Sartre's diagnosis of the various types of bad faith, however, allows us to recognize that this circuit is prone to constant disintegration, meaning that the centripetal and centrifugal moments in the circuit fall apart. Good faith affirms facticity while trying to deny the centrifugal motion of our meaning-giving activity. BF2 tries to deny the constraints that the world imposes

on us centripetally. But these are pathological forms of selfhood precisely because they are exploiting the disintegration in an effort to deny or overcome disintegration.

Consider, for instance, the disintegration implicit in the very structure of faith as such.[13] Sartre's example is faith in Pierre's friendship. This faith involves, on the one hand, a thetic perception of Pierre as friendly. Friendship thus shows up immediately as a property intrinsic to Pierre. Insofar as faith involves this positional consciousness, to have faith in Pierre's friendship is for Pierre to show up as an implicitly trustworthy being. My faith in Pierre's friendship is manifest not in reflexive thought, but in my motivations and dispositions to rely on Pierre, to stand by him, defend him, etc. 'If I *believe* my friend Pierre likes me,' Sartre explains, 'I will see his friendliness as the meaning of all his actions. Belief is a distinctive consciousness *of the meaning* of Pierre's actions' (Sartre 1976: 104/115).

At the same time, faith in Pierre's friendship involves a non-thetic recognition that exercising faith in Pierre is something I *do*; that nothing necessitates my treating him as a friend; and that I can never be absolutely, apodictically certain of Pierre's friendship for me. Sartre explains:

> I *believe* means both: I am persuaded of it – and – I simply believe it. I believe in Pierre's friendship. This means, at the same time, that 'I would rather be cut to pieces than to think that he is not my friend' and that 'I am not certain of his friendship.' Hence if I solemnly tell him, 'I believe in your friendship,' I immediately give rise to the counterposition in myself, 'I am not sure.'
>
> (Sartre 1983: 492/475)

Both aspects are essential to faith. If it lacked an immediate (and thus largely passive) thetic perception of Pierre as objectively embodying his friendship for me, it might be possible to form the *belief* that Pierre is my friend. But my attitude toward his friendship would not be *faith*. When I have faith in Pierre's friendship, Pierre shows up immediately as trustworthy and supportive – he is genuinely delighted to spend time with me; we are able to talk openly about trials and shared concerns, etc. – and I experience my faith in him as a necessary response to the evidence of his friendship.

But nor would it be faith if it lacked the (normally non-thetic) consciousness that my faith in Pierre's friendship is something I *do* – that I reveal and perpetuate my faith in him by acting in the light of his friendship. I become aware of my faith as an action in certain situations – perhaps when I choose to depend on Pierre to keep his promises even in circumstances where I know that most people do not keep their promises. The very experience of needing to act to perpetuate our friendship shows that my faith in his friendship is not merely an immediate response to objective properties he possesses. We can thus see that the person in good faith (i.e., BF1) denies that being Pierre's friend involves experiencing Pierre and his actions in the light of my own project of being his friend, while the person in BF2 denies that her project of being Pierre's friend depends on the way Pierre shows up to me.

Sartre generalizes this account of a disintegration, and argues that all significant human beliefs, affects, and actions involve just such a twofold disintegrating structure. If my beliefs, affects, and actions were merely something I will, then there would be no responsiveness to them, and no facticity to constrain them. There would be no epistemic norms to govern these attitudes, no sense in which they are normatively responsive to the state of things. Conversely, if my beliefs, affects, and actions were merely passively produced in me, in a way that is independent of my will, then there would be no taking responsibility for them, and no room to surpass them creatively.

Thus, to act in good faith is a bad faith effort to eliminate the tension and the instability produced by the disintegrating structure of the self. The real antithesis to bad faith, then, cannot be good faith. Instead, Sartre tells us in a footnote to *Being and Nothingness* that a radical escape from bad faith is only possible through authenticity, a project which 'presupposes that decomposed being repossesses itself' (Sartre 1976: 117/106; translation modified). Authenticity, in other words, faces up to the decomposition in human existence and embraces it not as an obstacle to authentic selfhood, but as an enabling condition of it.

For instance, Sartre tells us, an authentic faith in Pierre's friendship 'would be to maintain the tension' insofar as 'faith becomes an act of willing and acting at the same time that it is aware of its limits' (Sartre 1983: 492/476). By

actively being a friend to Pierre, I give significance to our shared form of life. But the willfulness of this undertaking is tempered by my recognition of the 'limits' to my ability to define that significance on my own. There must be a factical basis to sustain me in acting, and I must act in such a way that the significance can continue to disclose itself. Friendship becomes, from this perspective, something both active and passive in turns – 'an intentional choice to do something (to make a friendship) and, from this perspective, to allow each moment its concrete development' (Sartre 1983: 492/476). This is an example of the 'valid coordination' between facticity and transcendence that Sartre insists must be possible.

If selfhood calls us to take responsibility for ourselves, authentic selfhood will involve neither a flight from nor a denial of the disintegration of human existence and the ambiguity that it brings. Because of the disintegration, it is never possible to say once and for all who I am, what my actions mean, what I ultimately want, and what is at stake in my decisions. Good faith proceeds as if there could be an objective answer to all these questions. But there only could be an objective answer if human beings had the being of the in-itself. However, since human selfhood necessarily takes the form of a disintegrating circuit of ipseity – the constant tension between our transcendence and what is given in the world – it is impossible for us to be in-ourselves. Our discussion of good faith shows that to be an authentic self in the Sartrean sense cannot mean that one possesses a determinate and clear identity, fitting neatly into a fixed set of categories. Instead, to borrow from de Beauvoir's *Ethics of Ambiguity*, to be authentically ourselves we have to find a way to live with 'the painfulness of the indefinite questioning' (de Beauvoir 1948: 133–4). Being human means to be involved in a constant and restless activity of defining ourselves in response to the world, without ever reaching a final verdict on who we are.

## ORCID

*Mark A. Wrathall* https://orcid.org/0000-0001-6140-9799

## ENDNOTES

[1] References to Sartre will include first the page number to the French text, followed by the page number to the English translation.

[2] Not all of these commentators endorse good faith to the same degree. Webber, for example, recognises good faith as a form of bad faith. Nonetheless, he suggests that among the different possible attitudes of bad faith, good faith stands out as normatively superior to the others (Webber 2009: 94). McNulty treats sincerity as a form of bad faith, but holds that 'good faith is facing the reality of our situation' (McNulty 2025: 159).

[3] Zheng (2005) provides a detailed discussion of good faith. While he suggests that 'bad faith distorts reality more severely than good faith does, and accordingly, it is bad in comparison with good faith', he ultimately argues that 'they both fundamentally involve "corrupted" modes of being' and that 'only in a meager and relative sense should we prefer good faith to bad faith ethically' (Zheng 2005: 70). Our classification of bad-faith attitudes differs from Zheng's and we hold that good faith is not even trivially better than bad faith.

[4] We agree with Manser (1987) and Bell (1989) that there is no normative difference between bad and good faith, but they do not discuss the distinction in much detail. Bell points out that 'Paradoxically, even good faith is for Sartre a matter of bad faith' and that 'both good and bad faith play off one part of the definition of man against the other. Both try to introduce into the being of man a principle that is straightforwardly applicable only to the in-itself—the principle of identity' (Bell 1989: 42, 45). Manser notes: 'There is nothing to choose between sincerity and *mauvaise foi*,' and 'Sartre makes it clear that the role of the example is to show that the structures of the two are fundamentally the same, that both are attempts to create a gulf within the human being between facticity and transcendence' (Manser 1987: 62).

[5] Our thanks to an anonymous referee for proposing this possibility.

[6] We suspect that McCulloch's proposal is best construed as positing two different restricted quantifiers.

[7] Davenport has recently proposed a three-fold distinction that might be more successful in modeling Sartre's discussion. In addition to being$_t$ (for transcendent being) and being$_f$ (for factical being), Davenport proposes being$_w$ (for 'trying to "be" *wholly* something or other'). See Davenport 2024: 444.

[8] Our solution, like Morris', draws on Sartre's distinction between two different kinds of relations and two different kinds of negation. Our approach can be viewed as a friendly amendment to Morris's approach, although we draw this distinction somewhat differently than Morris does. The key difference is that Morris believes it is unnecessary to disambiguate

between different forms of predication – that all the necessary work can be done by observing that 'the concept of truth does not apply straightforwardly to ambiguous reality' (Morris 1997: 477). This goes a considerable way toward explaining the paradox, since claims made about an ambiguous reality can be true in one sense and not true in another. But there are different senses in which a claim can be untrue about an ambiguous reality. Take for instance Sartre's observation that the waiter is in one sense 'not a waiter'; he is in quite another sense 'not a diplomat' (see Sartre 1976: 119/134). Both claims are about an 'ambiguous reality,' so the difference between them cannot be down to that. We can easily make sense of the distinction, however, if we see that 'I am not a diplomat' (when uttered by the waiter) is an external negation, while 'I am not a waiter' is an internal negation. He is not a diplomat because he at present possesses no properties – no skill, no training, no dispositions, no official status – that would allow him to determine himself as a diplomat. He is not a waiter in a different sense: even though he possesses all the characteristics that allow him to be a waiter, and even though he takes up the role of a waiter, as a human reality he is not ultimately determinable in terms of that social role.

[9] Facticity can be thought of as an objective fact, but one that has been taken up into our understanding and given significance for our form of existence: 'My facticity is only an indication that I present to myself of the being I am obliged to keep company with, in order to be what I am. It is impossible to grasp it in its brute nakedness, because we only encounter it when we have already reclaimed and freely constructed it' (Sartre 1976: 199/134).

[10] Sartre 1966: 28/10, quoting Husserl 1976: 200/221.

[11] We will focus primarily on Sartre's treatment of good faith in *Being and Nothingness*. Sartre mentions good faith in other works from the period such as *Existentialism is a Humanism* and *The Notebooks for an Ethics*. While we think that the treatment of good faith in these works is consistent with our reading, we do not have the space to discuss these passages here.

[12] Santoni acknowledges that Sartre's account of good faith in *Being and Nothingness* states that good faith and bad faith share the same ideal, but he argues that a positive reconstruction of good faith is possible. He admits that 'To many readers of Sartre, this account of good faith as an ontological attitude may sound strikingly similar to what many regard as Sartre's conception of *authenticity*' (Santoni 1995: 87). The key distinction between authenticity and good faith, for Santoni, is that good faith is a spontaneous, pre-reflective attitude, whereas authenticity is a reflective conversion: 'good faith prereflectively accepts the ambiguity of consciousness' (Santoni 1995: 80).

[13] Sartre identifies a number of different dimensions of disintegration in human existence. One source of disintegration is found in the way that the meanings of my actions are fixed both by my intentions, and by the reactions of others. What my action means for me is not the same as what it means for the other, but each aspect has 'equal dignity' in defining the action (see Sartre 1976: 92/101). Another source of disintegration is the way that the meaning of an action depends both on the agent's intention, but also on the change the action produces in the factical situation (see Sartre 1976: 552/660–1).

## BIBLIOGRAPHY

Bell, L. (1989), *Sartre's Ethics of Authenticity*. Tuscaloosa: The University of Alabama Press.

de Beauvoir, S. (1948), *The Ethics of Ambiguity*. New York: Citadel Press.

Catalano, J. S. (1980), *A Commentary on Jean-Paul Sartre's 'Being and Nothingness'*. Chicago: University of Chicago Press.

Davenport, J. J. (2024), "Sartre and Frankfurt: Bad faith as evidence for three levels of volitional consciousness," *European Journal of Philosophy 32*: 432–458.

Gordon, L. R. (1999), *Bad Faith and Antiblack Racism*. Amherst: Humanity Books.

Husserl, E. (1976). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Erstes Buch, Husserliana volume III/1*. Den Haag: Martinus Nijhoff.

Husserl, E. (1984), *Logische Untersuchungen*, vol. 2, *Husserliana* vol. XIX/1. New York: Springer Science+Business Media.

Manser, A. (1987), *Sartre: An Investigation of Some Major Themes*. Aldershot: Avebury.

McCulloch, G. (1994), *Using Sartre: An Analytical Introduction to Early Sartrean Themes*. London: Routledge.

McNulty, J. (2025), "Bad faith as true contradiction: On the dialetheist interpretation of Sartre," *Philosophy and Phenomenological Research 110*: 150–171.

Morris, K. J. (1997), "Ambiguity and Bad Faith," *American Catholic Philosophical Quarterly LXX*: 467–484.

Morris, K. J. (2008), *Sartre*. Oxford: Wiley-Blackwell.

Santoni, R. (1995), *Bad Faith, Good Faith, and Authenticity in Sartre's Early Philosophy*. Philadelphia: Temple University Press.

Sartre, J-P. (1945), *L'Age de Raison. Paris: Gallimard, translated as* The Age of Reason, *trans*. Eric Sutton. New York: Vintage International 1992.

Sartre, J-P. (1966), *La Transcendance de L'Ego: Esquisse d'une Description Phénoménologique. Paris: Librairie Philosophique, translated as* The Transcendence of the Ego: A Sketch for a Phenomenological Description, *trans. Andrew Brown*. London: Routledge, 2004.

Sartre, J-P. (1976), *L'Être et le Néant*, Collection Tel edition. *Paris: Gallimard, translated as* Being and Nothingness, *trans. Sarah Richmond*. Abingdon, Oxon: Routledge, 2018.

Sartre, J-P. (1983), *Cahiers pour une Morale. Paris: Éditions Gallimard, translated as* Notebooks for an Ethics, *trans. David Pellauer*. Chicago: Chicago University Press, 1992.

Sartre, J-P. (1995), *Carnets de la drôle de guerre: Septembre 1939 - Mars 1940*, Nouvelle Édition. *Paris: Éditions Gallimard, translated as* War Diaries: Notebooks from a Phoney War 1939-40, *trans. Quintin Hoare*. London: Verso, 1999.

Webber, J. (2009), *The Existentialism of Jean-Paul Sartre* (New York: Routledge).

Zheng, Y. (2005), *Ontology and Ethics in Sartre's Early Philosophy* (Lanham: Lexington Books).

**How to cite this article:** Wrathall, M. A., & von Knobelsdorff, W. (2026). 'A Perpetually Disintegrating Synthesis': Sartre on Bad Faith, Good Faith, and the Projects of Selfhood. *European Journal of Philosophy*, 1–20. https://doi.org/10.1111/ejop.70081