

4-25-2013

# Akrasia and the Elusive Self

Frank R. Faries

*University of Missouri-St. Louis*

Follow this and additional works at: <http://irl.umsl.edu/thesis>

---

## Recommended Citation

Faries, Frank R., "Akrasia and the Elusive Self" (2013). *Theses*. 195.  
<http://irl.umsl.edu/thesis/195>

This Thesis is brought to you for free and open access by the Graduate Works at IRL @ UMSL. It has been accepted for inclusion in Theses by an authorized administrator of IRL @ UMSL. For more information, please contact [marvinh@umsl.edu](mailto:marvinh@umsl.edu).

Akrasia and the Elusive Self

Frank R. Faries

M.A., Philosophy, University of Missouri – St. Louis, 2013

B.A., Philosophy, University of Missouri – St. Louis, 2011

A Thesis Submitted to the Graduate School of the University of Missouri – St. Louis in partial  
fulfillment of the requirements for the degree  
Master of Arts in Philosophy

May 2013

Advisory Committee

Gualtiero Piccinini, Ph.D.  
Chairperson

Berit Brogaard, D.M.Sci, Ph.D.

Eric Wiland, Ph.D.

### **Abstract**

The concept of *akrasia*, or weakness of will, involves knowingly acting against what one judges best. This seems like a paradox of irrationality. However, I argue *akrasia* implies a notion of the self as a single, agentive mental object which persists across time, has a particular character, and is the subject of experience. This notion corresponds to the “folk” notion of the self invoked in everyday self-experience. However, the folk notion of the self is not only unnecessary for self-experience, it is also the source of the air of paradox surrounding *akrasia*. The novel view of the self introduced herein—the Multiple Occupant (MO) theory of the self—not only vindicates everyday self-experience, but also resolves the paradoxical nature weakness of will. Furthermore, the theory is suitable not only for establishing the diachronic unity of a self (even one that can act *akratically*) over time, but for establishing the *disunity* of a self in cases of dementia, Dissociative Identity Disorder, and Alzheimer’s disease.

*Keywords:* *akrasia*, weakness of will, self, self-experience, unity of self, rationality.

### **Introduction**

It is often the case that ubiquitous concepts permeate ordinary language. One such concept is the feeling of self-experience “which is as fully present at any moment of consciousness in which it *is* present, as in a whole lifetime of such moments” (James, 1890). This feeling leads us to posit an object, a self, which operates in the manner which we feel it operates in self-experience. This “folk” notion of the self influences ordinary language usage, and much of the way we think and feel about human beings in general. One area where this influence has encroached is the concept of *akrasia*, often called weakness of the will, or acting against one’s best judgment. It is the influence of the folk notion of the self which accounts for the air of paradox, and the apparently unsolvable nature of *akratic* action. In fact, the folk notion of the self is central to the standard conception of *akratic* action.

This influence oversteps the boundaries of the self-experience. In fact, there is good reason to believe the folk notion of the self is just plain wrong. We need only account for the unity of self-experience at a time and across time. Folk notions of the self go far beyond what is given in self-experience. Thus, a minimal view of the self can explain the unity of self-experience at a time. Furthermore, lateralized brain activity suggests that unity of self-experience across time is more complicated than originally thought. In response, I propose an original thesis about the nature of the self across time: the Multiple Occupant theory of the self. Not only does this

vindicate everyday self-experience, as well as the unity of self-experience across time, the MO theory offers a way to resolve the seemingly paradoxical nature of akratic action.

The rest of the paper will go as follows. Sections 1 and 2 will provide a literature review on the folk notion of the self and akratic action, respectively. Section 3 clarifies how the folk notion of the self is implicit in the literature on akratic action. Section 4 argues that a minimal view of self-experience is all that is required for selfhood, and goes on to show why the folk notion of the self is much too bloated to achieve this. Section 5 will present the Multiple Occupant view of the self, and Section 6 will show the application of this novel thesis to the case of akratic action.

### **1. The Folk Notion of the Self**

There are indelible aspects of human experience which underpin the way in which we engage the world. Our fundamental experience of the world is mediated by the inescapable feeling that we are inhabitants of our bodies. This feeling is the source of perennial disagreements in philosophy. To what, exactly, does this feeling correspond? Where shall we locate its source? That is, philosophy seems concerned with the metaphysical status of this feeling. Indeed, the feeling appears to have led to many terminological disputes as well. Philosophers have yet to reach a consensus on what to call this feeling. Some candidate terms are awareness, consciousness, the soul, the self, and many more besides. Furthermore, no one is quite sure what this putative entity does. Does it issue commands to the body which it inhabits, like a charioteer to his horse? Or is it merely an ineffectual bystander to the events going on around it?

In spite of the uncertainties about this feeling, one indubitable fact is its ubiquity. And, like most ubiquitous entities, it is a lightning rod for philosophical controversy. Ubiquitous concepts tend to permeate ordinary language, leading opposing philosophical camps to cry 'foul' when they feel the infringement has crossed a critical threshold. For my part, I cannot claim to provide any resolution to the philosophical controversy regarding this feeling. However, I do

believe I can trace the influence of this feeling on folk notions about the self. This infiltration is unwarranted, and my intention is to find the proper boundaries of this feeling and selfhood.

First, it might be best to give this feeling a name. The choice of name may be equal parts investigation and baptism. Nevertheless, it will provide a single point of reference to eliminate any potential for confusion or terminological dispute. I find the term “self-experience” sufficiently descriptive and acceptably neutral. The “self” in “self-experience” here is not meant to invoke any particular connotation about the self. Rather, “self” refers only to the subjective aspect of this feeling as being owned by the experiencer. A suitable replacement might be the prefix “my”, and hence, “my-experience”. Because my intention is to generalize the application of this term, however, I wish to avoid the use of self referential terms like “my”, and thus “self-experience” is more suitable to my purposes.

Now that we have a name, we must give a more specific description of the feeling. Here I must defer to William James’ *Principles of Psychology*, in which he states self-experience “is found to consist mainly of the collection of these peculiar motions in the head or between the head and throat” (James, 1890). That is, in attending to any idea or sensation, we are not only cognizant of our minds reaching out and engaging with that idea or sensation, but of the various adjustments and executions of sense organs involved in the particular mental act. Furthermore, as James later remarks, those thoughts or feelings which make up our self-experience have a particular warmth and character about them that tells us they are distinctly ours. Self-experiences at different times seem to share this same warmth. Finally, as James concedes, the nature of this self-experience leads us to think of ourselves as a thinker issuing fiats of the will, in comparison to which all external change is merely transient. That is, the phenomenological character of self-experience presents an overwhelming temptation to posit a *self*: a persistent, unified, agentic mental object that consciously thinks and feels in accordance with certain dispositions.

This is the folk notion of the self that most people tend to invoke in everyday language usage. It is the source of the intuitions about questions of personal identity. It is the evaluative

basis in determinations of dementia and dissociative identity disorder. It is the notion to which you might defer when posing to yourself the nebulous question, "What do I really want?" It is the sort of notion that seems vindicated by everyday self-experience, and it can be dissected into seven parts. Following Galen Strawson (1999), the folk notion of the self is: (i) a subject of experience, a conscious thinker and feeler; (ii) a thing, in some robust sense, and what's more a (iii) mental thing, that (iv) is single at any given time, and during any unified or hiatus-free period of experience, but (v) continues to exist across hiatuses in experience (i.e., a persistent thing); it is also (vi) an agent with (vii) a particular character or personality.

## 2. The Paradox of Akrasia

Akrasia is a negative predicate referring to the lack of a particular character trait, *kratos*, or self control. A person who exhibits akrasia shows a lack of self-control or a weakness of will. Therefore, akratic action is action that exemplifies this lack or weakness. I agree with Alfred Mele (1992) that, standardly conceived, act X is akratic just in case it is *free*, *intentional*, and contrary to *better judgment*. Needless to say, in contemporary philosophical discussion, akrasia has the air of a paradox. It seems like a paradox of rationality that we can freely and intentionally act against better (or best) judgment. As Hume has famously noted, "[w]hatever has the air of a paradox ... is often greedily embraced by philosophers" (Hume, 1739, p. 26). After investigating these three characteristics of akratic action, I hope to determine if philosophers are simply being a bit too greedy when it comes to akrasia.

In considering possible kinds of akratic actions, there is an intuition that the degree of akrasia we attribute to an act X is inversely proportional to the agent's compulsion to X. A heroin user is less akratic in getting a fix against his better judgment after five years of addiction than he was in trying heroin for the first time against his better judgment five years prior, *ceteris paribus*. In this case the heroin addict is compelled to use heroin in a way that the thrill-seeking first-time

user is not. Therefore the addict's action is less akratic than the thrill-seeker's. Compulsion to act violates the condition of freedom, and reinforces this condition as a constituent element of akratic action.

A second intuition that arises in cases of akratic action is that the act must be *intentional*. If I hold that it is best that I do not eat cake (because of my diet), and come home to find what looks and tastes like a steak dinner waiting for me, but upon eating it I am told that it was actually cake batter formed and flavored to appear to be a steak dinner, this would not be a case of akratic action. Primarily this is so because I did not intend to eat cake when I ate what I thought was a steak dinner. Arguably, a case can be made that my freedom was also compromised in some sense by my lack of knowledge. Even if this is true, we can still see how intention factors into akratic action. It must be noted that although the act must be intentionally chosen, it need not be intentionally *akratic*. A woman who (intentionally) runs six miles every day may know that in doing so she is damaging her knees without intending to damage her knees. In the same way, if I eat cake (this time shaped and flavored like cake) against my better judgment I may know I am acting akratically without intending to act akratically. I think that we can put this assumption to bed, but I am aware that it might be a contentious point for some (cf. Bratman, 1984).

The third and final point regards the agent's judgment of what is best. While the first two points serve to establish the agent's act as *rational*, it is in acting against better judgment that the act takes on its distinctly akratic character and is thus *irrational*. As mentioned above, akratic action is taken to be a paradigmatic example of irrationality, and it is this characteristic that makes it such. But the term "better judgment" (or "best judgment") seems heavily loaded, and it will be best to unpack this term before moving forward.

Trouble is, this is a vague term, and any allusions to better or best judgment in the literature are sketchy at best. Indeed, Frank Jackson certainly seems to make the matter all the more confusing. His account of akratic action allows acts that may be in *accordance* with best judgment. Jackson argues that legitimate cases of weakness of will can occur in cases where the

agent fails to do what she deems is the *wrong* action. His example for this argument is a committed Catholic woman who is raped and becomes pregnant. She knows it would be wrong to get an abortion, yet she nevertheless decides to get one. When the time comes, she becomes wrought with guilt and fails to follow through with her decision. “Had she been more strong willed” Jackson claims, “she would have gone through with the abortion; hence her failing must count as one of weakness of will. She acts incontinently, while doing what she judges best” (Jackson, 1984, p. 4). The case is interesting not only because there are two nested akratic actions, but because these actions seem to “cancel each other out”. Again, there is an air of paradox about akrasia which Jackson rightly tries to bring to the surface.

In fact, I think Jackson’s example highlights a broad distinction between two ways in which we can interpret what is meant by best or better judgment. If the choice to abort her fetus is a legitimate case of weakness of will, then where does the failure lie? In fact, I believe there are two akratic actions here, each with their own corresponding best judgment. First, as Jackson believes, that the woman is unable to abort her fetus is a legitimate akratic action. The best judgment which it violates is the decision she made to have the abortion, at the time of having the abortion. This is a particular judgment held at a particular time; an *episodic* judgment. To commit akratic failure at this point is to commit a failure of rationality. If we hold that what we decide or intend is necessarily what we judge best at that time, then it seems irrational that we can act in spite of what we intend (or judge best) to do. However, the decision to have an abortion is itself an akratic action. The best judgment which it violates is the judgment to abide by the precepts of the Catholic church. This judgment is *diachronic*: it spans over a broad range of decisions she will make throughout the course of her life. This akratic failure is a failure of character. The fact that Jackson only acknowledges one case of akratic action is evidenced in his failure to acknowledge the two judgments at work. As Amelie Rorty (1980) states, “[t]o say that the akrates violates his preferred judgment about what is best, is ambiguous between saying that he deviates from his judgment, *however arrived*, about the preferred course, and saying that he deviates from



a preferred course which he underwrites because it is the most rational course” (p. 337). Here I read Rorty, in the former interpretation, to refer to episodic judgment (i.e., the Catholic woman’s judgment at the time of the abortion, however she came to decide it), and in the latter to diachronic judgment (i.e., that following the tenets of Catholicism is the most rational course for her life). That Rorty’s usage of the term “rational” corresponds to a category other than failure of rationality is, I hope, not due to a deficiency in my interpretation, but rather betrays the ambiguity that plagues the literature with regard to that term as well.

So it seems we have two kinds of better judgments which correspond to two broad types of akratic action. Episodic judgments correspond to what Mele (1987) calls *strict incontinent action* (SIA): “action A is a strict incontinent action if and only if it is performed freely and intentionally and, at the time at which it is performed, its agent consciously holds a judgment to the effect that there is good and sufficient reason for not performing A at this time” (p. 7). Episodic judgments, then, are consciously held judgments that figure directly into action-guiding cognition. Mele refers to these kinds of judgments as Decisive Better Judgments, for they commit an agent to action, presumably on pain of irrationality. Importantly, these judgments need not be explicitly endorsed.

Diachronic judgments tend to take on a moral or prudential character, in light of their broad scope. “Best judgment” here refers to “the judgment that one reaches, having taken into account all the reasons one judges to be relevant, as to what would be best for one to do in a given situation” (Arpaly, 2000, p. 490). Diachronic judgments are more like the kind of “better judgment” spoken of in akratic failures of character: they reflect judgment in light of “all things considered”.

Jackson’s example shows how these two kinds of judgments can come apart. That he does not acknowledge these two kinds of judgments as distinct belies a powerful assumption which underpins much of the literature on weakness of will. That is, Jackson takes episodic

judgments to be reflective of diachronic judgments. This conflation, as I will argue below, is one example of a particular view of the self is being smuggled into talk of akrasia.

### **3. How the Folk Notion of the Self is Implied in Akrasia**

Jackson's conflation of diachronic and episodic judgments is but one symptom of the influence of a particular view of the self on notions of weakness of will. Two other symptoms involve the implications of the freedom condition in cases of akratic action, as well as the relation of paradigmatic cases of akrasia to character and personal identity. Together, these elements favor a view of the self as something quite like that which figures in self-experience, as given above. In particular, these elements stress conditions (v-vii) of the folk notion of the self: the persistence, agency, and personality conditions.

To show that the conflation of diachronic and episodic judgment necessarily implies a persistent self requires only a simple argument. Jackson glosses over the fact that somehow the Catholic woman came to form the intention to abort her fetus, in spite of the fact that her best judgment in life is to be a Catholic. A charitable interpretation of this gloss might be that Jackson sees the woman as simply having changed her mind. That is, her diachronic judgment B, to be Catholic (and thus not choose abortion), changed to a diachronic judgment A, to abort. The case of akratic action thus arises for Jackson only in the action X, the choice not to abort at time  $t$ , in relation to the episodic judgment A, that she should abort. That is, X is not akratic when considered against judgment B. This is not a problem for Jackson, as the woman, at time  $t$ , does not hold judgment B. That is, Jackson does not acknowledge two instances of akratic action because, for him, the diachronic judgment simply *is* the episodic judgment. What the woman deems best across times (A), simply *is* that which she concludes it is best to do at time  $t$ , and against the judgment of which she chooses X. If this were not the case—if B were what the woman deemed diachronically best—Jackson would be forced to recognize the decision to A as

an instance of akratic action, and thus the decision to X as an akratic action against the episodic judgment A, which is itself an akratic action formed against diachronic judgment B. This is the position I hold, and I argue that it arises due to the recognition of the distinction between diachronic and episodic best judgment, which itself cannot occur in a persisting, continent self. That is, a persisting object such as that posited by the folk notion of the self cannot exhibit both A and B as judgments, and thus recognize both instances of akratic action. A defining feature of a persisting self is the collapse of the distinction between diachronic and episodic selfhood. The distinction can be highlighted by thinking about oneself as stretched over one's lifetime, that is, diachronically, and thinking about oneself *at this very* moment, episodically. According to the folk notion of the self, there is a single entity which is identical in both scenarios, and persists, unchanged, throughout the former. Maintaining the episodic/diachronic distinction allows that what exists episodically need not be identical to what exists diachronically. But the folk notion of the self does not recognize this distinction. Therefore akratic action that does not recognize the distinction between diachronic and episodic best judgment implies an object which persists such as that posited by the folk notion of the self.

By the condition of agency in the folk notion of the self, I mean something like the feeling that one is controlling or intentionally producing one's thoughts. This seems inexorably tied to the freedom and intentionality conditions of cases of weakness of will. To intentionally choose an action without being compelled to do so, one must feel that one has the capacity to have chosen otherwise. To illustrate this, we need go no further than a common objection to cases of akratic action; what Alfred Walker (1989) calls "The Enslavement Argument". This argument attempts to undermine the possibility of akratic action on the grounds that these actions are irresistible and thus not free. In addition to attacking the possibility of akratic action, these objections threaten the view of the self as agentive.

The Enslavement Argument has taken various forms, but generally speaking, it posits a distinction between motivation and evaluation. The deliberative process that precedes cases of

akratic action produces a dissuasive best judgment which we set for ourselves, and yet we still act against. That is, we are motivated in a way that is contrary to our rational evaluation of a situation. Alternatively, we might say that our evaluations, or best judgments, motivate us in ways that are different from, and presumably less powerful than, non-evaluative motivations. Thus if the evaluative, or rational, aspect of our judgment is unable to overcome our “passions”, or non-deliberative motivations, then our agency is in some way compromised, and creeps dangerously close to compulsive, or unfree, behavior. Indeed, the very notion that the will (or, roughly, the control of thoughts and actions) can be *weakened* implies some agent being rendered ineffectual.

Finally, regarding the relation of the personality condition of the folk notion of the self to akratic action, we need look no further than everyday, man-on-the-street ideas about weakness of will. That is, akratic actions, if they exist at all, mark a point on the continuum between continent, rational behavior by a person of a general character or disposition, and clear cases of a *change* in character. They are a way to make sense of how people can act “uncharacteristically” without complete loss or change of character. General intuitions are by no means a bulletproof standard by which concepts are defined, but they serve as a benchmark for the intended meaning of a term, or the role which a concept is thought to serve. In the case of akrasia, this role is to fill in the narrative gaps in a person’s character left by uncharacteristic action. (Who decides that the action is uncharacteristic is another question, but I feel this point equally stands whether it is oneself or others who make this call). When the frugal mom splurges on an expensive new handbag, even though she overtly asserts that she should not, we do not say that she ceases to be frugal, or that she is not the same person she was before. Indeed, character traits are hardly ever supposed to encompass every action ever undertaken by someone.<sup>1</sup> A woman can be “frugal”, a man “temperate”, or a child “calm”, while still succumbing to the occasional spending spree, tantrum, and outburst, respectively.

---

<sup>1</sup> Aristotle might disagree here.

In this section I have shown how the folk notion of the self is implicit in cases of akratic action. The conflation of diachronic and episodic best judgments implies the self as persisting. The concept of freedom in cases of akratic action implies the self as an agent, choosing its thoughts and actions. Finally, paradigmatic cases of weakness of will serve to establish and reinforce the narrative identity or personality of the self. Defenders of the possibility of akratic action, and incontinent action more generally, hold fast to this notion without good reason. I will argue in the next section that this notion of the self is a vestige of archaic thinking, and unsupported by current empirical science. In subsequent sections, I hope to show that in giving up on the folk notion of the self, we need not give up on the possibility of akratic action.

#### **4. Why the Folk Notion of the Self is Wrong**

In attempting to determine what kind of inductions we are licensed to make about the self, it is best to seek the necessary and sufficient conditions of self-experience. If one of the ways which we take our self-experience to be can be dispensed with, there is no reason to cling to it. The motivation here is not simply parsimony; if a particular phenomenon can be explained in two different ways, we should prefer the way which is in greater accord with empirical evidence. In the same vein, we should be suspicious of any explanation that conflicts with our everyday experience. In most cases, the manner in which we explain a phenomenon will do little to change its appearance to us. A straw placed in a glass of water looks bent whether you believe it looks that way due to the refraction of light or the machinations of straw-demons. However, if straw-demons seem so unlike anything vindicated by everyday experience, we are justified in being doubtful of their existence.

Self-experience presents experience of the “specious present”, the short duration of time of which we are immediately aware. How long is this duration? Perhaps anywhere from a few seconds to a minute, as William James surmised? The precise answer is irrelevant for our purposes; we need only stipulate that there is some duration, SP, during which we can establish a

cross-sectional, or episodic, unity of a particular mind at a particular time. In fact, James may have meant for the specious present to be a bit longer in duration than SP, but the particulars are unimportant. All we need is a functionally relevant unit of time to which we can rightly ascribe to an agent a thought, belief, desire, etc. It is in this duration of time that self-experience operates.

What is the minimal form of self required to capture self-experience during SP?

Following Strawson (1999), I recommend that we abandon (v-vii) above. We can thereby establish the minimal conditions of self-experience of SP as being a mental thing (in some sense) that is unified at SP and consciously thinks and feels. This sounds reasonable, but there are those who might not let me so easily dispose of concepts like persistence, agency, or personality. An argument is in order.

First, we deal with personality. This one should be easy to do away with. Personality and character traits can be conceived as dispositions to act in a certain way. A person at SP may be disposed to act in certain ways, but to claim this as a personality the person must exhibit this disposition over multiple SPs. Indeed, it may be that a single person, alone in the universe, has no distinct personality without other people to reflect it. We see personality clearly when we see it in others, but not so clearly in ourselves. Furthermore, that we are in a mood at SP does not entail that we are aware that we are in a mood at SP, though this is commonly the case. Finally, in my own experience it is common to have dreams in which I feel like I am myself, but act completely out of character, yet during the dream, this strangeness of personality does not bother me, nor am I even aware that it is strange. Anecdotal evidence aside, we have good grounds to let go of personality as a necessary condition of self-experience.

Agency might not be so easy to relinquish. That one is consciously in control of one's thoughts seems self-evident to some. For his part, I feel Strawson does not do enough to dismiss such a vital concept. He believes people are equally justified in adopting a view of their mental lives "in a Rimbaud- or Meursault-like fashion, i.e., almost entirely as something that just happens to them, while others naturally think of themselves as controllers and intentional

producers of their thoughts” (Strawson, 1999). Though I am sympathetic to his example, I am certain others are not. In support, I offer the simple examples of (day)dreaming, so-called ‘stream-of-consciousness’ associations, and lucid dreams. Someone speaking or writing in a ‘stream-of-consciousness’ style can make associations that seem to come from nowhere, the genesis of which they feel no agency over. Furthermore, in dreams, it is common to feel like an observer seeing strange thoughts. Lucid dreams are characterized by the feeling of agency about the images in dreams, suggesting that, by default, dream images are spontaneous. In any case, these examples should serve to support Strawson’s point that the feeling of agency is not required for self-experience.

Finally, the elimination of persistence is going to be a jumping-off point for some. It is here that I must bid those with traditional views of the ‘soul’ adieu. For any that remain who hold a thesis of some kind of mind-brain identity (reduction, supervenience, epiphenomenalism, etc.) perhaps persistence will be the easiest to cast off. This is well within their right, as self-experience at SP remains silent about what happened *before*, and what happens *after*, SP.

Thus, we see that self-experience, the primal engagement with the world that is the foundation of the folk view of the self, has shed its fetters and chains. We have no need of persistence, agency, or personality. What remains is the form of self-experience at SP—the self as a Subject of Experience that is a Single MENTAL Thing, or SESMET. This is Strawson’s term, and I will follow his usage. He believes, as I do, that SESMETs satisfy the necessary and sufficient conditions of our self-experience at a time. However, an important point to raise is this: we don’t experience single, isolated SPs. Even if we are willing to forego persistence as a condition of self across time, there is an important sense in which there is a unity in our self-experience across time. Any acceptable notion of self, therefore, must account not only for unity of self-experience at a time, but unity (in some sense) across time.

We can imagine all the experiences of a SESMET at SP as a kind of “time slice” of that mind, or a Total Temporary State (TTS), to use Grice’s (1941) term. The cross-sectional or

episodic unity of a SESMET at SP is determined by establishing which mental states are parts of the TTS of that mind, and which are not. If two people are looking at a red ball, part of the TTS of each person will be the experience of the red ball. For our purposes, let us set aside differences in perspective and individual histories of the two people; we will assume their experiences of the red ball are identical. We want some way to say which red-ball-experience belongs to which person. Furthermore, if we conceive of a person's mind as a chronological connection of TTS' held together in a certain way, then we need a way to establish the longitudinal unity of a mind. That is, we need a way to connect all the SPs that make up the life of a person. Thus, self-experience presents us with episodic unity and diachronic unity. Thus any necessary and sufficient conditions for self-experience (and thus a licensed version of the self) will have to account for both of these.

### **5. SESMETs and the Multiple Occupancy View of the Self**

“The assumption of one single subject is perhaps unnecessary; perhaps it is just as permissible to assume a multiplicity of subjects, whose interaction and struggle is the basis of our thought and our consciousness in general” (Nietzsche, 1967).

Theories of the identity of self across time can be categorized along a number of dimensions. C.D. Broad gives an excellent categorization in his 1925 *Mind and its Place in Nature*. The first division cuts at the most prominent aspect of the question of personal identity: the presence or absence of a center. The center, according to Broad, is “a certain particular existent ... which stands in a common asymmetrical relation to all the mental events which would be said to be states of a certain mind, and does not stand in this relation to any mental events which would not be said to be states of this mind” (Broad, 1925, p. 558). Thus, there are



those theories about the unity of the mind which posit this center—either in the form of a Pure Ego, or some kind of central event—and those which deny it.

Language use, at first pass, might seem to favor central theories. “This certainly suggests that “I” is the proper name of a certain existent which stands in a common asymmetric relation to all those contemporary mental events. We say further: “I, who am now doing and feeling these things, was yesterday doing, thinking, wanting, and feeling such and such other things.” And this certainly suggests that “I” is the proper name of something which existed and was a centre yesterday as well as to-day” (Broad, *ibid*, p. 584). This center is where language places all the thoughts, beliefs, desires, and fears of a single mind. When a man, deep in thoughts, asks aloud, “What do *I* really want?” the ‘I’ seems to refer to the thing underlying the longitudinally unified succession of TTS’ in his mind. This sounds a lot like a center, and to ask about what kinds of desires range over the entire longitudinal unity implies that there is something permanent or unchanging. That is, the colloquial use of self-referential terms like ‘I’ or ‘me’ at least *implies* the identity of self across time.

However, as Reid noted, there are few things which are truly identical between two successive mental states. Each of the distinct states of a mind “are all successive in their nature, like time itself, no two moments of which can be the same moment” (Reid, 1785). Furthermore, if this center has nothing to do with the body in which it is contained, then the same person should be able to withstand changes to anything other than the center. And indeed, as Locke’s (1690) prince and pauper will attest, living beings can survive changes in matter without a change in identity. However, there does seem to be a point at which non-central properties seem to matter. To complicate matters, most of the historical problems about personal identity assume that the answer to the question, “Are X and Y the same person?” can only be “yes” or “no”. That is, the literature on personal identity seeks to determine which (or to what degree) changes in substance or memory result in a loss of personal identity.

Appeals to ordinary language, theoretical disputes, and thought experiments, however, have not yielded any substantial answers about the identity of self, only warring intuitions. In fact, the tensions about personal identity arise from the real borderline cases, not from thought experiments. The groundbreaking work of Sperry and Gazzaniga on split brains have shed light on cases of personal identity, and in turn spawned their own thought experiments.

Sperry and Gazzaniga experimented on subjects who had received commissurotomies—or the severing of neural tissue separating the left and right hemispheres of the brain. These subjects exhibited remarkable effects in their perception. In a slightly simplified case adapted from Gazzaniga, et al (1962), a subject is shown the word “face” in his right visual field, and asked what he saw. Because the left hemisphere is dominant for verbal processing, the subject is able to verbally respond with the word “face”. Subsequently, the subject is shown the word “face” in his left visual field, and asked to report what he saw. This time, the subject is unable to give a verbal report of the word, but instead can draw a face. This is because the information was delivered to the right hemisphere, and cannot be shared across the now-severed corpus callosum to areas controlling verbal processing. This suggests that there are two streams of consciousness occurring in split brain subjects. What makes this remarkable is that, despite these two streams of consciousness, the subjects, by and large, act normally. Their actions appear unified, and they report feeling a subjective unity. That the subject has two distinct visual input streams which appear to be unable to communicate information between them does little to effect the overall subjective self-experience, and is only drawn out in atypical experimental conditions.

It is obvious that if I do not respond to the conditions in the split-brain experiments in the same way (and, presumably, if I have not had a commissurotomy) then I do not possess two independent streams of consciousness in the same way that the split-brain patients do. However, what the research does show is that there is nothing, in principle, that could prevent it from being the case, as far as self-experience is concerned. Remember, Sperry and Gazzaniga’s subjects did not report any change in subjective unity, and their behavior was fairly normal. In fact, when

asked about why the split-brain subjects did not notice the “missing half” of their visual experience after the surgeries, Gazzaniga himself has explained via a metaphor of a pipe organ. “The thousands or millions of conscious moments that we each have reflect one of our networks being “up for duty”. These networks are all over the place, not in one specific location. When one finishes, the next one pops up. The pipe-organ-like device plays its tune all day long. What makes emergent human consciousness so vibrant is that our pipe organ has lots of tunes to play” (Gazzaniga, 2008, p. 321). I find this to be an especially vivid image, and one you might expect given his research. In principle, split brain subjects could conceivably experience alternating SPs which make up the chronological succession of SPs in their lifetime: each SP handing over the next to the opposite half. All the while this passing of self-experience back and forth between hemispheres happens under the radar of subjective awareness.

Split-brain cases gnaw at some of the fundamental intuitions about personal identity, or the diachronic unity of self over time. What has happened to the subject? Does his identity belong to his left-brained or right-brained self? Does it belong to some combination of both? Or, even worse, does the subject completely lose his identity? The two brain halves each appear to have their own stream of consciousness. If, upon separating my left and right hemispheres, each were put into its own body (both of which are exact genetic duplicates of the now brainless me), which one would *be* me? This thought experiment urges us to abandon talk of an unchanging self, and adopt a more flexible approach. Given the split brain cases, central theories seem hard to maintain. Instead, as Derek Parfit recommends, talk of personal identity should be replaced with a language that allows for split-brain cases.

Rather than being a relationship of strict identity, diachronic unity of self over time is a matter of degree. We say not that the self at  $t_1$  simply is the self at  $t_2$ . Instead, the self at  $t_1$  need only share a sufficient degree of unity with the self at  $t_2$  to say that the self has *survived* from  $t_1$  –  $t_2$ . Survival is a sufficiently robust notion to capture what matters in cases of personal identity, while granting the inherent differences between two consecutive selves. When we use everyday

language to speak of personal identity, they are not directed at strict identity, but “judgments of personal identity do derive their importance from the fact that they imply psychological connectedness” (Parfit, 1971). Thus we may say that mental states X and Y belong to the same self if they have sufficient psychological connectedness. This relation can be one-one, one-many, or many-one, and need not presuppose identity. Most importantly, psychological connectedness is a matter of degree. When we ask about matters of personal identity, then, we are trying to ascertain the requisite degree of psychological connectedness.

In most cases, the relative similarity of a person’s behavioral and external characteristics makes sameness of identity the default assumption. If a person looks the same, acts (generally) the same, and the possession of the right memories can be inferred from their actions, that person is generally assumed to be the same person. Some paradigmatic cases of loss of personal identity might help mark the boundary of psychological connectedness separating survival of identity from death. Two kinds of cases which are especially illustrative are Alzheimer’s disease and dissociative identity disorder (DID).

Alzheimer’s patients show signs of increasing loss of memory capabilities. In earlier stages, these capabilities are typically short-term (i.e., difficulty remembering the name of a new person, memory lapses about everyday objects or common words), while in later stages the damage to memory affects procedural and narrative memory. The memory loss is diagnosed based on the behavior of the subject, but also on self-report and other symptoms like increased confusion and frustration suggesting that, for a patient with Alzheimer’s disease, all is slipping away on the inside.

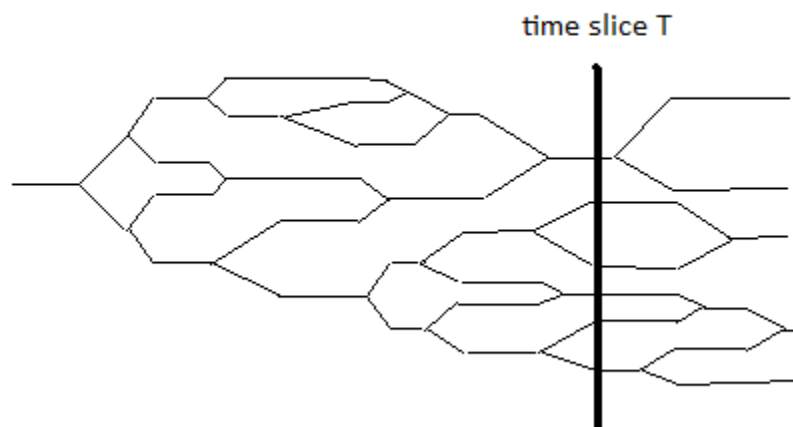
Regarding the latter condition, *Diagnostic and Statistical Manual of Mental Disorders* notes four criteria for diagnosing DID. “(1) Two or more distinct identities or personality states are present, each with its own relatively enduring pattern of perceiving, relating to, and thinking about the environment and self. (2) At least two of these identities or personality states recurrently take control of the person's behavior. (3) The person has an inability to recall

important personal information that is too extensive to be explained by ordinary forgetfulness. (4) The disturbance is not due to the direct physiological effects of a substance (such as blackouts or chaotic behavior during alcohol intoxication) or a general medical condition (such as complex partial seizures)” (*DSM-IV-TR* (2000) 4th ed., text rev.). These criteria are behavioral—they are a way for external observers to determine that the subject is switching from one personality, or ‘alter’, to another. That is, in addition to effects on memory, the behavior of the patient has become so unlike their ordinary behavior (habits, posture, manner of speech) that they appear to have adopted a different personality state.

I believe that what makes these paradigmatic cases of identity loss so painful to friends and families of the patients is the fact that the loss of their loved one to the disease is *not* an all-or-nothing affair. Alzheimer’s disease is so painful to watch because the disease consumes slowly, eating away at the identity of the victim bit by bit. It is disturbing to watch DID in advanced stages because the recurrence of the original personality becomes a rare occurrence, and thus the degree to which the patient is their old self shrinks. Indeed, it seems these paradigmatic cases of loss of identity clearly show that identity is in fact a misleading term, and what we see in these cases are diminishing degrees of psychological connectedness.

A conception of sameness of self as a matter of degree, flexible enough to accommodate division and fusion, is going to take on a character which might conflict with the conception of diachronic unity of self over time ingrained in the language of one-one identity. Fusion of selves, while arguably an easier case than division, is still neither a clear case of survival nor a failure to survive. We can imagine a successful reversal of the commissurotomy, successful treatment of DID, or a cure for Alzheimer’s. The flexibility of this notion gives us a wide theoretical space into which hypotheses about the nature of the self can be cast, some of which do not fit nicely in the theoretical space implied by ordinary language. I intend to cast my own here: the Multiple-Occupant (MO) theory of the self.

This hypothesis is a combination of the imagined beings in Parfit (1971) and Shoemaker (1988). First, we imagine that a single self can divide into two selves as a result of causal influence from the environment. (There is an obvious way in which this might be true, if the causal influence is the performance of a commissurotomy and successful bodily hemisphere-transplantation.) This is not unlike the split-brain case, and we cannot clearly call this death or identity of self, but there is a distinct degree of continuity among the selves which allows us to say that the self has survived. Second, we state that any two selves with sufficient continuity and compatibility may fuse as a result of causal influence from the environment. Again, the extent to which the resulting single self is psychologically continuous with the previous two will be a matter of degree. During any single time slice SP, the TTS will contain as many selves as have resulted from the various divisions and fusions that preceded the moment in time. The set of selves present during any SP can pleasantly be conceived as “chords” in Gazzaniga’s pipe organ metaphor. All of this happens, as it were, in the same body. Just as severing the corpus callosum results in two distinct streams of consciousness in the same body, so it is with my hypothesis, on an exploded scale. Figure 1 offers an illustration.



**Figure 1.** There are six selves at time slice T, all of which are connected to the initial self by varying degrees.

Each SP, on this view, will be marked by the fusion and division of various branches of selves in the chord. The subject need not be consciously aware that an external causal influence has caused a fusion or division of any self, just as he need not be aware that the chord at SP<sub>1</sub> is not the same as the chord at SP<sub>2</sub>. The whole of experience is the handing-off of an SP from one chord to the next, all under the radar of the subject in question.

## 6. Resolving the Paradox of Akrasia

To understand how the MO theory relates to akrasia and irrationality, we must refer back to the fact that fusions and divisions were posited to occur as a result of external causal influence. What I have in mind here is something like the effects of priming studies. As a toy example, consider the work of Bargh and Pietromonaco (1982). In this study, subjects were consciously primed with negatively and positively valenced words, and then asked to interpret a situation in which the intentions of the actor in the situation are vague. Subjects were more likely to attribute hostile intentions to the actor if they were primed with more negative words. I will argue that the priming results in a division. The division represents two distinct (and possibly contradictory) sets of dispositional properties. Perhaps at time  $t_1$ , subject  $X$  has only one self  $A$ . This self is a set of dispositional properties which always “looks for the good in other people”. At time  $t_2$ , upon being primed by negatively valenced words, there is a division. Now there are two sets of dispositional properties:  $B$  and  $B'$ .  $B$  is exactly like  $A$ ;  $B'$  is like  $B$  in all ways but one—it has the disposition to attribute negative intentions to people. At time  $t_3$  the subject (and thus both  $B$  and  $B'$ ) is shown the ambiguous act and at time  $t_4$  gives a response. The decision that is made at time  $t_4$  is the only means of determining whether  $B$  or  $B'$  acted. Because  $B'$  acted, at time  $t_5$  the subject realizes he acted out of character, and  $B$  and  $B'$  fuse together (and perhaps guilt or shame is a byproduct of this process).

Using this framework, the application of the MO self to problems of akrasia should begin to emerge. Let's take the case of the cupcake over the holidays. At the outset, single self ( $Z$ ) is

committed to the diet. That self splits when primed by the sight of the cupcake. Two sets of dispositional properties result, with one main difference between them: the first self (D) wants to remain committed to the diet, and the second (P) wants to eat a cupcake. (Here it can be mentioned that, if division results in two selves similar in all but this one respect, it is no wonder why we can't "feel" our selves splitting.) When the cupcake is placed in front of me, a choice must be made. If I eat the cupcake, then P did the deciding. Perhaps later (presumably after P and D have fused) I feel guilty about it. Alternatively, if I did not eat the cupcake, then D did the deciding. Later I may look back and realize that, although the cupcake looked and smelled delicious, I am glad I held firm (and this may be the result of fusion after resisting temptation). We often go out of our way to ensure that our past actions fit in with our overall view of ourselves. Confabulation and cognitive dissonance are good examples of this. Perhaps pride and guilt might be other examples.

Finally, it must be noted that this framework is also able to accommodate paradigmatic losses of self. In recollecting past actions, we evaluate how 'in-character' those actions were for us. The extent to which the act seems out of character is one rough and ready way to evaluate the degree of psychological connectedness between the deciding self and other selves. Once this degree crosses below a minimum degree of psychological connectedness, we claim that the self is lost. (Perhaps the chord is out of tune?) Certain traumatic events can prime a division so dramatic that the resultant selves have *very little* in common. These selves are "felt" to be distinct from a first-person perspective, and behaviorally adjudged to be significantly distinct such that a case of DID is diagnosed. Also, prior branches in the longitudinal unity of the MO self can be forgotten. When a significant number of these branches dissolve, we can say that the present self crosses below the minimum degree of psychological connectedness with past selves, and conclude that the diachronic unity of the self has been lost to Alzheimer's disease.

It may be objected here that I have multiplied entities unnecessarily. At least one initial motivation was parsimony, was it not? In fact, I find the MO theory to be trimmed more neatly



under Ockham's razor than the alternative. Whatever any Center theory of the self lacks in combinatorial complexity it more than makes up for in metaphysical baggage. As mentioned before, we have no license to posit the self as any kind of persisting entity. The MO theory merely respects everyday self-experience by not letting our imagination take us well beyond the bounds of sense. Furthermore, I find it a virtue of the MO theory that it can respect the subtlety of akratic action, yet exhibit the firm boundaries required to accommodate paradigmatic loss of self in cases of dementia.

On this view, akratic action is no longer a paradox. Strict incontinent action, or what Audi (1990) calls "last-ditch akrasia", is, strictly speaking, impossible. I eat the cupcake at time  $t_1$  because, of the chord of selves present at time  $t_1$ , the individual self that acted deemed it best to eat the cupcake. However, this is not a bother; paradigmatic cases of akratic action do not refer to last-ditch akrasia. If, of the chord of selves present at time  $t_0$ , the individual self that acted judged it best not to eat the cupcake, and, of the chord of selves present at time  $t_2$ , the individual self that acted judged that eating the cupcake at  $t_1$  was in error (or even if it didn't), then we still have a genuine case of akratic activity that all but the most stubborn will have to accept.

## **Conclusion**

It should be clear at this point that not only is the folk notion of the self unsuitable for cases of akratic action, it is unsuitable as a theory of the self, the very thing it proposes to explain! If the folk notion is derived from everyday self-experience, it takes its share and more. Instead, the MO theory provides a streamlined framework for rational and irrational activity. Furthermore it is vindicated by everyday experience. Finally, it is a theory suitable not only for establishing the diachronic unity of a self (even one that can act akratically) over time, but for establishing the *disunity* of a self in cases of dementia, Dissociative Identity Disorder, and Alzheimer's disease.

## References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Arpaly, N. (2000). On acting rationally against one's best judgment. *Ethics* 110 (3):488-513.
- Audi, R. (1990). Weakness of will and rational action. *Australasian Journal of Philosophy*, 68 (3): 270-281.
- Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, 43:437-449
- Bratman, M. (1984). Two faces of intention. *Philosophical Review*, 93:375-405.
- Broad, C. D. (1925). *The mind and its place in nature*. New York: Harcourt, Brace, and Co., Inc .
- Gazzaniga, M. (2008). *Human: The science behind what makes us unique*. New York: HarperCollins.
- Gazzaniga, M., Bogen, J. & Sperry, R. (1962). Some functional effects of sectioning the cerebral commissures in man. *Proceedings of the National Academy of Sciences USA*, 48:1765–1769.
- Grice, H.P. (1941). Personal Identity. *Mind*, 50(200): 330-350.
- Hume, D. (1739). *A treatise of human nature*. New York: Oxford University Press.
- Jackson, F. (1984). Weakness of will. *Mind*, 93(938): 1-18.
- James, W. (1890). *The principles of psychology*, Cambridge, MA: Harvard University Press.
- Locke, J. (1690). *An essay concerning human understanding*. Menton: Scholar Press.
- Mele, A. R. (1987). *Irrationality: An essay on akrasia, self-deception, and self-control*. New York: Oxford University Press.
- Nietzsche, F. W. (1967). *The will to power*. Walter Kaufmann (trans). New York: Random House.
- Parfit, D. (1971). Personal identity. *Philosophical Review* 80:3-27.

Reid, T. (1785/2002). *Essays on the Intellectual Powers of Man*, Derek Brookes (ed.), University Park: Pennsylvania State University Press.

Rorty, A. O. (1980). Where does the akratic break take place? *Australasian Journal of Philosophy* 58(94): 333-346.

Shoemaker, S. (1988). On knowing one's own mind. *Philosophical Perspectives*, 2:183-209.

Strawson, G. (1999) The Self and the SESMET. *Journal of Consciousness Studies* 6 (4): 99–135.

Walker, A. F. (1989). The problem of weakness of will. *Noûs* 23 (5):653-676.