The Phenomenon of Bad Faith as Evidence for Three Orders of Identification in Volitional Consciousness

A Critique of Frankfurt and Sartre

Table of Contents

I. From Transcendental Apperception to the For-Itself

a. Review of the "I" of transcendental apperception in the *Critique* as pure subject;

b. If we think ofperson' as a substance concept, then indeed (as Hazel Barnes notes) we cannot treat Sartre's for-itself as a person. But it *is* related to the notion of person as pure Kantian subject. Response to Thomas Flynn on this point.

c. Difference from Husserl's and Kant's transcendental ego: relation to transphenomenal in-itself. Similarity to Kant's `Refutation of Idealism' argument (despite Sartre's disclaimer of this similarity). Mediation of Kant's RI argument through the Fichtean `Anstoss' (reference to Dan Breazeale's paper on the meaning of the Anstoss). Relation to thecircuit of selfness' in Sartre's "Immediate Structures of the For-itself"

d. The "non-personal consciousness" as the source of desire vs the "Ego" or Psyche in Sartre

e. The modal extension of the for-itself as the basic manifestation of freedom. Thomas Flynn's analysis of the freedom of consciousness in Sartre (transcendental vs. `noetic' freedom). Relation to the analysis of "the being of possibilities" section of Sartre's "Immediate Structures of the For-itself"

f. Critique of Kant and Sartre on the reflexive immediate self-awareness of consciousness (p.5-30 in *Being and Nothingness*): the reflexivity of consciousness is a necessary but not sufficient condition for the *unification* achieved in judgement. The synthesis achieved in thought has to be attributed to freedom of the will, rather than to the reflexivity of consciousness—refer to Frankfurt's fruitful discussion of this distinction in his "Identification and Wholeheartedness"

g. Conclusion: there must in fact be *two* immediate, non-cognitive relations of consciousness to itself, if it is to be a transcendental subject or forensic `person' (not `person' in the psyche/substance sense): 1. the reflexivity of consciousness which is also present in animal consciousness, and 2. the *higher* reflexive relation of a transphenomenal freedom to itself, which is what adds *mineness* and the unity of an `I' to reflexive consciousness. I will call consciousness with this two-fold reflexivity *volitional consciousness*.

II. 'Bad Faith' vs `Wantonness'

(A) Inward and Outward-Looking Negations

Jean-Paul Sartre begins his famous chapter on "Bad Faith" in *Being and Nothingness* by noting that a human being "can take negative attitudes towards himself" (p.86). He contrasts this immediately with "the nihilation of a possibility which another human reality projects as *its* possibility" (p.86), or the relation in which one person enslaves or coerces another: [1] in this case, the negative attitude is taken by one person in relation to another. Sartre notes, with an obvious allusion to Kierkegaard, that irony is similar although it involves more "inwardness of consciousness" (p.87): it is an *outward-looking* negation, in which the person herself negates what she is communicating, but also to some extent communicates that very negation. [2] "Self-negation" is different in that it is *inward-looking*. Although forms of self-negation are "diverse," Sartre chooses for his phenomenological analysis a type of self-negation he claims is "essential to human reality:" namely bad faith or "mauvaise foi."

After making this shift, however, Sartre immediately returns to another kind of essentially outward-looking negative agency, [3] the kind for which bad faith is most easily mistaken: namely, *lying*. Sartre emphasizes the importance of this distinction: we can describe bad faith as "a lie to oneself" only if "we distinguish the lie to oneself from lying in general" (p.87). Lying to others is essentially different than bad faith, because "the inner disposition of the liar is positive" and "rests on a truth" which the liar knows and intentionally misrepresents: "The liar intends to deceive" (p.88). Harry Frankfurt essentially agrees with this Sartrian account. [4] In an essay called "On Bullshit," which is more serious than its title might suggest, Frankfurt notes that if the liar does not utter a

---

[1]Frankfurt's definition of coercion agrees with Sartre here. For Frankfurt a person is coerced when their actions *or the desires on which they act* are made by some external force to violate the will of their own higher-order volitions. As we also saw, my willing to act on motive Y involves envisioning Y as a first-order possibility in my personal world: the agent of coercion who makes me act on motive X instead thus nihilates a *possibility of mine*, a possibility projected by my own higher-order will as the one on which I want to act.

[2]In this respect, it would be interesting to contrast irony with simply *lying*: for the liar also inwardly denies what she is communicating to the would-be victim of the lie. The ironist appears to be different in that—in some forms of irony, at least—she is trying to communicate *both* the false representation and the truth at the same time, in a kind of double-meaning to different audiences. The intended victim will be fooled by the misrepresentation, while a third-party or another audience will grasp the truth. The liar, by contrast, does not intend to communicate the truth to anyone.

[3]That lying is "outward-looking" is what Sartre means when he says that "The lie is a behavior of transcendence" which is "also a normal phenomenon of what Heidegger calls the '*mit-sein*'" or being with other persons (p.88).

[4]*Except* for the references to Frankfurt's article "On Bullshit," all references herein to essays by Harry Frankfurt refer to page numbers of the reprinted versions in H. Frankfurt, *The Importance of What We Care About* (Cambridge University Press, 1988).

statement which he *knows* to be false, at least "he himself believes that the statement is false and intends by making it to deceive."[5] Frankfurt also points out that the whatever the liar explicitly misrepresents, implicitly he "must inevitably misrepresent his own state of mind" as well:

> ..someone who lies about how much money he has in his pocket both gives an account of the amount of money in his pocket and conveys that he believes this account. If the lie works, then its victim is twice deceived...[6]

Similarly, Sartre sums up his own interpretation of lying as follows: "The ideal description of the liar could be a cynical consciousness, affirming truth within itself, denying it in his words, and denying that negation as such" (p.87). This much already shows us that only a being capable of reflective consciousness and explicit awareness of its own psychic states of belief (as objects) could be capable of lying.

However, this "doubly negative attitude" of the liar (as Sartre calls it) applies not only to his misrepresentations, but also to his will: he must *will* to misrepresent to his listener the maxim on which he acts. Lying involves a kind of misrepresentation of one's own for-itself, or one's volitional *identification* in Frankfurt's sense. The liar's intention to deceive "explicitly exercises a regulatory control over all his attitudes" (p.88): this shows that the intention to lie is really constituted as a second-order volition to arrange the first-order desires expressed in one's actions in such a way that they combine to give the appearance of earnestness. But it is not just a higher-order intention to deceive: it is also an intention to deceive the victim about one's higher-order intentions themselves. As Sartre says, the liar plays the role of a character who intends to tell the truth—i.e. the role of someone who *identifies* with this intention, in Frankfurt's sense. But of course, this expressed second-order desire is "not recognized by the liar as *his* intention:" he is `wholeheartedly' (again, in Frankfurt's sense) identified with the will to act on the motive of misrepresenting both outward facts and his own authentic higher-order will. Therefore, lying is an action only possible for a *person* in Frankfurt's sense, i.e. a being with reflective consciousness capable of authentic identification with its own first-order maxims through second-order volitions.

(B) `Bullshit' and Wantonness as Outward-Looking Non-Thetic Attitudes

The same may be said for what Frankfurt calls "humbug"—another more subtle yet still

---

[5]Harry Frankfurt, "On Bullshit," *Raritan*, Vol. VI, No. 2 (Fall, 1986): p.83. An abridged version of this amusing and insightful essay was published in *Harper's Weekly*, February, 1987, and the full essay was reprinted in Frankfurt's *The Importance of What We Care About.* My references to this essay all cite the original *Raritan* version and pages.

[6]Frankfurt, "On Bullshit," p.83.

essentially `outward' form of misrepresentation—which Frankfurt distinguishes from lying. In lying, as we saw, the liar directly intends to deceive his audience about the putative facts which he misrepresents (even if these are facts about himself); in doing so, he indirectly intends (through the higher-order volition constitutive of lying) to misrepresent his own state of mind and his own will. By contrast, drawing on Max Black's *The Prevalence of Humbug,* Frankfurt portrays "humbug" as a sort of `deficient' type of lying, or an action "short of lying," where the speaker directly intends to misrepresent her own state of mind and will, although in doing so she may correctly represent what she takes to be the facts relevant to her topic. Frankfurt's example is a bombastic Fourth of July orator whose statements about the greatness of the our country and the divine inspiration of the Founding Fathers are not intended as lies:

> But the orator does not really care what his audience thinks about the Founding Fathers, or about the role of the deity in our country's history, or the like...what makes the Fourth of July oration humbug is not fundamentally that the speaker regards his statements as false. Rather, just as Black's account suggests, the orator intends these statements to convey a certain impression of himself...He wants them to think of him as a patriot, as someone who has deep thoughts and feelings about the origins and mission of our country.. [7]

The humbug directly intends to misrepresent his own concerns, feelings, and will, although he does not explicitly *lie* about these: rather, he misrepresents them indirectly through communicative actions the direct content of which need not be lies at all. In fact, Frankfurt's model of the will affords a rather precise formulation of what "humbug" in this sense is: [8]

> Person P acts as a "humbug" if and only if, in some (truthful) communicative action A, P does not identify with the *cares* C which A suggests that P has, but rather P identifies only with the will to misrepresent herself as caring about C.

Frankfurt's interest in humbug is that it provides a kind of simulacrum for the state of the will which is essential to "bullshit." Like humbug, Frankfurt argues that what makes statements bullshit has nothing directly to do with the truth-value their maker takes them to have. But bullshit aims at misrepresenting one's own will only in one general respect: namely as including the concern for truth generally expected in both strategic and cooperative types of communicative action. To illustrate this, Frankfurt cites a report according to which Wittgenstein once criticized a woman who said she felt as bad as "a dog that has been run over." As Frankfurt speculates, Wittgenstein must have been

---

[7]Frankfurt, "On Bullshit," p.85-6.

[8]And in this respect, I am surprised that Frankfurt did not have recourse to his own hierarchy model of the will to give such a precise definition of humbug in the paper.

piqued because her statement is "mindless" in a special sort of way: it is "not germane to the enterprise of describing reality." [9] The problem, as Frankfurt is quick to recognize, is not that her statement is an inadvertent slip in a serious attempt to describe her current state to Wittgenstein: rather it is flippant in the sense that "she is not concerned with the truth-value of what she says." As Frankfurt says,

> That is why she cannot be regarded as lying: for she does not presume that she knows the truth [about what a run-over dog feels like], and therefore she cannot be deliberately promulgating a proposition that she presumes to be false. Her statement is grounded neither in a belief that it is true nor, as a lie must be, in a belief that it is not true. It is just this lack of connection to concern with truth—this indifference to how things really are—that I regard as the essence of bullshit. [10]

Bullshitting, we might say, consists in communicating without *caring* about the truth at all, either flippantly, or to appeal to what listeners want to hear or what will move them, or for other reasons, including sheer perversity: "The fact about himself which the bullshitter hides..is that the truth-values of his statements are of no central interest to him." [11]

But what is really interesting about bullshitting, in fact, is that although it is *outward-looking* in the Sartrian sense, it is not really a `thetic' *negation* in any sense at all. Perhaps the bullshitter hides his lack of care about the truth, but in the most authentic and revealing form of bullshit, he doesn't even care about hiding his lack of concern for veracity.

On the strength of his analysis, Frankfurt arrives at two interesting conclusions: (1) that "bullshit is a greater enemy of the truth than lies are" and (2) that "the contemporary proliferation of bullshit" has its deepest sources in the modern attachment to anti-realist skepticism which erodes our faith that we have "any reliable access to objective reality." [12] Frankfurt notes that as a result, an "ideal of *sincerity"* or honest representation of oneself has begun to replace the ideal of truth. But he concludes with the Kantian point that "As conscious beings, we exist only in response to other things," and that "facts about ourselves" are no more objectively accessible to us than facts about the external world: hence "sincerity itself is bullshit." [13]

---

[9]Frankfurt, ibid, p.88-89.

[10]Frankfurt, ibid, p.90. As he adds later, bullshit is "phony" in the sense that "although it is produced without concern for the truth, it need not be false." It is thus analogous to a phony product, which may work as well as the real one, but was produced illegitimately (p.94).

[11]Frankfurt, ibid, p.97.

[12]Frankfurt, ibid, p.98-99.

[13]Frankfurt, ibid, p.100.

At this point, there is an immediate temptation to say that Sartrian "bad faith" consists in a kind of Frankfurtian "bullshitting" *of oneself*, and to compare Sartre's own critique of "sincerity" as itself in bad faith to Frankfurt's.[14] Nevertheless, the temptation to draw immediate comparisons must be resisted, for the relation between "bullshit" in Frankfurt's sense and "bad faith" in Sartre's sense is more complex than it might first appear.[15]

In fact, "bullshit" in Frankfurt's sense is more closely related to his earlier description of *wantonness* than to bad faith in Sartre's sense. We might say that Frankfurt's "bullshitter" is *wanton* with respect to the truth, since he or she has neither the second-order volitions of the liar nor of someone who cares about truthfulness.[16] What wantonness has in common with bullshit in its truest form is that neither is an active, `thetic' negation affected by consciousness. Recall that in his famous contrast between a "wanton" and a "person" in the sense of someone capable of moral responsibility, Frankfurt argued that the latter must have what the former lacks (by definition): namely, higher-order volitions (or *cares*, in the later versions of the theory) through which he or she is *identified* with first-order states of will:

> The essential characteristic of a wanton is that he does not care about his will. His desires move him to do certain things, without its being true of him that he wants to be moved by those desires or that he prefers to be moved by other desires. The class of wantons includes all nonhuman animals that have desires and all very young children. Perhaps it also includes some adult human beings as well [such as the addict who is neither willing nor unwilling]. In any case, adult human beings may...act wantonly, in response to first-order desires concerning which they have no volitions of the second order, more or less frequently.[17]

"Wantonness" in this sense has a certain superficial similarity to at least some types of Sartrian bad faith. Since the wanton acts by "the liberty of anarchic impulsive behavior" rather than in "the autonomy of being under his own control," which requires authoritative identification with a

---

[14]After reading "On Bullshit," it is at least hard not to suspect that Frankfurt has been reading his Sartre.

[15]As we will see, what Sartre and Frankfurt mean by "sincerity" is not precisely the same, either.

[16]I suspect that Frankfurt's original interest in bullshit arises from its close connection to the problem of wantonness which he identified "Freedom of the Will and the Concept of a Person." In that 1971 essay, Frankfurt allowed that the wanton could deliberate rationally about his desires. But in his 1987 essay "Identification and Wholeheartedness," Frankfurt changes his position, and acknowledges that the wanton cannot engage in deliberation, because "reasoning involves making decisions concerning what to think" (p.176). I think it is the analysis of bullshit which has prompted this change. What this analysis shows is that bullshit is wantonness with respect to the truth, and that therefore any *complete* wanton must also be a bullshitter, which entails that they lack an essential interest in veracity necessary even for instrumental deliberation.

[17]Frankfurt, "Freedom of the Will and the Concept of a Person," p.17.

coherent set of one's affective states,[18] the wanton seems to `live in his facticity' in much the same way as Sartre's famous waiter who tries to "be immediately a café waiter in the sense that this inkwell *is* an inkwell" (p.102). The waiter tries avoid realizing that he performs the actions required to be a café waiter freely, "separated *by nothing*" from the role, the "imaginary café waiter" he aims at (p.187).

And yet, Sartre's waiter is *not* a wanton in Frankfurt's sense, because to be a wanton literally is to be a non-person for Frankfurt, a being wholly summed up by its being-in-itself in Sartre's sense. Here Sartre sees something that Frankfurt's account leaves out: a `wanton waiter'[19] is what the waiter *wants to believe* that he is, but in his bad faith, he cannot actually succeed in being wanton. By definition, wantonness is a *state* of the human being which can only be appreciated from the *third-person* point of view: for Frankfurt, the wanton could not *identify with* her wantonness, and hence it cannot have a first-personal significance for her. Wantonness, as we see, shares with bullshit not only the lack of thetic negativity, but also *outward-lookingness.* It is not inward-looking in the way bad faith is, because wantonness is not an *intrapersonal structure affected by `consciousness' [or transcendental `personality'] itself.* By contrast, Sartre sees that bad faith is not a "state" (p.89), i.e. not a `state of mind' or an aspect of the "Ego" in sense of a psychic object—or what I have elsewhere called the `public self' of actions and attributable first-order desires, preferences, and motivations.[20] Rather, bad faith is actually a *project*—or `care' in Frankfurt's sense—of the transcendental `person' or "consciousness" itself:

> ...consciousness affects itself with bad faith. There must be an original intention and a project of bad faith; this project implies a comprehension of bad faith as such and a pre-reflective apprehension (of) consciousness as affecting itself with bad faith (p.89).

This is the apparent paradox of bad faith which Sartre sets out to explain. As an account of this undeniable phenomenon of the human will, Frankfurt's notion of "wantonness" would be as inadequate as the Freudian conception of "the censor" which Sartre so mercilessly reduces to absurdity in the first section of the chapter on "Bad Faith" (p.90-94).

(C) The Central Problematic of Wantonness vs Bad Faith

The problem with Frankfurt's account of wantonness is that he remains indecisive on whether

---

[18]Frankfurt, "Identification and Wholeheartedness," p.175.

[19]Read this in the same way as Frankfurt's famous *wanton addict.*

[20]"The Person as Will rather than Mind," unpublished.

it is the actual lack of second-order volitions, or the *impossibility* of such volitions for the animate being in question, that makes it wanton. If the former, and Frankfurt usually implies, then a "person" as contrasted with a "wanton" is being who happens to have volitional identifications, rather than a being with the *capacity* to identify. Sartre's analysis of bad faith brings this problem to the surface because it raises the possibility that leading a dissolute life, or having no persisting second-order volitions to authorize and (to the degree possible) to control one's first-order desires, may itself be an identifiable "style of life" (p.90) as Sartre says, which is *tacitly* chosen by the person.[21] What exactly `tacit' might mean in this context will emerge as we go on.[22] But even unexplicated, the hypothesis is enough to show that for any adult human being past the age of reason who *appears* to be a "wanton" in Frankfurt's sense, it is possible that that human being is actually in bad faith, and so may be a "person" in the forensic sense after all. Rather than indicating their inability to identify, their outward wantonness may be a sign of their unwillingness to use their *capacity* for authentic identification.

Indeed, it is a problem for Frankfurt's theory that it implies the possibility that some human beings who have developed the cognitive equipment necessary for forming higher-order volitions, and who may be thought of as *capable* of forming such volitions, may simply lack them and thereby literally fail to be *persons* in the forensic sense.[23] Our intuitions suggest that personhood is not a contingent `property' in this sense; one cannot just accidentally have or fail to have it. Traditionally, the quickest way to accomodate this intuition is to adopt a substance conception of personhood:

---

[21]Kierkegaard's deliberate aesthete in *Either/Or* Vol.I is clearly a relevant example here.

[22]As Sartre suggests in arguing against the Freudian account of bad faith as "a lie without a liar" (p.92), the pleasure and anguish which accompany the "symbolic and conscious satisfaction" of an allegedly unconscious drive are impossible "if consciousness does not include—beyond the censor—an obscure comprehension of the end to be attained as simultaneously desired and forbidden" (p.94). Such an "obscure comprehension" comes close to what I mean by a *tacit* identification. The difference is that I explicitly regard the tacit "comprehension," which allows for the unity of the phenomena Sartre and Freud are concerned with, as a *volitional* relation which is not only pre-thetic, but also not equivalent to any mode of the basic reflexivity of consciousness (see §I of the paper). Moreover, against the Kantian tradition, I regard such a tacit volitional relation as the bearer of a sort of pre-conscious 'expectation' or tendency, which is nevertheless not a mere impulse or involuntary drive.

[23]This has led more than one commentator to assume that Frankfurt's theory is "normative" in the sense that it makes distinctions between persons in the `metaphysical' sense who live up to an ideal of fully autonomous personhood, and those who do not. In that reading, "person" in Frankfurt's terms would be a synonym for "autonomous." But that is not the right interpretation, because for Frankfurt, "persons" can authentically identify with the most immoral maxims for action, and can do so on the motive of pure desire, in fact. Frankfurt uses "person" in a *constitutive* sense *within* the practical: "persons" have the identifications necessary to be more or less "fully" persons in whatever ethical or normative sense we use for evaluating integration over a whole life. Life-plan normative conceptions of "person" are derivative from and dependent on a more fundamental *non-evaluative*, *constititive* forensic characterization of personhood. Such a concept of personhood is neither a non-practical `metaphysical' theory, nor an evaluative theory of morally autonomous or ethically integrated personhood.

personhood becomes a property one has essentially because what *makes one* a person at all is identified with what gives one self-identity over time. But this is a confusion Frankfurt, like the existentialists, rightly resists. The condition of personhood (or *ipseity*) are prior to the conditions of its identity over time—otherwise, we have no reason to believe that the latter establishes conditions of *personal* identity over time, rather than the identity of something non-personal over time. Existentialism offers a different explanation for the intuition that personhood is not a contingent property: personhood is not a *sosein* at all, i.e. not an essence that is *contrasted* with its multiple contingent instantiations; rather, each individual *is* personhood, which consists in a freedom for (or openness to) multiple possibilities for one's moral identity or character. In its freedom, existing personhood has the `modal extension' otherwise reserved to essences, but this openness itself already has the concrete individuality of an instantiation of a species-kind: it is "in each case mine," as Heidegger puts it.[24] Thus personhood is an *existential*: it cannot be *had* contingently by a bearer of properties, because to have it is rather to *be* it, i.e. for personhood to be the actuality that exists to bear properties in the first place. Thus, although the substance account and existentialist account of personhood are in diametric opposition, they at least agree that personhood cannot be had contingently. For Frankfurt to deny this is therefore philosophically precarious at best.

Sartre's theory of bad faith depends, as we have seen, on such an existentialist conception of transcendental `personhood.' As a result, Sartre's view implies not only that any apparently adult rational wanton may be in bad faith instead, but rather that *no* adult human being who is capable of the kinds of volitions essential to moral personhood is entirely without them, however he might appear. Sartre's theory implies that when we are confronted with an apparently "wanton" adult human addict, who is perhaps even quite a good calculator of what he needs to do to support his habit, our intuition that this is a *person* is stronger than our impression that he entirely lacks the volitional *projection* unavoidable for a "for-itself": rather than conclude that he is not a transcendental person, we conclude that deep within himself, he at least tacitly *identifies* with his current situation—that he is playing at being a `wanton,' which he cannot actually *be*. Otherwise we could only conclude that when an apparently wanton addict becomes either a *willing* addict identified with his desires, or an unwilling addict committed to fighting them (whether successfully or not), that change is wholly *external* to the free projection of the transcendental person: Frankfurt's theory

[24]This is my own rather elaborate way of explaining how freedom has essentially to do with the Heideggerian notion that Dasein's essence is its `existence,' in Heidegger's sense of that term.

implies that such changes are alterations in which a forensic person *comes into being*, not alterations by which the forensic person *affects themselves*. And in the span of human experience, we know this is not true: a theory which implies that it is, is itself in bad faith.

A neutral critic might justly respond that in that case, since the notion of "wantonness" proves to be invalid, the full paradox evident in bad faith returns. To do better than Frankfurt, Sartre will have to explain how the paradox of bad faith is possible in a way that *also* preserves the essential *insight* in Frankfurt's contrast between "persons" and "wantons:" namely, the insight that through higher-order volitions, it is possible to authentically identify or alienate one's own first-order desires, and that this capacity is essential to personhood. I will argue that by this standard, Sartre's own account of bad faith fails, because Sartre is led to the conclusion that authentic identification with and/or alienation of certain Ego-states (i.e. first-order desires) through the higher-order volitions implicit in the "consciousness (of)" the for-itself is impossible. As it turns out, the problem of Sartre's denial of authentic *alienation* as a possible *intrapersonal* structure is even clearer than the problem with his equivalent denial of the possibility of authentic intrapersonal *identification*. For a sense that something must be wrong with Sartre's account arises almost inevitably (for all but the most hard-nosed Sartians) when we realize that Sartre's model appears to have no room for *unwilling* addicts in Frankfurt's sense—i.e. for people who *authentically alienate* a first-order desire or impulse which they are nevertheless moved by. For Sartre, any such attempt to deny that the addictive desire represents who one wants to be, in the deepest sense, must immediately and with gross lack of charity be diagnosed as *denial* in the bad sense—as bad faith. Sartre has no way to allow that one may *ever* actually be *coerced* by desires and impulses on which one nevertheless acts. In this, Sartre stands alone, against not only the accepted `minimal psychological model' of Western common law and precedent, but against the testimony of human experience in general.

Sartre's model has no room for authentic (rather than bad-faith) alienation of one's own operative desires and impulses because he has no room for authentic *identification*.[25] The cause of this more basic problem, as I will try to show, is that Sartre confuses the fact that the transcendental "for-itself" is always free to alter its original project with an impossibility of *being* anything through this project. It turns out that Sartre's analysis of bad faith in terms of essential structures of *consciousness* is not sufficiently discriminating to allow us to distinguish the for-itself's *authentic*

[25]In Frankfurt's scheme, authentic alienation of a first-order desire D1 depends on authentic identification with a different (ideal) first-order will D2, that rules out D1. Thus alienation is not only defined in terms of identification; the former actually depends on the latter in an ontological sense.

identification with aspects of its Ego from bad faith in which the for-itself attempts to "be" these aspects of its factical Ego. When we have shown that Sartre's "patterns" of bad faith can be explained by hierarchical structures of volition which serve to explain Sartre's own diagnoses in the process, the distinction between authentic identification/alienation and bad faith will finally become clear.

This analysis has the virtue that it allows us to understand clearly why Sartre's theory of "bad Faith" itself tends to gives rise to conflicting intuitions. What is usually perplexing to the sympathetic reader in the famous chapters on "Bad Faith" is that: (1) it *seems right* or `rings true' that people actually do consent to the roles they wish to treat as essential to themselves, and they are actually responsible for actions they may want to pass off as mere events, such as their `happening' to leave their hand in another's caress; but (2) it does not seem right that no one could or ever has acted as a truly unwilling addict, or been compelled by imminent threat of death to act on a motive of fear which they honestly and authentically abhor. At this point, Sartre has gone too far with his Kantian refusal to see that `ought-implies-can' need not always be applied in the fashion of *modes ponens,* but can also sometimes have a *modes tolens* application: `*cannot*' sometimes does implies lack of culpability, even with respect to one's own `inner' springs of action. The reader's conflicting intuitions result from the fact that Sartre is *right* to deny that people can be wantons, but *wrong* to deny that they are ever authentically alienated from (or identified with) a role they have knowingly played or an affective aspect of their psyche which has resulted in performances on their part.

Thus both Frankfurt and Sartre are wrong. Frankfurt's belief that "It is possible for a human being to be at times, and perhaps even always, indifferent to his own motives" [26] is dubious, because it excuses from moral responsibility and forensic personhood certain apparently `wanton' adult human beings who in fact have no excuse. We cannot believe that these so-called `wantons' themselves played no part in their failure to activate their capacity to form higher-order volitions, or to take an "evaluative attitude" towards their own motives. Once we take the step—which Frankfurt has been unwilling to take—and accept that one must have some degree of negative liberty in the *higher-orders of the will* in order to even be capable of identification through higher-order volitions, it becomes obvious that no one is really wanton unless they are *totally incapable* of the cognitive and

[26]Frankfurt, "Identification and Wholeheartedness," p.164.

volitional acts required for higher-order willing (e.g. the animal or the infant).[27] Sartre's account fully acknowledges this reality of human existence, but it goes to the other extreme: Sartre's treatment of all putative cases of authentic alienation as instances of bad faith denies to persons who really do have an excuse that excuse to which they are entitled.

And both Sartre and Frankfurt are right, because each has *a part* of the truth. Harry Frankfurt has initiated the first successful phenomenological account of how authentic identification is possible and how moral responsibility not only for actions but also for *motives* is constituted—but in doing so, he implies that adults capable of the kind of *intrapersonal volitional identification* required for full moral responsibility, whom we naturally think of as persons, may be "wantons" who are accidentally without forensic personhood. In attempting the first sustained diagnoses of the crucial phenomenon of bad faith, Sartre sees why wantonness in Frankfurt's sense is impossible for a being capable of the kind of higher-order volitions Frankfurt describes, but in the process Sartre convinces himself that authentic identification will always be false "sincerity," a form of bad faith, while the notion of `authentic alienation' between the subject and its psyche becomes the form of bad faith in which one hides in one's transcendence. In other words, Sartre sees why for human consciousness, wantonness is always a disguise of bad faith, but in doing so, he confuses certain other important volitional possibilities with types of bad faith. Frankfurt, on the other hand, has room for authentic identification and alienation of one's own first-order affective and preferential states, but no room for bad faith. The truth is that the *same source* from which bad faith originates can also be the source of intrapersonal authentic identification and alienation. My analysis will make clear not only why these two structures are *different*, but how they both originate from the for-itself.

III. A Volitional Interpretation of Sartre's Patterns of Bad Faith

(A) The Coquette's Love of Ambivalence

Although the example of the "young woman" who is in bad faith about her date's sexual interest in her—traditionally called the `coquette'[28]—is the first Sartre introduces, out of all the `cases studies' in the chapter it actually presents the most complex pattern of bad faith. Recall that

---

[27]To this extent, we can see that what is dubious in Frankfurt's characterization of the wanton is entirely a symptom his resistance to any 'principle of alternative possibilities' for moral responsibility—even if such a principle applies to the causation of higher-order volitions rather than to the causation of one's actions.

[28]I am uncomfortable with this term, because its implications seem more than a little bit sexist, or at least insensitive. For that matter, Sartre's "paederast" seems similarly insensitive. Nevertheless, I have chosen to keep these terms for fear that my squeamishness itself might be in bad faith.

Frankfurt distinguishes the wanton who has no real identity or second-order self from a person who is not "wholehearted" or who authentically identifies with conflicting second-order volitions. Frankfurt describes the latter type of person as "*ambivalent*," because "there is no unequivocal answer to the question of what the person really wants."[29] Similarly, Sartre says of the coquette that "she does not quite know what she wants," because on the one hand she looks to her companion for a feeling of respect for her freedom, and on the other hand she looks for a suggestion of bodily desire, since "she would find no charm in a respect which would be only respect" (p.97).

However, this case is filled which complexities which Sartre's compressed description does not arrange in clear order. Two different *first-order* motives seem to be involved, the description of which is complicated by their reference to her companion's affective states. Moreover, both these motives or desires affect how the woman regards her companion and herself. I will describe all three aspects of both the relevant first-order motives.

Her first desire (D1) is the desire *not to* be desired sexually but to be liked purely as a friend: "the desire cruel and naked would humiliate and horrify her" (p.97). Notice that although D1 involves an *iteration in its description*, it is not a second-order volition in the relevant sense, i.e. the sense of being a movement of *identification* with a first-order desire. Rather, it is only a complex first-order preference, or if you will, an *action*-maxim with an `end' that refers to other desire-states. Since her companion's behavior betrays suggestions of his desire, it is D1 which motivates the coquette to regard him as "a thing" without desires and *consciousness* in Sartre's sense—since to acknowledge that he has any conscious desires would unavoidably mean recognizing the sexual ones. So "She has disarmed the actions of her companion by reducing them to being only what they are; that is, to existing in the mode of the in-itself" (p.97). Finally, this desire D1 makes no sense unless the woman acknowledges herself *as body*—as that which in D1 she wills not to be threatened by the sexual desires of another. D1 thus reduces both parties to their facticity.

The second desire (D2) is the desire *to* be desired sexually by the other: "at the same time this feeling [of her companion] must be wholly desire;[30] that is, it must address itself to her body, as object" (p.97). The `*as object*' is important in this passage because it tells us how the coquette regards herself in D2, and consequently how, when moved by D2, she can accept that her companion has desires. Although D1 requires her to reduce him to a thing without desires, in D2

---

[29]Frankfurt, "Identification and Wholeheartedness," p.165.

[30]Sartre of course is using *desire* to mean 'sexual desire,' while I have been using the term desire the way Frankfurt uses it, namely to stand for all sorts of non-cognitive impulses and preferences.

> ... she permits herself to enjoy his desire, to the extent that she will apprehend it as not being what it is, will recognize its transcendence. Finally, while sensing profoundly the presence of her own body...she realizes herself as *not being* her own body and contemplates it as though from above, as a passive object to which things can happen but which can neither provoke them nor avoid them (p.97-8).

So D2 moves the woman to regard her companion's desire not merely as the brutal impulse of a beast but as a free approval of an affection, i.e. as a second-order volition rather than a mere first-order impulse. But because Sartre is not thinking in terms of *orders* of will which allow for authentic intrapersonal identification, he thinks of the coquette, when motivated by D2, as *falsely* turning the man's desire into a kind of "transcendence," a free state of consciousness which *is not* the desire it intends. Moreover, the woman can appreciate this more complex `affection' for her body, because she again deceives herself by abstracting from the very body that the other is desiring: his desire is directed at an object that has nothing to do *with her self.* Sartre thus conceives D2 as falsely motivating the coquette to regard both parties as purely being their transcendence, or `not being what they are' in the terminology of *Being and Nothingness.* [31]

Given this interpretation of the coquette's conflicting desires, what is her *attitude towards* the conflict of D1 and D2? It is the answer to this question which will tell us what constitutes her bad faith. For bad faith cannot be simply the existence of conflicting first-order desires, however complex their reference to the desires of another. A person can openly acknowledge such a conflict with no self-deception. Bad faith is an *inward-looking* deception of some kind, and so it must involve some *intrapersonal* relation. Moreover, as we saw, to be a deception it must be a relation *effected* by the for-itself, or transcendental `person': so it must be an intrapersonal relation of the *volition*, and not simply of modes of conscious intentionality (such as, for example, *doubting* a belief or judging about a desire). But then, since bad faith must be an intrapersonal volitional relation, and it cannot be simply the relation of one desire to another within a *single order* of the will—e.g. the *mere existence* of a conflict between D1 and D2—it follows necessarily that bad faith must be an intrapersonal relation between volitions and/or desires of *different orders* of identification.

Yet none of the relations of this kind which Frankfurt has previously distinguished seem to constitute the coquette's *bad faith* with respect to D1 and D2. This woman is difficult to classify according to Frankfurt's criteria, because she seems to occupy a strange intermediate status between being `wanton' and being `ambivalent.' We cannot say that her bad faith consists in simply *lacking*

---

[31]however, we may say that D2 is really the coquette's desire for the other to have a certain complex of second-order and first-order motives with regard not to her, but 'only' to her body.

any higher-order volitions with respect to D1 and D2. If she were simply wanton, presumably one of these two desires would win out, or if their strengths *happened* to be perfectly balanced, she would just vacillate back and forth, flirting madly one moment and then backing off the next. But in neither case would she be involved in trying to postpone the point at which either D1 or D2 won out, or to preserve both in an uneasy symbiosis. The situation is clearly more complex than a merely *wanton* relation to D1 and D2 allows. The coquette is not entirely passive in her volitional relation to D1 and D2—as Sartre says, she is effecting a *project* of bad faith.

Recalling the distinction between wantonness and the authentic conflict of true *ambivalence*, could we then say that the coquette's bad faith consists in identifying with two opposed second-order volitions, $V_2$1 and $V_2$2, which approve of acting on D1 and D2, respectively? Not really, because Frankfurt treats ambivalence as an explicitly recognized state, which the person can only resolve through "a radical separation of the competing desires".[32] To be ambivalent rather than wanton for Frankfurt, it seems to be necessary to have *full-blown* second-order volitions that aim at different first-order desires as candidates for the role of being one's motive for acting.[33] Frankfurt also describes ambivalence as a hinderance: "The disunity of an ambivalent person's will prevents him from effectively pursuing and satisfactorily attaining his goals."[34] But this is certainly not the case for Sartre's coquette, who is pursuing her goal very effectively by affecting herself with bad faith. But most importantly, we cannot regard the coquette as openly ambivalent in Frankfurt's sense, because she precisely *does not* want to acknowledge either motive as a desire she *freely identifies* with.[35] She is not `authoritatively choosing' volitions that approve two conflicting first-order wills,

---

[32]Frankfurt, "Identification and Wholeheartedness," p.170. A conflict among second-order volitions is equivalent to what Charles Taylor usually calls a "strong opposition" that cannot be resolved by *ranking* the opposed volitions in a preference-ordering. The choice between such "strongly opposed" options is "strong evaluation" or identification, in Frankfurt's sense.

[33]In his 1991 Presidential Address to the American Philosophical Association, entitled "The Faintest Passion," Harry Frankfurt describes ambivalence as a state where the two "conflicting volitional movements" are "both *wholly* internal to a person's will rather than alien to him; that is, he is not passive with respect to them" (p.8—my italics). The example he gives is from Augustine's *Confessions*, where Augustine remarks that "it is...no strange phenomenon, partly to will to do something and partly to will not to do it" (p.9). [Eastern Division Meeting, Dec. 29, 1991].

[34]ibid, p.9.

[35]In his Presidential Address, Frankfurt himself suggests that "self-deception" is always a strategy someone takes to conceal from himself some aspect of his personality that he is not satisfied with—an aspect that "he cannot wholeheartedly accept." It follows that self-deception is an attempt to *escape* ambivalence (p.15). If we took bad faith as self-deception in *this sense*, then it would immediately follow that bad faith is not ambivalence but an attempt to escape it. However, as we will see, bad faith cannot be "self-deception" in Frankfurt's sense, since Frankfurt argues that self-deception is always motivated by the *necessary* desire to be wholehearted (p.14). Bad faith will turn out to be a counterexample to Frankfurt's rather irenic hypothesis that "no one can be wholeheartedly ambivalent" (p.14).

but rather trying to deny that she has two different desires she must decide between at all. When her companion takes her hand, she finds that

> To leave the hand there is to consent in herself to flirt, to engage herself. To withdraw it is to break the troubled and unstable harmony which gives the hour its charm. The aim is to postpone the moment of decision as long as possible. We know what happens next: the young woman leaves her hand there, but she does not notice that she is leaving it. (p.97)

Sartre's description shows that it is crucial for the young woman's project that she not authentically identify with either D1 or D2. Thus there is an important difference between honest ambivalence and bad faith. Nor can the woman's bad faith be described as merely repressed ambivalence, for her effort is precisely not to acknowledge the kinds of volitions necessary for ambivalence. She is not directly active in authorizing her conflicting first-order desires, in the way genuine ambivalence requires. But nor is she simply passive or `wanton' with respect to those conflicting desires. *What is she, then?*

(B) Sartre's Paradigmatic Structures of Bad Faith and Authenticity

Before I give my own explanation, I wish to consider Sartre's. We still know what we already concluded: that bad faith must be an intrapersonal relation between volitions and/or desires of different orders. Without knowing any more than this, we can infer that if there were *no differentiation* between different `orders' of volition—or (more accurately) between the kinds of intrapersonal `identification' created by different orders of volitional `consciousness'—then bad faith could not occur at all. This is the real meaning of Sartre's dictum that "If man is what he is, bad faith is forever impossible" (p.101). The question is, precisely *how* is man `not what he is?' Do we need only one division to account for bad faith, or must further relevant divisions of order be posited within the volitional consciousness to account for bad faith?

It is important to approach the issue this way, because it locates a point of broad agreement that for bad faith to be possible, it must be true that, in Sartre's poetic terms, "the principle of identity must not represent a constitutive principle of human reality and human reality must not be necessarily what it is. What it is must be able to be what it is not" (p.101). But all that this playfully oxymoronic statement really means is that there must be *some* ordinal division within the volitional consciousness that is "human reality." The problem is that it does not tell us *precisely* what this division must be.

This transcendental point about the *general* conditions for the possibility of bad faith is therefore to be distinguished from *specifically* analyzing bad faith in terms of a *two-orders*

hypothesis, which is what Sartre actually does. Sartre's general principle that volitional consciousness must *in some way* not be `what it is,' if bad faith is to be possible, *does not*, contrary to the impression Sartre gives, really entail *his specific* two-tiered approach to explaining it.[36]

Sartre's argument that bad faith essentially "utilizes the double property of the human being, who is at once a *facticity* and a *transcendence*" (p.98) can be read as referring to such a relation between volitions of different orders. In this relation, facticity corresponds to desires that are immediately operative in actions (including bodily impulses), while the transcendence of the for-itself, which is *wholly identified* with its original project, corresponds to a negative freedom for volition of *some* higher-order which is (a) unique, (b) incapable of supporting simultaneous but inconsistent `original projects' of this order. Hence we may say that the for-itself corresponds to the order of the will which is necessarily (c) the origin of all `identification,' fully and immediately identified with itself. This is true for the for-itself even though in the anguish of its freedom it exists as the pre-thetic awareness of the permanent possibility of changing its project: it cannot give up its original project without replacing it with a new one, which is all the substance it has, and its anguish consists precisely in the paradox that this original project with which it is *fully* and completely identified is still not immutable for it. In Frankfurtian terms, we might say that the for-itself is the highest-order level of the will—a level which is necessarily wholehearted, but always an *anxious wholeheartedness*.[37]

Sartre never explicitly defines bad faith in terms of a single relation between facticity and transcendence (in their volitional forms), but he does suggest that different `misalignments' between them, all sharing certain features, account for the varieties of bad faith. Although facticity and transcendence "are and ought to be capable of valid coordination,"

> ..bad faith does not wish either to coordinate them or to surmount them in a synthesis. Bad faith seeks to affirm their identity while preserving their differences. It must affirm facticity as *being* transcendence and transcendence as *being* facticity... (p.98).

---

[36]My own argument will be that the three-tiered approach gives us a much more adequate basis for explaining Sartre's own case studies.

[37]Precisely because he has several objections to this Sartrian theory of the origin of identification, Frankfurt has recently tried to develop a different (but probably inadequate) account of the origin of identification which ultimately depends on what he calls "satisfaction." His doubts about Sartre's approach, which motivate his own pursuit of an alternative account, have mainly to do with its apparent implications (1) that the 'personally possible' volitions which make up one's 'personal world' would be limitless, and (2) that total arbitrariness would result in the history of one's volitions, which would make cares and volitional necessities impossible. But explaining how to accommodate the insights in this Frankfurtian critique of Sartre is a *corrective* project [carried out *fully* in the paper on Molinism] that should occur only once we have established that the for-itself, as a third-order will, is the origin of full identification. At the same time, I postpone my critique of Frankfurt's own theory of satisfaction-based full or complete identification.

To understand these twin formulae for bad faith, it is crucial to realize that they are always intended to contrast not just with one another, but more fundamentally with what Sartre takes to be the true relation between transcendence and facticity: namely, *being what I am* in my facticity, acknowledging my first-order character, but *in the mode of not being it*, or acknowledging also that I cannot subsist in any first-order state of myself, since I transcend it and have the power to change it. For Sartre, `authenticity' can only mean acknowledging that transcendence and facticity are related in *just this* twofold way: the "valid coordination" of the two means openly affirming this *inescapable truth* about the relation of transcendence to its facticity. I must not try to live wholly in the one or the other, since in fact I am a particular irreducible relation between them. Every type of bad faith, by contrast, attempts to deny or escape from precisely this twofold fact: namely, that I am inescapably `identified' (in some unclarified sense) with my first-order self (or Ego), although I can never be *fully identified* with it in Frankfurt's sense, since I `am it' *only* in the mode of `not-being it,' or having a free higher-order of the will which transcends it.

This conception of the single possible authentic relation of transcendence to facticity appears to work well up to a point: it gives Sartre two basic forms of bad faith that can be understood fairly easily. Since my actual existence is a "double property" of being transcendence and facticity, each in different ways that are mutually interdependent, every type of bad faith, which is essentially a denial of what I am, will involve a double error. Thus, affirming "transcendence as being facticity" (as Sartre puts it) means trying to live *wholly* in my transcendence: "I flee from myself, I escape myself" or deny my facticity, when I affirm that "what I really am is my transcendence." But in doing so, Sartre says rather cryptically, I must "affirm here that I *am* my transcendence in the mode of being a thing" (p.99). Why does this follow, especially when I am trying to escape facticity? Because transcendence lives only by negation of the in-itself: when I try to subsist entirely as for-itself without acknowledging first-order desires and affects as *mine* in any sense, then the negation collapses and the freedom I am trying to live wholly within turns into a thing. The man who "in the face of reproaches or rancor dissociates himself from his past by insisting on his freedom and on his perpetual re-creation" (p.100) also seems to fall into this category. In general, I will indicate this type of bad faith which tries to escape into transcendence as $T \rightarrow F$ (to indicate the inevitable `facticizing' of transcendence it involves).

The other paradigm of bad faith is that exemplified most clearly in Sartre's second famous illustration: the waiter. The human being in this case is trying to *"be* a waiter," or to be wholly

identical with the set of motives for action which society thinks should be automatic for a waiter. Sartre suggests the `factical' character of the role in which he is trying to lose himself, by describing this waiter as "chaining his movements as if they were mechanisms" and giving to himself "the pitiless rapidity of things" (p.101-102). These are again largely *poetic* descriptions, which serve to suggest the *chthonic* associations of the in-itself in his actions, their disturbing approximation to `lifeless mechanism' (to use Erich Fromm's symbol for the `necrophilous'). Sartre thus suggests how unnatural to human existence he finds such reduction to pure *functional role*. By enforcing such reduction through its expectations, modern society thus exercises a *perverting* influence on human beings. For example, a "tradesman" such as a grocer or a soldier finds that "society demands that he limit himself to his function as a grocer, just as the soldier at attention makes himself a solider-thing..." (p.102). These striking passages and the powerful critique they imply show us not *wanton* grocers, waiters, and soldiers, but rather tradespeople deliberately *trying* not to have the kinds of personal cares for their jobs or anything else which would put them `above' mere outward behavior in accordance with predictable impulses.

In this type of case, the result is the opposite of what we found in *T→F* bad faith. Because the transcendental `I' cannot *just be* a waiter, my attempt to lose myself in this facticity leads to my being the waiter "only in *representation*," or in the freedom of what Kant calls reproductive imagination. Thus "I cannot be he," i.e. this complex of first-order motives and behaviors, but rather "I can only play at *being* him; that is, imagine to myself that I am he" (p.103). The attempt to realize "the being-in-itself of a café waiter" thus fails, because the fact that the whole performance is supported by the free projection of the for-itself, which confers the value or first-personal significance that my rights and duties as a waiter have for me, creeps back into my active instantiation of the role. It becomes apparent that I *doth protest too much* that I am only a waiter. The transcendence trying to take cover under the guise of a waiter-machine is revealed, because *no disguise* of this type can ever perfectly cover it. Thus we may refer to this paradigm of bad faith by the locution *F→T* (to indicate that a facticity becomes shot through with transcendence). Sartre's example of the man who "arrests himself at one period in his life" (p.100) and tries to identify himself wholly with *being* that person, also fits this F→T paradigm.

The results Sartre expects in each of his two straightforward types of bad faith, T→F and F→T, thus simply follow the logic inherent in his nearly-Kantian conception of the authentic relation of the for-itself to the factical `self,' i.e. to the Ego of theoretical `inner sense' and practical *arbitrium sed sensitivum*. In both types of bad faith, the proper `authentic relation' in Sartre's own sense can only be

restored by acknowledging that one really is_1 (at the first-order level) what one is not_2 (at the higher-order level). Sartre indicates what the waiter would think if he were authentic as follows:

> Yet there is no doubt that I *am* in a sense a café waiter—otherwise could I not just as well call myself a diplomat or a reporter? But if I am one, this can not be in the mode of being in-itself. I am a waiter in the mode of *being what I am not* (p.103).

Hence the first formula for authenticity **(A1): I am_1 what I am_2 not**.

Authenticity can *equivalently* be put the other way around: **(A2) I am_2 not (only) what I am_1**. Yet this is *not* the same thing as wholly denying one's facticity, which is a form of bad faith. The difference between this second formulation of authenticity and bad faith becomes apparent in Sartre's discussion of the paederast, who *rightly* (in Sartre's view) recognizes that homosexuality as a complex of actions and first-order motives or attitudes is not "a destiny," because "a homosexual is not a homosexual as this table is a table" (p.107). And Sartre explicitly says that the "undeniable comprehension of truth" in this realization would have led to authenticity rather than bad faith if the paederast had simply used it to *qualify* his homosexuality with the point that "human reality cannot be finally defined by patterns of conduct" (p.108). Sartre's way of putting the point is again poetically compact:

> He would be right actually if he understood the phrase "I am not a paederast" in the sense of "I am not what I am."....But instead he slides surreptitiously towards a different connotation of the word "being." He understands "not being" in the sense of "not-being-in-itself." He lays claim to "not being a paederast" in the sense in which this table *is not* an inkwell. He is in bad faith (p.108).

To understand the distinction at issue here, we have to translate it as follows. The homosexual would have been authentic had he aimed at the view that `I am not_2 (only) what I am_1.' But rather, he is in bad faith because he tries to lay claim to not $being_1$ what in fact he is_1.

This, however, is a negative formulation of the T→F form of bad faith. Moreover, it is not entirely adequate, because it suggests that one could overcome this form of bad faith by admitting simply that one is_1 what one is_1. This formula should actually stand for that *part* of authenticity which Sartre indicates when he says that I am in a sense a waiter, since I am not the diplomat, etc. But by itself, however, this formula would be misrepresentative, because it does not indicate the *whole* of the human reality.[38] The two equivalent formulae A1 and A2 count as formulae for Sartrian authenticity because they (supposedly!) capture the *whole reality* of volitional

[38]Moreover, on its surface, I am_1 what I am_1 looks too much like the *bad faith* of the waiter, who claims that he coincides entirely with his waitering. We have yet to show how it actually differs from the structure of his bad faith.

consciousness, constituted by the immediate relation of a transcendental for-itself to the facticity of a transphenomenal in-itself. To give us an opposition commensurable with authenticity, then, the formulae for bad faith must likewise indicate the *misplacement* of the *whole being* (which I will indicate with the subscript `w' for convenience). Like the F→T form illustrated by the waiter, the paederast's bad faith is really an attempt to locate the *whole* of his existence in one pole of the authentic transcendence-facticity relation of reflexive, volitional consciousness. Thus we should say that he *wholly* denies that he is a paederast, because he is trying to locate the whole of himself in his transcendence, i.e. he is trying "to take refuge in a sphere where one is no longer anything but a pure, free regard" (p.110). The formula for the paederast is thus as follows:

BF2[39]: I am_w *not* what I am_1, because I am_w my $not\text{-}being_2$ anything that I am_1.

This is the *structure* of the bad faith whose inevitable *upshot* is indicated in the pattern T→F. Using this same notation, we can now also state the problem with the waiter in a way that makes clear why his bad faith is not precisely the same as simply acknowledging that he is_1 a waiter, as opposed to being a grocer. The formula for the waiter is:

BF1: I am_w what I am_1, because I am_w *not* my $not\text{-}being_2$ anything that I am_1.

The result of this form of bad faith is the paradigm I have labelled F→T.

Other equivalents for A1 and A2, then, are the denials of BF1 and BF2 respectively. I can restate A1 negatively using the `w' subscript by saying: I am_1 what I am not_w. Similarly, A2 becomes I am not_w what I am_1. This says that I am not *wholly* what I am, or that I do not completely coincide with my factical (or first-order) self. Note that is *not* equivalent to the type of bad faith which says that I am *wholly not* my first-order volitional consciousness, or that I *wholly fail* to coincide with it, which is BF2 again: I am_w *not* what I am_1. Note also that BF2 and BF1 are *not* equivalent, although Sartre does not emphasize this point. They are two distinct ways of diverging from what is really a single form of authenticity. Although we have distinguished A1 and A2 to clarify the contrast with BF1 and BF2 respectively, A1 and A2 really are two ways of saying the same thing:

[39]I have labelled this BF**2** because it is the perversion of the second formulation of authenticity, i.e. A2. Similarly, BF1, which comes on the next page, is the perversion of A1.

A0: I am not$_w$ either what I am$_1$ or (my not-being$_2$ anything I am$_1$).

Sartre's own analysis of authenticity suggests that what I am$_w$ corresponds to neither of his two orders of volitional consciousness. Note, however, that we have no positive way to state what my reality *does* fully correspond to: in other words, there is no available positive reformulation of A0 in Sartre's system. Although Sartre would take this as evidence for his view that human reality is an *irreducibly* negative phenomenon, it is always a sign of trouble when a philosopher is forced to say that the most basic possible analysis of an ontological structure is a purely negative formulation. As Husserl knew well, this means we do not have a fully adequate analysis, however valuable the formulation in hand. [40] My own diagnosis of the problem is that a positive statement of what I am$_w$, or what *full identification* in Frankfurt's sense originates in, will only become possible when we make a further distinction in the orders of volition involved in human consciousness.

(C) The Question of Sincerity

However unwieldy these formulations using subscripts may look, they are congenial not only because they help us understand what Sartre actually took the contrasts between bad faith and authenticity to be, but also because without them, the even more complex locutions which Sartre introduces towards the end of the chapter become completely confusing. For example, in critiquing sincerity, Sartre argues that the goal of the sincerity norm is

> To bring me to confess to myself what I am in order that I may finally coincide with my being: in a word, to cause myself to be, in the mode of the in-itself, what I am in the mode of "not being what I am" (p.110).

This takes some decoding. The last part of this passage is a reference to the factical self of outward behavior patterns, and habits and impulses of the first-order will: "what I am$_1$ in the mode of not being$_2$ what I am$_1$." In other words, this is an extended or fully formal way of referring to one's facticity which recognizes that, *authentically* speaking, I am it only "in the mode of not being what I am" (A2). Sartre interprets sincerity as the demand that I *coincide* with this facticity (thus formally designated), which means that I should lay claim to being$_w$ what I am$_1$. But if the norm of sincerity

---

[40]We reached the same result with Frankfurt when it became evident that wholehearted or complete 'identification' could not consist in simply having second-order volitions, since (to put it in our new locution) the authentically ambivalent person is not$_w$ either of the second-order wills they are$_2$. [See the Personal Will paper, §IV]. I am still postponing discussion of Frankfurt's own solution, which is that the person is$_w$ any coherent set of their second-order volitions and 'volitional necessities' which they rest *satisfied* with.

cannot mean anything else but this—an interpretation that I will challenge in the next section—then of course sincerity is *immediately* equivalent to a form of bad faith—namely, BF1. This is all that really underlies Sartre's apparently stunning conclusion that "the essential structure of sincerity does not differ from that of bad faith..." (p.109). Sartre's belief that sincerity = BF1 is also the reason why he goes straight into the example of the waiter, who illustrates BF1, after introducing the question of sincerity. Finally, this also explains why Sartre pits the "champion of sincerity" against the paederast as interlocutors. In this case study, the paederast resists the demand of sincerity by way of focusing on the irreducibility of his transcendence, which precisely militates against BF1, but in the process he falls into the opposite error, namely BF2. Thus, given his own interpretation of sincerity, Sartre is able to ask ironically, "Who is in bad faith? The homosexual or the champion of sincerity?" (p.107). The irony is that neither is more or less in bad faith than the other, since they represent BF2 and BF1 respectively.

Sartre's conclusion on sincerity is actually part of a larger argument that "in order for bad faith to be possible, sincerity itself must be in bad faith" (p.112)—i.e. I must not in fact be_w anything that I am_1 or even anything that *could* be_1. This makes sense of his point in the difficult discussion of courage and cowardice. Sartre argues that it is only because of the negativity which prevents me from coinciding wholly with the cowardice I actually display (in the habitual first-order will on which I act), that I can even think of escaping into my transcendence:

> The condition under which I can attempt an effort in bad faith is that in one sense, I *am not* this coward which I do not wish to be. But if I *were not* cowardly in the simple mode of not-being-what-one-is-not, I would be "in good faith" in declaring that I am not cowardly (p.111).

To translate, Sartre's hypothesis is that it is only because in fact I am not_w what I am_1 (a version of A2) that it is even possible in the first place to deceive myself into believing that I am_w not what I am_1. The structure of the authentic relation is the ground of possibility for bad faith, because it is a specific way of fulfilling the most primordial necessary condition that volitional consciousness not consist of a single order. Thus Sartre concludes this middle section of the chapter by saying that "the condition of the possibility for bad faith is that human reality...in the intra-structure of the pre-reflective cogito, must be what it is not and not be what it is" (p.112). This means that bad faith is possible because we are what the formula for authenticity says we are: we are_1 what we are not_2 and we are not_2 what we are_1.

In the passage on cowardice just quoted, Sartre also points out that there is a "simple mode" of not $being_1$ what one is not_1, but I cannot claim that I "am not" a coward in that sense. So why,

then, does sincerity have to consist in a demand that I admit myself to be$_w$ what I am$_1$? Why, instead, doesn't sincerity simply consist in admitting that I am$_1$ the coward that I am$_1$—which is true, but not a type of bad faith? The answer is that this recognition, although it would not be *inauthentic*, can never constitute a recognition of the *whole* of what I am. It could be the sincere truth for a *wanton* coward, but this I can never be, because I am always is some way not$_2$ what I am$_1$. Sincerity is a requirement that relates to what I am$_w$—it is the demand that has to do with my full or complete `identification,' in Frankfurt's sense.

But what precisely does sincerity require of me relative to what I am$_w$? Sartre introduces sincerity as the "demand" that "it is necessary that a man be *for himself* only what he *is*," and complains immediately "But is this not precisely the definition of the in-itself..?" (p.101). But as we have seen, sincerity cannot simply be the demand for honesty about what I am$_1$, if Sartre's own critique of it is to make any sense. So he argues that "if candor or sincerity is a universal value" it "posits not merely an ideal of knowing but an ideal of being...that we *make ourselves* what we are" (p.101). Here we see the point at which Sartre's conception of sincerity reaches a deeper level than Frankfurt's, which remains an ideal of personal *honesty* (that will sometimes replace the ideal of truth). Sartre's analysis suggests that sincerity must be what Frankfurt would call an *ideal of complete identification*, or `integration' of oneself, so that one *is* wholeheartedly identified with the cares and projects that actually do express one's `true self.' But Sartre realizes that this would be a vacuous demand if there were not more than one relevant `order' of selfhood or identification in volitional consciousness: "what *are* we then if we have this constant obligation to *make ourselves* what we are?" (p.101). As he sees, the demand "supposes that I am not originally what I am" (p.105-6)—i.e. I am not$_2$ what I am$_1$.

This leads to what we may call a *metaprinciple* for sincerity, which Sartre at least implicitly sanctions. Sincerity is a demand for a kind of action rather than merely some cognitive form of honesty about oneself; in that case, it can only be meaningful if it is a demand that I be$_j$ what I already am$_i$, where *i* and *j* stand for distinct orders of volitional identification, and $i < j$.[41] This is a general principle, but Sartre is basically pursuing a two-tiered hypothesis in his analysis of phenomena of human volitional consciousness. So for Sartre, in effect this metaprinciple implies

[41]In fact, it is clear that if sincerity is to be defined this way in terms of existential *identification*, and the definition is to be more than vacuous, then there must be more than one order of volitional identification *above the minimal identification* that exists for every personally actual and personally possible conscious state 'of mine.' This requires that we distinguish at least three orders of volitional identification.

that sincerity requires that I be_2 what I am_1—which turns out to be a formula for bad faith (namely, BF1). Thus Sartre asks, "I...make up my mind to be my true self without delay" but "what does this mean if not that I am constituting myself as a thing?" (p.106). Sincerity as an existential demand can *only* mean this for Sartre, not because it is the only possible interpretation of sincerity that conforms to our metaprinciple, but because Sartre works with what he takes to be the *only two* relevant orders of volitional consciousness. It is thus a symptom of Sartre's two-tiered approach that sincerity *must* be a form of bad faith for him. As Sartre himself says, since "consciousness is not what it is" (p.105) and can never become wholly what it is, the "ideal of sincerity" is a "task impossible to achieve" (p.105).[42]

But that does not make sincerity superfluous for him. Rather, Sartre suggests that sincerity always has an ulterior *purpose*: the ultimate goal of the demand for sincerity is precisely to effect the F→T modulation that results from BF1. The "champion of sincerity" is thus "not ignorant of the transcendence of human reality," when he confronts the Other with this (insincere) demand for sincerity, for "the critic asks that the man be what he is in order no longer to be what he is" (p.108). This means that "the critic demands of the guilty one," in this case the homosexual, that he be what he is_1, so that in trying to reduce himself wholly to this facticity, his own freedom from it will force its way to the surface, as it did with the waiter. Similarly, Sartre analyzes a `sincere' confession of evil motives as a project of bad faith with the same kind of ulterior intention:

> The man who confesses that he is evil has exchanged his disturbing "freedom for evil" for an inanimate character of evil; he *is* evil, he clings to himself, this is what he is. But by the same stroke, he escapes from that *thing*...in confessing it, I posit my freedom in respect to it; my future is virgin; everything is allowed to me (p.109).

This is presumably Sartre's critique of the hypocrisy so often evident in the Christian practice of confession—but it is now extended to a *ressentiment*-style claim that there is *necessarily* an ulterior motive of bad faith in any such action, that it is impossible *ever* to confess any motive that was freely incorporated without intending to deny precisely the identification with it which the structure of one's being actually makes it impossible to remove: so the denial will always be in bad faith.

Equivalently, Sartre also argues that `sincere' identification with something else than what I am_1 will always be in bad faith. To see this, let us return to his discussion of the coward who wants

---

[42]Of course, in saying this Sartre has in mind that sincerity is impossible because volitional consciousness must have *more than one level*. The irony is that Sartre is not considering the possibility that consciousness might be even *more* differentiated than his account implies. Thus it does not occur to him that perhaps it is not the plurality of orders in volitional consciousness, but only his *restriction* of this plurality to two levels, that makes sincerity impossible.

to be courageous. As we saw, this `unwilling coward' can apprehend himself in bad faith as "not being cowardly" (p.111) or as being *wholly* alien from his cowardice, only because it is true that he does not *entirely* coincide with being$_1$ cowardly: his BF2 builds on an authentic truth about the inevitable separation within volitional consciousness. But Sartre also adds that "bad faith is not restricted to denying the qualities which I possess" (in my first-order self or Ego); rather

> It attempts also to constitute myself as being what I am not. It apprehends me positively as courageous when I am not so. And that is possible, once again, only if I am what I am not; that is, if non-being in me does not have being even as non-being. Of course necessarily I *am not* courageous; otherwise bad faith would not be *bad* faith. But in addition my effort in bad faith must include the ontological comprehension that even in my usual being what I am, I am not it really and that there is no such difference between the being of "being sad," for example—which I am in the mode of not being what I am—and the "non-being" of not-being courageous which I wish to hide from myself. Moreover, it is particularly requisite that the very negation of being should be itself the object of perpetual nihilation. If I *were not* not courageous in the way in which this inkwell is not a table...if it were not on principle impossible for me to coincide with my *not-being-courageous* as well as with my being-courageous—then any project of bad faith would prohibited to me (p.111-112).

Let me go through this crucial passage very carefully. Sartre is trying to explain how the `unwilling' coward's (UC's) bad faith not only lets him alienate his cowardice, but also lets him *identify* with the courageous motives he wishes he had. Sartre acknowledges the presupposition that `UC' is not$_1$ courageous. But then he tries to emphasize the *difference* between this negative description of UC's first-order state, and the *general* fact that my authentic structure entails that I never coincide entirely with *what I am*$_1$: thus we should read him as meaning `even in my usual being what I am$_1$, I am not it wholly.' His example is that I am sad only "in the mode of not being what I am," which (as we have seen) is one of the formulae for the authentic relation: I am$_1$ sad in the mode of not$_w$ being what I am$_1$. Sartre says that there is no difference between this authentic relation and the "'non-being' of not-being courageous" which I actually exhibit and wish to hide from myself. The point here is that although UC is not$_1$ courageous, it is *also true* that UC is not$_w$ what he is not$_1$ (i.e. courageous) in the same way as he is not$_w$ what he is$_1$ (i.e. cowardly). So what is occurring when UC feels that in some sense he is courageous? Sartre's answer is that UC is making false use of the fact that he is not$_w$ the cowardice that he is$_1$. But rather than simply trying to escape wholly into transcendence, he is identifying with a courage which he does not have in any sense.

Sartre's explanation is that this step occurs through a kind of `negation of the negation.' But it is rather difficult to figure out precisely which negation UC is negating. The best interpretation seems to be that Sartre means UC recognizes that he is not$_w$ his not-being$_2$ what he is$_1$—i.e. that he

does not wholly coincide with the transcendence that stands out from his cowardly facticity. In itself, this is an authentic recognition (contained in A0) which is the antidote to BF2, the bad faith which wishes to escape into transcendence. So the final account of UC is as follows: UC flees from the facticity he actually is_1 into his transcendence, but then in a dialectical reversal, he negates the negation, or recognizes that he cannot reify his transcendence or coincide wholly with it, which he falsely takes as justification for identifying himself with *another* first-order state he doesn't actually have.

But if this were the case, then UC would be moving through *both* types of bad faith in one moment, proceeding dialectically from BF2, through the negation of the negation, to BF1, through which he finally loses himself in a facticity he does not even $possess_1$. This, it must be admitted, would be a remarkable feat indeed, for even Sartre's pre-reflective cogito does not make it possible to be in both types of bad faith at once. Moreover, it would still not be clear on this account how the unwilling coward, through his double-movement, manages to avoid simply regaining his first position, and losing himself in his *cowardice*. When I negate my cowardice by a leap into transcendence, but negate the negation itself and return to facticity, why should I end up in a different first-order state than the one I started with?

(D) A Deconstruction of Sartre's Explanation of `Bad Faith'

It should be clear by now that although Sartre's approach is admirably systematic, as I have shown, it begins to break down rather badly when it comes to the `sincere' confessor and the unwilling coward. Sartre's model seems to have a plausible explanation for how one tries to alienate the first-order motives one actually has, but he has no very persuasive explanation of how this alienation can occur *through identifying instead* with an ideal first-order state one wishes that one had. His explanation is itself paradoxical, since it suggests that such a person goes through *both paradigms* of bad faith in one movement: UC both escapes wholly to his transcendence in order to alienate his cowardice, but only then to escape wholly into a facticity he does not have? This contradicts the expectation we receive from the T→F paradigm, which predicts that UC should end up *reifying* his `not-$being_2$' cowardly, and try to $be_w$ this not-$being_2$. But instead, he negates this not-being in identifying with $being_1$ courageous. Moreover, how is it possible to attempt to $be_w$ `what one *would* be_1,' rather than what one is_1? Obviously, Sartre is grasping towards something like Frankfurt's notion of second-order volitional identification, but Sartre's own terms are not adequate to express this phenomenon: BF1 as Sartre has explained it does not cover `losing oneself in a

longed-for *counterfactual* facticity.' And the even more convulted attempt to capture this phenomenon in the idea that the unwilling cowward achieves his identification with courage in bad faith by going simultaneously through T→F and then F→T proves to be completely incoherent in Sartre's own terms.

The difficulty can be located more precisely. As we saw, Sartre naturally allows that "In one sense, I am not this coward which I do not wish to be" (p.111). This refers to the authentic relation which makes me stand out from my cowardice, rather than the bad faith which would alienate me *wholly* from it. But how can we interpret the *additional phrase* "which I do not wish to be?" Sartre would presumably take this to be part of the motive the for-itself has in its project of believing that the subject is *wholly not* the coward that he is in society? But then another question arises: isn't it possible *to wish not to be* this coward that I am_1, and yet not be engaged in the project of BF1? If one says this is not possible, one contravenes the common testimony of human experience, and a good justification for the claim that it is not possible should be forthcoming. Sartre does not address this question, and so he does not offer and such defeater for this important possibility. Given that it is possible for me to wish that I were not a coward *without* falling into the error of trying to escape wholly into my transcendence, then again we must ask in what this alienation of myself from the cowardice consists? Sartre cannot deny that it is *more* than my simply not $being_w$ that coward that I am_1—for that only indicates my freedom to not to retain the original project which supports this cowardly set of impulses and affections I act on. But if, by hypothesis, this wishing that I were not a coward is also *less* than bad faith, then it occupies an intermediate position between authentically not $being_w$ what I am_1 and deceptively fleeing into transcendence, $being_w$ not what I am_1.

Further consideration shows that *both* of Sartre's paradigms of bad faith, BF1 which results in F→T and BF2 which result in T→F, are inadequate to explain some of the cases to which they are supposed to apply. The "champion of sincerity" and `sincere' confessor are supposed to be like the waiter, engaged in BF1. But as presented, the waiter's real aim *just is* to escape into his facticity. The fact that he goes through the F→T modification—that his attempt to be a waiter-machine turns into `playing' a waiter—is *an inadvertent result which frustrates his intention.* Yet Sartre describes the `sincere' confessor as *using* BF1 precisely in order to realize his transcendence of the `evilness' which he admits to *$being_w$*, or in which he wallows. But in that case, his real aim is precisely to bring about the modulation which *frustrates* the waiter. In fact, the ulterior aim Sartre attributes to the sincere confessor is incompatible with BF1 as exhibited by the waiter. The real intention of the champion of sincerity, as Sartre says, is to make the paederast wallow in $being_w$ a paederast,

precisely in order to bring about liberation from this facticity. It is as if the champion of sincerity want to *trick* the paederast into bad faith of the F→T sort, to bring about an end the paederast would not be anticipating when engaging in the bad faith. But the `sincere' confessor is doing this *to himself*—so we cannot separate his bad faith from an end *external* to its motivation. The sincere confessor, then, must be in bad faith BF2 *about* his own attempt to be in bad faith BF1—which is absurd. If we wanted to bring the waiter into line with the champion of sincerity, we could argue that the waiter is also wallowing in waiterdom also just in order punish himself and force himself to free himself from his role: he would then be employing the entire F→T paradigm for the same type of ulterior motive as the sincere confessor. But this would still not remove the problem, because in both cases, it turns out that what constitutes the bad faith is *not* adequately described as trying to be_w what one is_1.

Thus Sartre can only assimilate the phenomenon of sincerity to bad faith by describing it as a project in which BF1 seems to give way to BF2. And this is something that Sartre's own analysis does not allow. Sartre's account makes clear how trying to live wholly in one's transcendence inevitably means trying to *be* it wholly or completely, in the full self-identity of a thing; it also makes clear why trying to immerse oneself wholly in one's roles or deny that one is anything more than certain of one's first-order volitional states inevitably means *playing* those roles or desires with an irrepressible free distance from them. By themselves, the bad faith that slips from purified transcendence to reified facticity (T→F), and the bad faith which slips from pure facticity to playful transcendence (F→T), both make sense. But these two slippages are *not the same* as `slipping' from the F→T complex as a whole to the T→F complex as a whole, or visa versa.[43] And yet this is the nearest we can come to construing the sincere confessor in Sartre's terms: he slips *from* F→T *to* T→F—which is not a possibility predicted by Sartre's paradigms.

If we now turn back to the coquette's case, we can see that it is also not adequately captured by either of Sartre's paradigms for bad faith, because it is just the mirror-image of the sincere confessor structure. Just as the sincere confessor is supposed to have the same sort of bad faith as the waiter, but turns out *not* to, the coquette is supposed to have the same sort of bad faith as the paederast, but turns out not to.

The paederast illustrates someone in BF2 whose intention in fleeing to his transcendence is to

---

[43]Sartre does not make this crucial distinction clear, because if he did, the full difficulty of the coquette's case would become all-too-apparent.

escape or alienate his first-order Ego or social self: his deepest real aim is thus *frustrated* when he finds that his transcendence itself becomes reified. But the coquette is not frustrated by the equivalent modification in her own case: rather, she seems to intend it. In desiring only to be liked as a friend, she flees from her own facticity, not noticing the hand she is leaving in her companion's caress. She does not notice this "because it happens by chance that at this moment she is all intellect" (p.97). But in the very process, she reifies her transcendence: "she shows herself in her essential aspect—a personality, a consciousness" (p.97). This seems to conform to the T→F type of bad faith. Yet at the same time, as Sartre acknowledges, when she "recognizes herself as *not being* her own body," her intention is not *just* to escape into her transcendence (which would, in fact, mean choosing D1 over D2), but rather to be able to enjoy the Other's desire for that body without endangering any part of herself in responsibility for that body. Thus her deepest aim in fleeing to her transcendence is really to be her facticity *as a thing*, rather than as something for which she must take responsibility. She is trying to separate her facticity and her transcendence, in order to be *both* in the manner of a thing. But in that case, she is the inverse of the man who `sincerely' confesses to evil motives he wishes to overcome. He flees to his facticity precisely in order to escape *wholly* into his transcendence (which, as we saw, is not really explicable as bad faith in either of Sartre's senses), and the coquette flees to her transcendence precisely in order that she should become a *thing* in her facticity—a body which is caressed without her being able to accede or refuse.

Thus we must say that, in affecting herself with BF2, she really intends to affect herself with BF1. This is how she aims to satisfy both of her contradictory desires at once—both being_w her transcendence and being_w her facticity. But this is not a form of *bad faith* we can adequately explain simply through two orders of volitional consciousness. How would we even begin to express its logic? We would have to say that the coquette tries to be_w her not-being_2 what she is_1, *in order to* be_w what she is_1, like a thing. But this makes no sense at all—there is no explanation here for how such a project could be anything other than *immediately* self-defeating, which the coquette's bad faith clearly is not. Our distinction between levels worked to remove the apparent contradictions in Sartre's playfully paradoxical formulae for A1, A2, BF1 and BF2. But here, two orders of subscripts are not enough to remove the contradiction.

IV. A New Theory of Bad Faith

(A) From Two to Three Orders of Volitional Consciousness

We are now in a position to say not only why Sartre incorrectly diagnoses the unwilling

coward and sincere confessor as in bad faith, but also what is wrong with his analysis of bad faith itself, as revealed in cases like the coquette's. The central problem is evident in the inadequacy of Sartre's conception of authenticity. As we have seen, Sartre's interpretation of authenticity as `being what I am_1 but only in the mode of not $being_2$ it' does not allow for any authentic identification with first-order desires that differentiates my first-order self into three segments: desires I affirm, desires I make alien to myself, and desires that I neither authorize nor alienate (e.g. desires that I am `wanton' with respect to). Rather, Sartre's way of carving up the volitional consciousness responsible for bad faith only allows for <u>one</u> non-deceptive intrapersonal relation of my transcendence to my facticity: acknowledging my factical self (Ego, first-order will) as my own, and yet acknowledging also my `negative' distance from it, or the negative liberty which my highest-order project has with respect to it. *Authenticity* is thus defined as *this* transcendental for-itself's acceptance of its facticity, without reducing itself to its factical characteristics.

It is clear that by definition, such Sartrian authentic acknowledgement of my factical desires, roles, and Ego-states must be the same for all my first-order states of consciousness and volition. Authentic recognition of these states as my own involves no *discrimination* between them, classifying some as *affirmed* and some as *rejected* in the further, `special sense' Frankfurt has brought to our attention. And yet by authentic recognition of one's factical characteristics, Sartre seems to have in mind something more than what I have previously called the *automatic* `minimal' identification I have with any of *my* actual or personally possible conscious states. Sartre's analysis of authenticity in terms of the transcendence-facticity relation is thus too simplistic. It rules out from the start any conception of authentic identification with my first-order states that *differentiates between* these states. Thus Sartre has made it impossible for Frankfurtian identification or alienation to occur as as *authentic* act in his sense: any differentiation which implies that some of my first-order desires or motives are *more internal* to my `true self' than others is therefore something which, on first principles, will falsely appear to be a form of bad faith. This is why Sartre is forced to conceive the sincere confessor as in bad faith—although neither of his types of bad faith allow us to understand how the sincere confessor's project is possible.

Directly related to this undifferentiated singleness in Sartre's authentic relation of transcendence to facticity is the fact that Sartre cannot distinguish authentic identification in Frankfurt's sense from this minimal `Lockian' identification that comes with all my `representations.' Since the only type of authentic identification with my facticity which Sartre's account allows must be the same for that facticity *as a whole*, in this respect it *seems like* minimal identification, yet it

clearly is something more than mere *consciousness of* having these first-order desires, impulses, and habits. Sartre's notion of authentic acknowledgement of first-order states thus *hints* at Frankfurtian authentic identification based in second-order volitions, which clearly stand out against minimal identification, but he cannot explain or locate any such structure using only the two-term language of `transcendence' and `facticity.'

The problem with Sartre's model, then, is that it doesn't clearly distinguish *enough* orders of volitional identification. Indeed, the inadequacy we found in Sartre's own explanation of the coquette, who clearly is in some kind of bad faith, already pointed to this same conclusion. The orders of `identification through volition' which the concepts of "transcendence" and "facticity" mask actually include *three different levels.* While my facticity includes first-order desires, impulses, and action-maxims that are *minimally identified* with me (because they are actual or possible elements in my consciousness), and the for-itself represents a *highest*-order will that has no `states' but only free projects or highest-order volitions, there is also a middle `level' of the will, which consists of *second-order volitions* in Frankfurt's sense. Frankfurt's second-order volitions are neither Sartre's `facticity' or in-itself nor his transcendental `for-itself.'

(B) The Essence of Bad Faith

In this section, I want to show as briefly as possible that if we start from the assumption that we have all three of the above `orders of relevance' for identification in volitional consciousness, it is possible to give more penetrating and better unified explanations of *all* of Sartre's major case studies in §2 of the chapter on "Bad Faith."

It is the distance between the *highest* order of identification, which we found in the complete or total identification of the for-itself with its original project (see §III-B), and the first-order states of motivation such as the coquette's D1 and D2, which opens up the possibility of bad faith. It is this space that the coquette exploits, in order *neither* to identify with D1 at the expense of D2 nor to identify with D2 at the expense of D1. The coquette is *wholly identified with her third-order will to have no second-order volitions* which would authorize or alienate either D1 or D2, or select one of these desires as the one she wants to be her will (i.e. the one she will act on). This is why on the surface, she *appears* to be similar to a `wanton' in Frankfurt's sense, i.e. to be someone who simply has no second-order volitions to decide which of their first-order desires they want to be their will. But the coquette's bad faith is not genuine wantonness, for she *wills* precisely this lack of second-order volitions which makes her *appear* to be wanton. This also explains why she seems in some

way similar to the authentically ambivalent person, but is not. The coquette does not have the conflicting second-order volitions $D_2$1 and $D_2$2 that we ascribed to her ambivalent counterpart. But she does have conflicting first-order desires, D1 and D2. It is not just any set of first-order states she wishes *not to separate* into `authorized' and `alienated' first-order states (which would be the effect of taking a second-order stand in favor of one of them)—but it is precisely this beautifully conflicting set of first-order desires which she wishes to leave alone. Her original project of having no second-order volitions thus aims indirectly at the preservation of `wantonness' towards the D1-D2 complex: thus we may say that she is identified *indirectly* with their conflict, since it is this conflict itself that she wants to preserve, rather than breaking it up through any constitution of her second-order will. Thus we may say that she *indirectly* wills the conflicting first-order desires that a person in authentic ambivalence wills directly. The coquette's case reveals the general structure of bad faith, which is as follows:

> *To Be in Bad Faith with respect to a given complex of first-order desires $D_1$1..D_1n* is:
>
> (1) To fully identify with a highest order volition W_3 (i.e. third order will) *not to form* any second-order volitions V_2 with respect to $D_1$1..D_1n; or
>
> (2) To will to remain `wanton' with respect to $D_1$1..D_1n.

Definition (2) is intended as a shorthand for definition (1), although *strictly speaking*, we cannot call anyone "wanton" in Frankfurt's precise sense of the term if they will this very `wantonness.' In bad faith, then, we get a person whose volitional consciousness has a gap in it, as follows: W_3--_____--$D_1$1..D_1n. Where the second-order volitions should be, there is a *nothingness*, which is still something in that it separates the for-itself or highest-order will from in-itself or first-order Ego. It is interesting to note that Sartre himself recognizes this distance between the for-itself and the factical first-order Ego, but because he does have the notion of second-order volitions which could fill the gap, because they are distinct from both "transcendence" and "facticity," Sartre is forced to see the distance as an *unbridgeable gap.* This is why authentic identification and alienation of one's first-order states ends up being impossible for Sartre. For example, when he puts himself in the waiter's place, Sartre explains: "It is not that I do not wish to be this person or that I want this person to be different. But rather there is no common measure between his being and mine" (p.102). Sartre can see no independent link or `common measure' *between* my transcendence and this facticity ('this person') which would allow me to `be' him without that meaning that I am trying to lose myself in this facticity, or which would allow me to "wish him to be different" (i.e. to *alienate* this first-order

`person', in Frankfurt's sense), without that meaning that I am trying to escape wholly into my for-itself. In both cases, it is only second-order volitions which can provide the common measure between heterogenous `facticity' and `transcendence.' But bad faith is precisely the *nihilation* of these second-order volitions: it thus destroys the bridging relation, and this is why in bad faith, transcendence (the 3rd order) and facticity (the 1st order) really are separated in the way that Sartre has detected. What he cannot explain is why this nihilation is contingent, or why we have the capacity to bridge the gap and to avoid being in bad faith.

Remarkably, we will find that this *same structure* can be found in every case of bad faith. Although Sartre thinks of the coquette as escaping into her transcendence, it is because she wills not to identify with or alienate any of her conflicting first-order desires that, from one perspective, she might seem aloof from them. Yet, from another perspective, we might say that her will to be `wanton' with respect to them is an attempt to live *wholly in* the complex of her first-order states, or to be like an animal with no second-order volitions that authorize or alienate its immediate impulses. This is what we find in the case of the waiter, who is trying to coincide entirely with the complex of first-order desires that would constitute a perfectly functioning waiter-animal. He wants to *be* this first-order complex of preferences and motives, *and no more*; in other words, he wants to appropriate waitering as an `animal substance.'

But this is not the same thing as authentically identifying with being a waiter, in Frankfurt's sense. Sartre's waiter must be contrasted, let us say, with a chef—Babette[44]—who honestly and openly *cares* about preparing the most delectable meals she can, or who authentically wills_2 to identify with her desires_1 to prepare attractive and nutritious food that will please her customers. Babette is <u>not</u> trying to be *nothing more than* this complex of first-order desires or a chef-animal: rather, she freely and authentically embraces them as *the self she wills to be*. Sartre's waiter, on the other hand, is trying to convince himself that he just is his complex of impulses and affections appropriate to being a waiter—he is trying to be a *wanton waiter*. Of course, in this he cannot succeed, because no person capable of having second-order volitions towards his own first-order desires can ever really be a wanton in Frankfurt's sense, or an animal that entirely lives in its fluctuating first-order states. Rather, he *chooses* not to exercise his capacity to form second-order volitions, and therefore he is a `wanton' only in bad faith. He is playing at being a waiter in the following sense:

---

[44]I am taking this name and character from the film *Babette's Feast.*

> He wills$_3$ not to have any second-order volitions V_2 with respect to D_1w (the complex of desires and affections appropriate to a waiter), in order to *seem* to be wholly D_1w.

Thus the essential structure of the waiter's bad faith does not differ from the coquette's. The only difference is that she wants to let herself be a *conflicting* set of first-order desires, whereas the waiter is aiming to be wanton with respect to a single coherent set of attitudes and impulses. The bad faith consequently stands out more clearly in her case than in his.

Although the paederast is supposed to be the opposite of the waiter, he also exhibits the same structure. For the paederast identifies fully with the will *not to identify* (through a second-order volition) with the homosexual desires on which he acts, but he is in bad faith because he *also* refuses to will$_2$ not to act on these desires$_1$, or to authentically alienate his homosexual impulses. This is the sense in which we can say that he refuses to make a decision regarding these impulses: although he shrinks from adopting them as his own, he also refuses to take the second-order stance of alienating them. We may thus distinguish him from the *unwilling homosexual* who earnestly wills$_2$ to act on heterosexual desires$_1$ he does not have, and so alienates the homosexual desires$_1$ on which he acts. In Frankfurt's terms, this unwilling homosexual is legitimately said to be *coerced* by his homosexual desires—he honestly cares about overcoming them, and therefore is presumably willing whatever means are at his disposal to that end. Therefore we *cannot say* that this unwilling homosexual is in bad faith.[45] He may be trying to fight something that he should not—something he would be better off accepting—but he is honestly identified with his will to fight it, which is not the same thing as `fleeing into his transcendence.' Sartre's paederast, by contrast, wills$_3$ to have no second-order volitions$_2$ towards his homosexual desires D_1h.

If my explanation of these cases is right, however, then why do the waiter and the paederast come out as having *opposite* types of bad faith on Sartre's account? The opposition, I suggest, turns out to be no more than a difference in emphasis. Since the waiter does not will to alienate his role, he seems to be `identifying' with it in too fervent a way, and so he is described as escaping into his facticity. Since the paederast refuses to authorize his desires as `his own,' or refuses to make them `*internal*' to his identity in Frankfurt's sense, he seems to be fleeing into his transcendence, or refusing all authentic identification with his first-order states. But in fact, as well as refusing to

---

[45]By this I do not mean to endorse or approve the position of the unwilling homosexual with respect to such desires. He treats his sexual orientation as if it were a corrupt addiction which is to be overcome. Discovering that there are no 'cures' for this condition which are not far worse than the 'disease' might be one reason to say that this view, although traditional, is not one that ought to be sanctioned. But my intention here is only to point out that this unwilling homosexual's will is different in a crucial respect from that of Sartre's paederast.

alienate his role, the waiter also refuses to *authentically identify* with his desires to make the café clean and serve customers well, as the contrast to `Babette' makes clear. He refuses to think of himself as freely willing$_2$ to act on these desires$_1$. And the paederast, while refusing to authentically identify with his homosexual desires, also refuses to actually will to overcome these desires, which would alienate them from him. So in fact the bad faith of the waiter and the paederast is a single thing: it is a highest-order will *neither to alienate nor authenticate* their own first-order desires and impulses. They both will to be `wanton' with respect to their first-order states. [46]

(C) The Vindication of Existential Sincerity

Let me briefly recap the advances made in the preceding analyses. First, my account of bad faith in terms of three orders of volitional identification has allowed us to explain Sartre's case studies more perspicuously than his own model can. Second, my account shows that the structure which lies behind the different examples of bad faith is the same structure in every case: we have thus reached a *unified* conception of bad faith. Third, this unified conception explains the superficial similarity between real bad faith and mere Frankfurtian `wantonness,' as well as letting us see why certain cases of bad faith (those with conflicting first-order desires) will appear superficially similar to authentic *ambivalence* as well. Finally, and perhaps most importantly, our discovery of the ür-structure of bad faith lets us see clearly why the sincere confessor and the unwilling coward are *not* necessarily in bad faith after all.

As we saw, despite some torturous twisting, Sartre's account is not able to explain how the unwilling coward can claim to identify with a set of ideal first-order `courageous' impulses which he actually doesn't have. But the three-order model clearly shows how this is possible without *either* escaping into one's transcendence, or escaping into a subjunctive facticity. The unwilling coward wills$_2$ to be$_1$ courageous, which necessarily means that at the same time, he wills$_2$ not to be$_1$ cowardly. He may fail in this endeavor, but in that case he is coerced by his fear; he is not in bad faith. But what are we to say about his highest-order will? The answer is that it depends: we need to know more about him to tell. He might *also* will$_2$ to remain$_1$ cowardly, but in that case he is in real

[46]Note that in Frankfurt's terms, trying to live solely in one's transcendence would amount to solely identifying oneself with certain second-order desires that do not purport to be second-order volitions. Notice that Frankfurt's doctor who only wants to feel addictive desires is playing at being an addict, just like Sartre's inauthentic waiter plays at being a waiter. Sartre's analysis shows, however, that when we can change our first-order will and the activities flowing from it, but refuse to do so by identifying only with a second-order *desire* (rather than volition), the tacit higher-order *volition* to let our first-order will stand may be attributed.

ambivalence, not bad faith. If not, we may assume that his highest-order will is *tacitly* behind his volition to overcome his cowardice: he is fully identified with the will$_3$ to persevere in his volition$_2$ to become courageous$_1$.[47]

Note that such a movement of authentic and non-ambivalent identification with some projected first-order state is not the same thing as trying to be$_w$ some first-order state. The highest-order of a volitional consciousness is never wholly anything but itself (which it *cannot* escape), and so we might say that the unwilling coward wills$_2$ to be$_1$ courageous, but still in the mode of not-being$_3$ what he is$_{1\text{-}2}$. Strictly speaking, however, this is not a `mode' of the for-itself, but rather its existential structure. As a free projection, it *cannot* will to be$_w$ what it is$_{1\text{-}2}$: its identification with lower orders of volition is only possible through its full and *yet free* identification with itself. When one is trying to flee into one's facticity in Sartre's sense, one is not really willing to be$_w$ what it is$_1$ after all: rather, one is *fully willing*$_3$ not to will$_2$ not to be what one is$_1$—but still *inevitably* in the `mode' of not-being$_3$ what one is$_1$. Thus authenticity and bad faith are both options for the highest-order will that cannot be *absolutely* identified with anything but itself. This means, of course, that even when the person has no abivalence in their authentic second-order identifications, or is "wholehearted" in their "cares" in Frankfurt's sense, they can never reach the point where other conflicting cares, or other possible inward character, are absolutely impossible: the highest-order will always has counter-character possibilities in its horizon.[48]

In Heideggerian terms, the highest-order will, and the *ownmost meaning* with which it absolutely identified, has a freedom that makes it `stand out' (or *ek-sist*) even from its own second-order volitions and dispositions. The existence of this freedom itself is not up to the person, and it does not constitute an *authentic* or relation of itself to itself in any sense: rather, as Kantian spontaneity is the ground of possibility of willing autonomously or heteronomously, so freedom of the highest-order will makes wholehearted, ambivalent, and bad faith identificational structure

---

[47]Indeed, in order to keep consistent with Sartre's descriptions, I am misusing virtue-terms to some extent here. I should really say that the unwilling coward wills$_2$ to act on some specific first-order desire other than fear (e.g. the desire to preserve his nation against the foreign invader). For *courage* as such is a virtue, which means that (with Aristotle) its structure is already defined in terms of having a disposition to a second-order volition not to act on impulses of fear$_1$. Thus we cannot, strictly speaking, talk of courage or cowardice as first-order complexes that can be alienated or authenticated. Virtues and vices *eo ipso* include second-order authentic identification, so to speak. Thus one of the effects of my theory, which shows how to account for bad faith without leveling off the possibility of second-order volitions, is to vindicate the *possibility* of virtues and vices in the classical sense, within an existentialist framework. Bad faith turns out not to be a vice, but something deeper, and worse. And in the process, virtues (including medieval ones, such as penitence), turn out not to be in bad faith.

[48]This means that "volitional necessity" in Frankfurt's strict sense is impossible, but it need not lead to arbitrariness in the history of one's inward will to be a certain sort of person in one's actions, or to care about certain kinds of projects.

possible. Thus my necessarily being$_w$ what I am$_3$ constitute a fleeing into my transcendence. Rather, the highest-order will, as the origin of identification, *exists* necessarily as absolutely self-identified, and this highest *volitional* reflexivity is *the very ground of possibility* of *both* an authentic relation to my first-order states through positive second-order volitions, and of a bad faith relation to my first-order states, through the *lack* of second-order volitions. This makes good on my earlier promise to explain how bad faith and authentic identification/alienation have the same transcendental source.

If we take sincerity in Sartre's sense as an achieved intrapersonal relation of *volition* (rather than in Frankfurt's sense as honest reflection on one's will), then it is clear that sincerity is the true opposite of Frankfurtian `wantonness.' *Existential* as opposed to cognitive `sincerity' means taking a stand towards one's own first-order desires, determining which of one's first-order affections and impulses one wishes to act on, or in Kantian terms, willing the formal maxim of one's action-maxims.[49] Frankfurt's work reveals that there are two different albeit mutually implicatory ways of being `sincere' in this sense: by authorizing a specific first-order desire as *internal* to the person one wills to be, and by alienating a specific first-order desire or making it *external* to the self projected by one's second-order will. The unwilling coward does both at once, alienating cowardice and authorizing courage.

The sincere confessor presents a more complex case, because if we think of him as now *alienating* certain unjust first-order desires on which he has acted in the past, he may be *on the road* to a second-order disposition to act on right action-maxims, but is not necessarily there yet. The reflective recognition that he has preferences for his own self-interest, or other base impulses that have led to unjust, unlawful, or immoral actions, is a necessary condition for his *confession*, which, if it is `sincere' in our sense, must also involve his adoption of a second-order volition no longer to let these first-order affections determine the maxim on which he acts. He must renounce his base first-order preferences, through an act of the second-order will. Of course, making them *external* in

---

[49]Of course, we should note that Kant never saw the possibility of doing otherwise, i.e. the possibility of 'wantonly' acting on impulses and desires. Since to act on such affections, we must incorporate their ends into our action-maxims, Kant holds that we must also will the formal maxim (the motive for the incorporation) in the same action. This holds true even for heteronomous actions, which are still performed through *spontaneous* incorporation, with a 'ground' (or formal maxim) for the incorporation. This is why we are *really* contradicting ourselves, according to Kant, when we formally will the incorporation into an action-maxim of a first-order end which is not universalizable. Kant's two-level theory of practical reason and volition is the antecedent of Sartre's, despite the intervening sophistications of pre-reflective consciousness which Sartre acquires from Fichte and Hegel. Kant's development of the idea of the *highest maxim* in his *Religion* is also the antecedent of Sartre's 'original project,' but here Kant is already beginning to perceive that the highest order of the will is a *third* level. My definition of bad faith, in Kantian terms, would be the adoption of a highest maxim *not to will* the formal maxims of one's action-maxims.

this way does not necessarily mean that he will immediately stop acting on these motives, but if his second-order volition remains firm over time and becomes a second-order disposition, there is reason to think that it will have a causal effect on his actions. A further task is still required of him, however. Since he must act in the world, in alienating his base motives, he must also adopt a second-order volition to act on *other* first-order preferences, emotions, and desires which lead to right actions. It is only when he authentically identifies with just first-order preferences through such a second-order will-to-goodness that, in Kant's sense, he will do what he does *for the sake* of the good, rather than merely in accordance with it.

Notice that sincerity in this sense, as constituted by one's second-order volitions towards one's actual and possible first-order states, complies with the metaprinciple for sincerity which I gave in §III-C. But it is precisely the opposite of what bad faith has turned out to be. Sartre therefore erred when he said that "the sincere man constitutes himself as what he is *in order not to be it*" and concluded that "the essential structure of sincerity does not differ from bad faith" (p.109). Rather, the sincere person reflectively recognizes what he is_1, in order to $will_2$ not to be_1 it, and eventually in order to $will_2$ to be_1 something else. This does not, of course, mean that hypocritical confession or confession in bad faith is impossible. Another person may very convincingly $will_3$ not to $will_2$ to be_1 or not to be_1 what she fully recognizes that she is_1. Such a person is a confessor, but not a *sincere* one. She is wallowing in what she is_1, but with no real intention of alienating it and trying to overcome it. But clearly we cannot conclude from such cases, as Sartre and Nietzsche would have us conclude, that there is *necessarily and always* an ulterior motive in bad faith behind confession. That is an illusion which results from not recognizing the full extent of differentiation within volitional consciousness, the three unassimilatable orders of identification which are at work in bad faith. Thus the possibility of existential sincerity, or authentic `second-order' alienation and identification, is vindicated.

V. Conclusion: Self-Deception and the Highest-Order Will

In closing, I should mention the remaining problem we have still not solved. In the third section of his chapter on bad faith, Sartre notes that "we have not yet distinguished bad faith from falsehood" (p.112). Our new account of bad faith does distinguish it *in one way* from both lying and from `bullshit' in Frankfurt's sense. Lying involves a second-order will to act on deceptive desires and preferences, which is also a will to misrepresent one's own authentic identifications. Bullshit is pure wantonness with respect to the truth: it means acting on first-order desires that lead to

communication with no interest in the truth, one way or another. But bad faith is a highest-order project *not to have* the second-order volitions necessary either to be a liar or a sincere communicator of the truth: thus the person in bad faith *appears* to be (but is not in fact) engaging in a kind of `bullshit.' This result follows from an application of our general schema and its distinction of bad faith from both existential sincerity (which can include lying to others) and wantonness.

But what we have not yet explained is how this type of bad faith can be an *inward-looking* attitude, a self-deception. Sartre is right, of course, that actively identifying with a project of self-deception is a fruitless endeavor: "if I deliberately and cynically attempt to lie to myself, I fail completely in this undertaking" (p.89). But, although Sartre's misalignments between transcendence and facticity purport to explain what the *errors* are in different forms of bad faith, they still do not show how it is possible to *deceive oneself* in these ways, without one's pre-thetic consciousness (of) one's own project of deception ruining it. Bad faith cannot be an intentional object; it must be a way of consciousness's being when it has *other* intentional objects. In fact, Sartre argues that one must in some sense be in bad faith `all the way down.' If that were so, however, it is hard to see how bad faith and authenticity could form two alternatives that *remain as possible original projects* of the for-itself.

Accordingly, Sartre has the most difficulty explaining how bad faith as a pre-thetic `way of consciousness' can begin. He concludes that "the project of bad faith must itself be in bad faith," or that "At the very moment when I was disposed to put myself in bad faith, I of necessity was in bad faith with respect to this same disposition" (p.112). Sartre's two formula for bad faith, however, do not help us understand how such an apparent *compounding* of bad faith is possible. But in his §III, he introduces the new hypothesis that *faith* itself, which he diagnoses as type of irrational "belief" which depends on "non-persuasive evidence" and resolves "to count itself satisfied when it is barely persuaded,"[50] is the original point which allows one to be conscious (of) bad faith *without recognizing it as bad faith.* In this first step, apparently, one has removed oneself from the norms of critical rationality which govern communication aimed at *truth*, and instead taken this perverse `faith' as a basis for belief:

> This original project of bad faith is a decision in bad faith on the nature of faith. Let us understand clearly that there is no question of reflective, voluntary decision, but of a

[50]Although as a critique of *faith in general* Sartre's argument here is not very convincing, it does provide a good basis for criticizing Frankfurt's latest conception of full identification as satisfaction with one's will. Frankfurt does not take sufficient account of the fact that what *counts* as satisfying (since that is mainly a cognitive category) will be conditioned by the highest order will. There is no escaping the fact that it is only in the for-itself that full identification originates.

spontaneous determination of our being (p.113).

This original project thus *conditions* one's consciousness (of) bad faith so that it is not self-defeating, but is still not an authentic, intentional *alienation* or rejection of the norms of critical rationality.

At this point, I will omit detailed criticism of Sartre's portrayal of all faith as proto-bad-faith. The problem may not be that Sartre is wrong about the origin-structure of bad faith, but rather that when he finds there a tacit *will* to an *uncritical* existence—to acceptance without adequate reason for real commitment—he calls this "faith." Kierkegaard would call it aestheticism, a childish simulacrum of real faith. This is important, because in this light, what Sartre's deduction suggests lines up well with Kierkegaard: bad faith, as a will not to engage in the kind of volitional commitment that involves one *personally* in norms of reason (including moral laws), must be a state of the highest-order will prior to explicitly *facing* the primordial choice between the aesthetic and the ethical.

The compelling point in Sartre's account is not his willingness to reduce "faith" to this tacit `aesthetic' character, but rather his insight that for bad faith (in his *or* my sense) to be possible, the freedom which spontaneously creates its *fundamental orientations* to basic categories of being and their universal ideals[51] (in this case, *towards* or *away from* truth about oneself) must *not* be conscious (of) itself: far from being an object of thetic consciousness, the highest-order or whole will must even lack the pre-thetic reflexivity of the pre-reflective cogito. Yet this is not a result Sartre can openly embrace. For it implies the limits of his own attempt to understand `volitional' modes through the reflexivity and internal differentiation natural to translucent consciousness. At the end of the hierarchy, we find a free source of volition which is *not related to itself in any <u>mental</u> way at all,* i.e. a will completely unmediated by consciousness. This astonishing result cannot be avoided if bad faith is to be possible. But it is also crucial to recognize that this result does not mean that the transcendental `I' which is$_w$ this free source of volition has no relation to itself which we might describe an *analog* of awareness. Its reflexivity is *purely volitional*, not conscious, and yet perhaps this is precisely the kind of ontological relation which grounds the highest *tacit* awareness

[51]There are three such categories (or world-concepts), and three corresponding ideals of *modal veracity*: (1) relation of the physical universe, and the ideal of objective, necessary truth; (2) relation to other persons (mitsein) and the ideal of *normative validity*; and (3) the relation to the intrapersonal world and the ideal of subjective honesty (see Jürgen Habermas, "Towards a Theory of Meaning" in *Postmetaphysical Thinking*). But there is also a fourth trans-category, the Divine or Eschatological, which stands over against the other three, and through which we relate to the other three in *different ways* than we do through the modal veracity ideals. In our relation to the Divine itself, there is no similar world-concept or modal ideal of veracity—it would be necessary to make all of this clear to respond sufficiently to Sartre's critique of faith.

and expectations in personhood, such as the expectations of time, a self, others, and the indeterminate modal range of possibilities each of these involve. Only if such a tacit *volitional* relation to one ownmost or most primordial self can subsist without reflexive consciousness, is bad faith possible, either in Sartre terms or in the new sense I have described.