Nagarjuna's Refutation of Personal Identity

by Tyler Jungbauer

Abstract

In contemporary metaphysics of personal identity, there are two main competing sets of philosophical positions: Reductionism and non-reductionism. The former set of positions claims that personal identity is reducible to some set of sub-personal facts, usually psychophysical facts of some sort, while the latter set of positions contests such a reductionist claim by instead maintaining personal identity to be something irreducible and distinct from those sub-personal psychophysical facts of experience. In his magnum opus, the Mulamadhyamakakarika*, the Buddhist philosopher Nagarjuna (c. 150–250 CE) provides a definitive argument against both reductionism and non-reductionism about personal identity. In this paper, I will argue that the argument set forth by Nagarjuna refutes non-reductionism because this view entails that personal identity must be distinct from psychophysical facts, a result that suggests rather absurdly that knowledge of personal identity could never be ascertained on the basis of the sorts of facts that we are privy to in our experiences of the world. Furthermore, I will suggest that Nagarjuna's argument refutes two common forms of reductionism because both the psychological and bodily criteria of the person reduce personal identity to psychophysical facts, a reduction that implies the contradiction that personal identity is something unchanging and permanent and yet constantly changing and impermanent.*

Introduction

A prominent metaphysical concern in contemporary philosophy is the nature of personal identity. The essence of this concern is discerning what it is that allows someone to persist across time as the same, numerically identical person. In other words, the question of personal identity seeks to identify whether we persist across time as the same person, and if we do, which sorts of facts are relevant to defining this persistence (Siderits *Personal Identity* 1–2). Some philosophers are non-reductionists about personal identity and seek to ultimately reify the person as an irreducible entity that does not admit of reduction to more basic facts, while others are reductionists and suggest that personal identity is best construed as reducible to impersonal facts, usually psychophysical facts. On the other hand, other philosophers take an even more radical approach to this reductionist view and ultimately claim that personal identity does not exist, an approach that is called eliminativism. One such philosopher is the Buddhist philosopher Nagarjuna, whose classical work the *Mulamadhyamakakarika* offers an argument against both non-reductionist and reductionist accounts of personal identity. In specific, in this paper I will argue that the argument set forth by Nagarjuna refutes non-reductionism because (1) this view entails that personal identity must be distinct from psychophysical facts, a result that suggests rather absurdly that knowledge of personal identity could never be ascertained on the basis of the sorts of facts that we are privy to in our experiences of the world. Furthermore, I will suggest that Nagarjuna's argument refutes two common forms of reductionism because (2) both the psychological and bodily criteria of the person reduce personal identity to psychophysical facts, a reduction that implies the contradiction that

personal identity is something unchanging and permanent and yet constantly changing and impermanent.

To argue these points, I will first briefly outline the notion of personal identity as construed in the contemporary debate. I will also show how this notion of personal identity is equivalent in a qualified sense to the Buddhist notion of *svabhava* or essence. After doing this, I will outline Nagarjuna's argument against the existence of the self as laid out in the *Mulamadhyamakakarika*, and then I will show how this argument successfully refutes both non-reductionist and reductionist views of personal identity. Before concluding, I will consider how Nagarjuna's refutation of non-reductionism and reductionism suggests that his view is best categorized as eliminativism.

Metaphysics of Personal Identity: Reductionism and Non-Reductionism

In the contemporary metaphysical debate concerning the nature of personal identity, there are two main schools of thought that have attempted to answer the question as to what constitutes the nature of a person. On the one hand there is non-reductionism, which is the view that persons are ultimately real things that "make up part of the furniture of the universe," as it were, because persons cannot be reduced to other facts (Siderits *Personal Identity* 10). A non-reductionist holds that persons must be included in our "final ontology" of the universe as ultimately real metaphysical entities (9). As Siderits explains, the most basic version of non-reductionism holds that "a person is a separately existing entity, distinct from the person's brain and body and their experiences" (29). All in all, this view construes the person as something above and beyond the psychophysical facts of experience. Instead, the person constitutes the self, or "the agent of actions" and "the essence of the person" (29–30), something that is similar to the notion of a soul.

This view of the self as something essential or distinct from psychophysical facts of the world is similar to the notion of self that Buddhists adamantly reject, the so-called *atman* (Rahula 51). One of the fundamental doctrines of Buddhism is *anatta* or *anatman*, or the claim that the interdependent functionality of the five aggregates generates the *illusion* of an essential or independently existing self because we falsely believe and incorrectly project onto this inter-functionality of the aggregates the idea of some fundamental or essential self that exists above and beyond these more basic facts of the world (26, 55). In Buddhist metaphysics, the five aggregates, viz., matter, sensations, perceptions, volitions, and consciousness, are the fundamental components of being, or the basic psychophysical facts into which all things may be reduced in an ontological sense (20).[1] In other words, they are the building blocks of reality. Furthermore, Buddhist metaphysics construes these aggregates as being fundamentally ephemeral and impermanent (25), so the very idea of an essence, or *svabhava*, becomes highly problematic on this metaphysical view.

[1] There is some dispute in the literature about how to understand Buddhist reductionism, i.e., whether Buddhist reductionist claims are meant to be ontological or purely phenomenological claims. For an account of Buddhist reductionism as a purely phenomenological materialism, see, for example, Francisca Cho, "Buddhist Mind and Matter," pp. 422–434. For an account of Buddhist reductionism as an ontological claim, see Siderits, *Buddhism as Philosophy*, Chapter 3: "Non-Self: Empty Persons," pp. 35–37. In this paper, I will prefer Siderits's view of Buddhist reductionism over Cho's for three reasons: (1) Siderits's view seems to be the more common interpretation of Buddhist reductionism; (2) the focus of this paper is on the ontological status of the person, not on the phenomenology of personhood, so any appeal to Buddhist reductionist claims must be construed on the ontological, not merely phenomenological, level of analysis; and (3) Nagarjuna's reductionist claims appear to be ontological, not phenomenological claims, mainly because his arguments concern metaphysics, not phenomenology (see Garfield 142–143).

Indeed, as Garfield and Priest explain, on the Buddhist view *svabhava* or essence is a thing as it is in itself, independent from all other things, and hence not conditional on any other thing (6). As I will show below through consideration of Nagarjuna's argument, this strict notion of essence precludes the possibility of a self *qua* essence, which exists permanently and independently of all other things. To this end, I will show how this view of essence within Buddhist philosophy is more or less equivalent to the idea of personal identity within contemporary metaphysics.

On the other hand, *contra* non-reductionism, there is the ontological view of reductionism. A reductionist holds that persons are ultimately reducible to more basic impersonal facts, in which case persons do *not* belong in our "complete theory about the nature of reality" as fundamental entities (Siderits *Personal Identity* 9). Rather than holding that persons exist as fundamental ontological categories, reductionists hold that persons are reducible to those psychophysical facts of the world, such as the facts of the five aggregates. To say that something *A* is reducible to *B, C,* and *D* is to say that *A* is nothing above and beyond *B, C,* and *D*, or that *A* just consists of *B, C,* and *D*. Consequently, reductionists about personal identity do not necessarily mean to say that personal identity, or the self, does not exist; rather, what reductionists insist on is that personal identity exists, insofar as the more basic facts into which personal identity may be reduced exist. Reductionists about personal identity, then, mean to say that the existence of personal identity "just consists in the existence of things of a more basic sort," where these more basic facts constitute the ultimate ontological furniture of the universe, as it were (*Personal Identity* 9).

Even though reductionists maintain that persons are reducible to more basic facts, they still maintain the existence of persons as real entities, unlike eliminativists, who also reduce persons to more basic facts like reductionists, but who in contrast ultimately deny the reality of persons as real entities (Hanner 217). Distinguishing between reductionism and eliminativism is really a matter of semantics. In the discussion of personal identity, a reductionist about personal identity will say that ultimately we can reduce discussion of the "self" into more basic facts of some kind, but nonetheless the reductionist will maintain that we should still use the *term* "self" in our ordinary language because of its practical utility; in contrast, the eliminativist, *pace* the reductionist, will concede that we can ultimately reduce the "self" into more basic facts but also deny that we should continue to use the term self in our ordinary language, perhaps even if the term has some practical use.

In this paper, I will focus on two major reductionist views, which consist of two different criteria with regards to what it is that constitutes personal identity. These criteria consist of the psychological facts and the physical facts of a person, respectively, and perhaps also the relations that exist between these facts (Hanner 212). Other reductionist views do exist, but these are the only two I will consider, because they are the most relevant to Nagarjuna's argument against personal identity.

On these views, a person is either reducible to facts about one's body or facts about one's psychology (or perhaps both) (Siderits *Personal Identity* 22). Bernard Williams outlines both of these views in his classic article "The Self and the Future," but ultimately he prefers the biological or bodily criterion of personal identity because he thinks that the bodily criterion preserves the transitive nature of the relation of personal identity, whereas the psychological criterion fails to do so (Williams 179–180). Williams's thought experiment demonstrates that even if a person's memories, thoughts, and beliefs were replaced with those of another, they would continue to be the same person in virtue of having the same physical body. Thus, Williams would argue that person *A* is identical to person *B* if and only if both *A* and *B* share the same body (180). On the other hand, Locke's conception of personal identity as being reducible to consciousness is an example of a reductionist view that uses psychological facts as the relevant criterion for identity (Yaffe 387).

Locke, argues that personal identity is reducible to the continuity over time of consciousness and its states, including beliefs, thoughts, and memories, even if the body should change. In contrast to Williams, then, someone like Locke may contend that a person is the same as another if and only if they share, e.g., the same consciousness, regardless of whether this psychological fact is shared by, or occurs in, the same body. Notice that, while both Williams and Locke reduce the self to more basic and impersonal facts, they still maintain that the self is real, unlike an eliminativist, like Hume, who denies its ultimate existence (Hanner 217).[2]

Personal Identity and *Svabhava*

Now that I have outlined in detail both the non-reductionist and the two common reductionist views of personal identity, I will show how the notion of personal identity is equivalent in a qualified sense to the Buddhist notion of *svabhava,* or essence. As Siderits explains, the notion of personal identity is essentially this idea that we "take ourselves to be the kinds of things that can continue to exist for many years" (1). In other words, the idea of personal identity boils down to the idea that, despite the passage of time, the person we are at time t is numerically identical in some relevant way to the person we are at time $t + 1$. Whatever this relevant fact is that makes us the same despite the passage of time—whether it is some fact above and beyond the psychophysical facts of experience, or whether it is these psychophysical facts themselves—it is something that we must have in order to remain the same person across time. Naturally, the question reduces to which fact(s) is relevant for identity across time.

As I mentioned above, in Buddhist philosophy, the idea of *svabhava* refers to the nature of a thing as being independent of all other things and as therefore existing intrinsically and in itself (Garfield and Priest 6). If personal identity and *svabhava* are really just two terms for the same thing, then it must be the case that whatever has personal identity must have it if, and only if, it is an essence or possesses essence in some way. To demonstrate this equivalence, it must be shown that whatever has personal identity across time also must have essence in the sense of *svabhava,* and that whatever has *svabhava* must also have personal identity across time. To see that whatever has *svabhava* also has personal identity, suppose that x is an essence, so x is permanent, unchanging, and unconditional (Garfield Chap. XV, stanza 1). Thus, at time $t + n$, x is numerically identical to x at any earlier time t because x is unchanging and remains the same despite the passage of time. Hence,

[2] Some philosophers have interpreted Hume as offering a reductionist view of personal identity rather than an eliminativist view. For example, in *The Self: Naturalism, Consciousness, & the First-Person Stance*, Ganeri considers Hume's view of personal identity to be that "the stream of experience collectively provides mental states with both an identity and an owner" (40). In this case, Ganeri seems to construe Hume as saying that the self is *reducible* to the stream of experience, such that the self still exists, but *solely* as the stream of experience. In other words, on Ganeri's interpretation of Hume's account of personal identity, while Hume reduces the self to the stream of experience, he is nonetheless *affirming* the existence of the self in this reduction, instead of reducing the self to these other (sub-personal) facts simply to deny or altogether refute its existence. If Ganeri is correct, then Hume's view of the self would be more similar to Locke's psychological view of the self than it would be to the Buddhist no-self or eliminativist view to be presented below through Nagarjuna's argument against reductionism and non-reductionism. However, other philosophers have read Hume in a more eliminativist fashion; for example, Jay Garfield claims that "[i]n his *Treatise of Human Nature*, David Hume argues not only that we are not selves, but also that we don't even have an *idea* of a self—that when we talk about selves, we are literally talking nonsense" (Garfield *Losing Ourselves* 25). The interpretation here is that Hume thinks personal identity (which is what Garfield means by "self") is a "fiction," or a misapprehension, so the word "self" or "personal identity" does not refer to anything at all. It is just a word, without any meaning whatsoever, so whenever we talk about personal identity or selfhood, we are saying nonsense because we are using meaningless words to say something.

if *x* is essential or possesses *svabhava*, then *x* must have personal identity. In other words, essence or *svabhava* implies personal identity across time.

In contrast, suppose that *x* has personal identity across time, so *x* at time *t* is identical to *x* at time $t + n$ for $n \geq 0$. Despite the passage of time, *x* remains the same, so *x* must exist permanently and unchangingly during the interval of time from time *t* to time $t + n$. The only question that remains is whether *x* exists independently and unconditionally. Note that *x* at time *t* may be numerically identical to *x* at time $t + 1$, but *x* at time *t* may nonetheless fail to be numerically identical to *x* at time $t + 2$ due to certain intervening factors. For example, we might be numerically identical to ourselves for a single day, but perhaps for some reason when we go to sleep at night and wake up the next morning, we are no longer numerically identical with the person we were the day before. If we restrict the length of time of analysis to the single day before, then we would say that I did possess something essential *during that day*, something unconditional and independent of the change that occurred *during that day*. On the other hand, if we consider the day before in light of the entire previous week, then we would say that no such essential thing existed *during that week* because each day saw the construction of a new person.

Thus, whether numerical identity or personal identity implies essence or *svabhava* depends on the length of time that we are considering in our analysis. If we admit that a person comes into existence at some point in time *t*—perhaps at birth, or perhaps at conception, or some other time altogether—and similarly dies at some later time $t + n$, probably at the time of bodily death or brain death, then we cannot say that numerical identity implies the existence of essence, even though someone might have personal identity across the entire interval of time from *t* to $t + n$. This is because the Buddhist notion of essence is strict in its meaning so that essence must exist unconditionally, and this also means unconditionally with respect to time. Otherwise, essence would exist dependently on time, but as Nagarjuna points out, essence that exists dependently is a contradiction in terms (Garfield Chap. XV, stanza 1). Hence, even if someone has personal identity across their lifetime, this still would not qualify as *svabhava* because at some point the facts of this personal identity passed into existence and at another point they will pass out of existence, so such personal identity is still conditional and hence not genuinely essential in the sense of *svabhava*.

However, if we choose to disregard these inflection points, as it were, *t* and $t + n$, and focus solely on the intervening points in time—that is, on the points in time contained within the *open* interval $(t, t + n)$—then we might be able to talk about the existence of an essence in a restricted or qualified sense of *svabhava*. For example, if the non-reductionist is right, and I exist as a substantial or essential self above and beyond the psychophysical facts of my experience, and if we consider this essential self solely within the length of my lifetime, while disregarding the points in time at which I was born and at which I will die, then we can consider this self to be an essence in a qualified sense of *svabhava*, even though we cannot consider it to be an essence in the full sense of the term because ultimately this self comes into, and goes out of, existence depending on other facts of the world. Nonetheless, within the interval of my lifetime, if I do in fact exist as an essential self, then I exist unchangingly, permanently, and hence unconditionally within this interval of time. In a loose sense, then, rather than a strict logical sense, personal identity implies essence, and so in a loose, but non-strict sense, personal identity and essence are equivalent.

The equivalence between personal identity and essence here is "loose" because, for equivalence between personal identity and essence or *svabhava* to obtain, it would have to be the case that *both* essence or *svahava* logically entails personal identity, *and* personal identity logically entails essence or *svabhava*, but since the latter condition does not hold, it follows that personal identity and essence are *not* equivalent in the logical sense of the word. Nonetheless, if we ignore this strict logical

sense of the word, then we can alternatively restrict our analysis to only consider a limited or confined range of time, namely, the range of time defined by the open interval $(t, t + n)$. By doing this, then we *can* say that personal identity implies essence or *svabhava* because *during this open interval of time* anything with personal identity will necessarily be something unchanging, permanent, and unconditionally existent, all of which is to say that anything with personal identity on this interval will necessarily be something with essence or *svabhava* on this interval. Thus, if we hold our frame of temporal reference closed to the interval $(t, t + n)$, then on this interval personal identity logically implies essence or *svabhava*, and, since both conditions of equivalence defined above hold on this restricted interval of time, as a corollary it follows that on this interval personal identity is logically equivalent to essence or *svabhava*. This is what I mean by saying that personal identity and essence are only "loosely," but not strictly, equivalent, because their equivalence is temporally restricted to a specifically defined open interval of time.

Nagarjuna's Refutation of Reductionism and Non-Reductionism

It is finally time to consider Nagarjuna's argument against the existence of self as laid out in the *Mulamadhyamakakarika*. Nagarjuna outlines this argument in a single stanza of chapter eighteen. To wit, the stanza reads:

> If the self were the aggregates,
> It would have arising and ceasing (as properties) [sic].
> If it were different from the aggregates,
> It would not have the characteristics of the aggregates. (Garfield 48)

Nagarjuna's argument is really quite simple as he lays it out. It constitutes a *reductio ad absurdum* argument, or an argument to absurdity, argument in which he presumes for the sake of argumentthat the self *qua* essence exists. As his primary implicit premise, he assumes that if the self *qua* essence exists, then either (a) the self *qua* essence is identical to one or more of the aggregates, or (b) the self *qua* essence is distinct from the aggregates entirely. This premise is assumed because the argument is a *reductio*, so Nagarjuna presumes conditionally the opposite of what he seeks to prove.

Next, Nagarjuna reasons that if the self *qua* essence is identical to one or more of the aggregates, then because the nature of the aggregates is to be constantly changing (Rahula 25), the self *qua* essence must also be constantly changing. This contradicts the assumption that the self is an essence, which is by definition something permanent and unchanging (Garfield and Priest 6). Thus, Nagarjuna rejects the supposition that the self *qua* essence is identical to one or more of the aggregates. Importantly, notice that we can understand this supposition in two different ways, i.e., the self *qua* essence is either equivalent to the *sum* of the aggregates, or it is equivalent to one or more of the aggregates themselves. In either case, since if one of the aggregates is changing, then the sum of the aggregates must also be changing, it does not make a difference here whether we choose to equate the self with the sum of the aggregates or simply one or more of the aggregates themselves. Either way, we are still left with a contradiction.

Since we must reject (a) above, if the self *qua* essence exists, then it must exist distinctly from the aggregates. If this is the case, then the self *qua* essence would lack the characteristics of the aggregates, such as the properties of matter, perception, sensation, etc. There is no necessary contradiction here because we can coherently conceive of, or at least imagine, ourselves as being selves that exists distinctly and independently of all facts about our bodies, perceptions, sensations, etc. This notion is conceivable because it is (tentatively) logically equivalent to the notion of an *atman*

or a *soul*, and most of us would agree that these ideas, while perhaps not being accurate descriptions of our actual nature as beings, are still conceivable and coherent in themselves.In fact, Garfield, the translator of Nagarjuna's *Mulamadhyamakakarika*, has suggested in his recent book *Losing Ourselves* that our default understanding of ourselves is that we exist separately or distinctly from our body and psyche. In specific, Garfield says that we tend to think of ourselves as being "distinct from both [our] mind and [our] body," such that we are "the thing that *has* [our] mind and body, but that, without losing its identity, could take on another mind, another body, just like changing [our] clothes" (Garfield *Losing Ourselves* 9). This idea is nicely illustrated in the *Bhagavad Gita,* as described by Garfield:

> The Indian classic *Bhagavad Gita* (*Song of the Lord*) characterizes the relation between the self and the embodied person as akin to that between you and your wardrobe. Each day you might put on a new set of clothes, but you are still you, the bearer of those clothes; you are not in any sense identical to them, and you are the same individual who put on different clothes yesterday and who may put on new ones tomorrow. Just so, according to the *Gita*, you, the *atman*, put on a new mind and body in each life, but are never identical to any mind or body; instead, you are the *bearer* of that mind and body, which are just as much objects to your subjectivity as any external phenomenon. (Garfield *Losing Ourselves* 3).

The fact that we can make sense of this analogy of the wardrobe suggests that there is no contradiction in thinking that the self *qua* essence is something distinct from the aggregates.

Nonetheless, there is still something rather strange going on when we think this way about ourselves. As Garfield notes in his commentary on the *Mulamadhyamakakarika,* if we choose to bite the bullet and accept this result in order to uphold the existence of a self *qua* essence, then we are faced with the radical outcome that "whatever happens to my mind, body, memory, sensory experience, and so forth, is independent of what happens to *me*" (246). Such an outcome is too bizarre, Nagarjuna suggests, so we must reject it. Consider why this outcome is so bizarre: For one thing, if I exist as a self, distinct from body and mind, then what happens to *me* is completely independent of what happens to my body, or my mind, or my experience, and so it seems plausible on this point to think that even if my mind should be altered (perhaps due to brain injury), or if my body should be destroyed (as occurs during bodily death), or if my experience should be changed in some way (as occurs in schizophrenia or other mental disorders), nonetheless *I* am completely unchanged and unaltered. For someone who accepts the notion of transmigration, it makes perfect sense to think that one's self persists unscathed after the destruction of one's body or mind, but for most of us who do not accept the idea of transmigration, this implication seems too odd to stomach.

It is also not clear how the self could interact with the mind and the body if the self were distinct from these aspects of a person. In the *Milanda Panha,* which is a putative recount of a discussion between the Buddhist monk Nagasena and the Bactrian King Menander, King Menander asks Nagasena if there is "such a thing as 'The one who sees'?", referring to the notion of a self that exists independently of the mind and the body (Bhikku Pesala 50). In response, Nagasena explains that "'The one who knows'cannot be found" because

> O King, the living principle [the one who knows, i.e., the self] within cannot make use of whichever sense it pleases.... It is, O King, by reason of the eye and forms that sight and... other conditions arise, namely; [sic] contact, feeling, perception, intention,

> one pointedness, vitality [sic] and attention. Each arises simultaneously with its cause.... (Bhikku Pesala 50)

Nagasena's point here is that if the "living principle" were independent of the psychophysical facts of the body and mind, then the self as a distinct thing should be able to use any of the senses indiscriminately. For example, if I wanted to *see* with my *ears*, or if I wanted to *taste* with my *hands*, then I should be able to do this because I am a completely independent thing with regards to my body and my mind (50). As an independent thing, I am like a person standing in a room with windows, wherein I can move from one window to another to look outside from any different perspectives. Similarly, if I am an independent self, then I should be able to move around inside my body and mind to use each of my senses and sensory organs however I see fit. But this is clearly an absurd notion: We *cannot* see with our ears, or *hear* with our tongue, and this is precisely because each of these organs *conditions* the sensory forms that arise from it—this is what Nagasena means when he says that "[e]ach arises simultaneously with its cause" (50). I am *not* an independent thing with regards to my body or mind, because my own abilities or capacities are *dependent* on my body and my mind. For one thing, I can only see because I have eyes, nerves, and a brain, as well as a world that impinges upon my retinas, but without these things, I could not see, so my capacity to see is conditional upon the nature of my body, my mind, and even the world itself.

Thus, for these two initial reasons, as well as other similar ones that can be adduced, the self cannot be distinct from the aggregates either. In this way, we must reject (b) above, too, so that we must ultimately conclude that the self *qua* essence does not exist due to Nagarjuna's primary implicit assumption. In this way, Nagarjuna provides a valid argument against the existence of the self *qua* essence.

Let us consider how Nagarjuna's argument can be used to refute both non-reductionism and reductionism. First, consider the case of non-reductionism. As I have shown above, we can equate personal identity with essence in a loose sense, although we cannot equate these concepts in a strict, logical sense due to the considerations above. Nonetheless, if we restrict our analysis sufficiently, we can charitably grant a non-reductionist that they can loosely assert that the irreducible self of a person exists essentially in the sense of *svabhava*. In other words, such a self exists permanently, unchangingly, and unconditionally during a restricted interval of time, and in this way we can say that this self is an essence, but only loosely so. Given that such a self is in fact logically, although perhaps not metaphysically, conceivable, the question becomes whether we have any reasons to reject such a view. Clearly, I think that there are good reasons to reject this view, but in specific Nagarjuna provides a singularly good reason to reject it.

As presented explicitly in the stanza above, and including the suppositions (a) and (b), Nagarjuna's argument demonstrates that the self *qua* essence must either be identical to the aggregates or entirely distinct from them. Instead of talking about the aggregates, we can simply talk about psychophysical facts here because there is really no essential difference between the idea of the aggregates and the idea of psychophysical facts (Garfield 245). Indeed, as Garfield suggests, the aggregates are simply five arbitrary classes into which these psychophysical facts are divided (245). The non-reductionist will deny that personal identity is reducible to these psychophysical facts, so they will consequently deny that the self *qua* essence is reducible to the aggregates, too. This is because we can (loosely) equate personal identity with the self *qua* essence, and we can also equate psychophysical facts with the five aggregates. Hence, the non-reductionist will accept (b) above in Nagarjuna's argument. Therefore, since the non-reductionist's claim that personal identity is distinct from these psychophysical facts is equivalent to (b), the same problems arise with the non-

reductionist's claim as arise with (b). In other words, if personal identity is distinct from these psychophysical facts, then it follows, as it follows from (b), that whatever is the case about these facts is fundamentally independent from what is the case about *me* or my *self*.

In this way, the non-reductionist must essentially agree with Hume and Parfit that the self is inaccessible to introspection (Siderits *Personal Identity* 33). This seems utterly bizarre. Furthermore, if the self is not readily accessible to us through these psychophysical facts, the question arises how we would conclude or infer the existence of the self in the first place (34). We might posit the existence of such a self if making such an hypothesis would somehow allow us to explain some phenomena that we otherwise could not explain. However, early in the *Mulamadhyamakakarika,* Nagarjuna outlines the notion of dependent origination, which constitutes the claim that all phenomena are causally interrelated by a complex network of causal relations (Garfield Chap. I, stanza 1; Rahula 53). Dependent origination, which is the English translation of the Sanskrit term *pratityasamutpada,* did not originate with Nagarjuna, but is a long-standing idea within Buddhist thought in general (Jiang 28). This theoretical notion is commonly expressed in the formulae, "When this exists, that comes to be; with the arising of this, that arises," and "When this does not exist, that does not come to be; with the cessation of this, that ceases" (27). No phenomenon whatsoever exists or arises without some cause or condition for its existence or arising. In this way, dependent origination explains phenomena "by appealing to the conditions upon which they arise," rather than "treating things as either substantive or non-existent" (28). Hence, dependent origination expresses "a worldview that rejects the independence of beings and embraces their interdependence" (28). For this reason, when we talk about an "entity," such as an object, what we commonly mean is something with "substance-*svabhava*," or something that "is not artificially created and not dependent on anything else" (30). In other words, we commonly talk about the world in terms of self-defined, independent *things* or *substances,* like "matter" or "mind." However, Nagarjuna and other Buddhist thinkers replace this object-oriented ontology with a process-oriented ontology; as Radhakrishnan states:

> Causality is always self-changing or becoming. The essence of a thing... is its immanent law of relation. There is no *being* which changes. There is only self-changing or self-forming. We cannot say... that one thing is the cause of something else. For a thing is what it is, and it cannot become something else.... The causal evolution [of phenomena] is not to be viewed as a mechanical succession of movements.... It is the determination of the present by the past.... Causal relations are of the type of the seed growing into the tree, where the one is necessary for the other. (Radhakrishnan 313).

In contrast to the view that causation involves independent entities acting upon each other in causal relations, the notion of dependent origination conceives of causality as being the necessary determination of present conditions by past conditions, or more metaphorically, the present *ripening* of past conditions, like the ripening of a seed into an oak tree.

Seemingly, then, Nagarjuna would not assent that there is any good reason to think we need to posit such a self instead of simply referring to this causal network to explain the phenomena of our experience. If by hypothesizing the self we could explain some phenomenon that otherwise we could not explain equally well without hypothesizing the existence of the self, then we would have a *prima facie* reason to think that the self exists. However, by appealing to dependent origination, Nagarjuna effectively denies that there is *any* phenomenon that we cannot explain at least equally well in terms of the causal relations described by dependent origination, rather than by positing the existence of the self. In this way, Nagarjuna would likely agree that "the grounds for such an

inference [of the existence of the self *qua* essence] can instead be explained by causal relations among entities and events" (Siderits *Personal Identity* 42). For this reason, there does not seem to be a good reason to grant the non-reductionist the need for hypothesizing the existence of a self *qua* essence, especially when this hypothesis means that we must posit a metaphysical category that is irreducible to other metaphysical facts that justifiably exist. These considerations seem sufficient to undermine the non-reductionist's claims about the self *qua* essence.

Let us now consider alternatively the two reductionist views of personal identity outlined above, those of the bodily and psychological criteria, respectively (Siderits *Personal Identity* 21). Nagarjuna's argument allows us to refute both of these reductionist views by the same argument: Both the bodily criterion view of personal identity, and the psychological criterion view, equate personal identity with some psychophysical facts of experience. Insofar as personal identity can be construed as identical to essence in a loose sense, if personal identity is in fact equivalent to psychophysical facts, then we can say that the bodily and psychological criteria of personal identity are consequently bodily and psychological criteria of the self *qua* essence. In other words, we can simply exchange personal identity for the self *qua* essence. Similarly, the aggregates of Buddhist metaphysics are once again equivalent to what we would simply call psychophysical facts in modern philosophical discourse, so we can reduce the bodily criterion to the equivalent claim that the self *qua* essence consists fundamentally of the "endurance" across time of the aggregate of matter (Hanner 212). We can say the same things, without altering the main point of the argument, about the psychological criterion of personal identity. A reductionist who maintains that personal identity is reducible to some psychological facts about a person can be construed as maintaining equivalently that the self *qua* essence is reducible to the relevant mental aggregates of Buddhist metaphysics, such as the aggregates of perception, sensation, or volition.

Consider (a) of Nagarjuna's argument against the self *qua* essence. He reasons that if the self *qua* essence is identical to at least one of the aggregates, then this gives rise to a contradiction because the nature of the aggregates is to be constantly changing, while the nature of the self *qua* essence, if it exists, is to be permanent and unchanging. Similarly, if personal identity consists of a psychophysical fact, then for the reasons Nagarjuna gives in his argument the nature of these psychophysical facts must be the same as the nature of personal identity. Since the nature of the self *qua* essence and personal identity is to be unchanging and permanent, this must also be the nature of the psychophysical facts. On the other hand, if the nature of personal identity is to be unchanging and permanent, while the nature of psychophysical facts is to be changing and impermanent, then we certainly cannot equate personal identity with some psychophysical fact without contradicting ourselves. The question, then, is whether the nature of these psychophysical facts is consistent with the nature of personal identity.

To answer this question, consider what Nagarjuna says of the nature of phenomena: "All things lack entitihood, / Since change is perceived" (Garfield Chap. XIII, stanza 3). As Garfield notes, Nagarjuna means to suggest that nothing can have an essence, or some fundamentally permanent and unchanging nature, because if something did, then that thing could never change (209). But *everything* changes—all of our experience is flux and mutation of phenomena. Our bodies are constantly replacing themselves as cells are recycled and destroyed, and our minds are nothing but shifting streams of thought, sensation, perception, and so on. Therefore, it cannot be that essence exists, because there is a total absence of anything permanent, unchanging, and unconditional in our experience of the world. Instead, everything that exists must be comprised of those psychophysical facts that consist of "ever-changing physical and mental forces or energies" (Rahula 20), or "a complex causal series of impermanent, impersonal psychophysical elements"

(Siderits *Personal Identity* 37). These considerations are meant to suggest that the nature of the aggregates, as well as the nature of psychophysical facts in general, is to be changing, impermanent, and only conditionally existent.

Thus, if we equate personal identity with the bodily criterion, or physical facts about ourselves, then since the body is constantly changing, our personal identity, or the self *qua* essence, must also be constantly changing. The same conclusion follows, without altering the main point of the argument, if we equate personal identity with the psychological criterion, or psychological facts about ourselves. Furthermore, the same conclusion *also* follows if we equate personal identity with *both* facts about the body and facts about our psychology, or psychophysical facts in general. In this way, we cannot equate personal identity with psychophysical facts without contradicting ourselves because the nature of personal identity, or the self *qua* essence, just turns out to be inconsistent with the nature of psychophysical facts, or the aggregates. For this reason, (a) of Nagarjuna's argument against the self *qua* essence undermines the reductionist views of personal identity that attempt to define identity in terms of psychophysical facts about an individual.

Conclusion: Nagarjuna's Eliminativism about Personal Identity

After considering the views that Nagarjuna's argument undermines, namely non-reductionism and reductionism, the question certainly remains how we might classify Nagarjuna's own view of personal identity. Recall the distinction between reductionism and eliminativism above: After reducing personal identity to basic ontological facts, reductionism seeks to provide an account of the self or personal identity, while eliminativism does not bother to provide such an account of these things because it rejects their existence outright (Hanner 217).

While Nagarjuna does affirm the conventional existence of the self (Garfield 187), his argument outlined above—that is, his argument as explicitly stated in the stanza from chapter eighteen of the *Mulamadhyamakakarika,* and including suppositions (a) and (b) as described above—isclearly intended to undermine the ultimate reality of the self as conceived by both reductionism and non-reductionism. This conclusion should be clear because (1) Nagarjuna rejects that the self is equivalent to one or more of the aggregates, and (2) Nagarjuna rejects that the self is equivalent to something *other than* one or more of the aggregates. If Nagarjuna were a non-reductionist, then he would affirm the claim that the self is equivalent to something that is distinct from one or more of the aggregates, while if Nagarjuna were a reductionist, then he would affirm the alternative claim that the self is equivalent to one or more of the aggregates. But Nagarjuna does not affirm either of these claims, so it follows that Nagarjuna is *neither* a non-reductionist *nor* a reductionist. In this way, Nagarjuna refutes the existence of the self both as some final ontological category to be included in our final picture of the nature of reality, and as something that still exists but is reducible to more basic psychophysical facts of the world. Nagarjuna also does not attempt to provide some alternative account of the self after refuting its essential existence, and this fact gives further credence to the claim that Nagarjuna is not a reductionist. For these reasons, Nagarjuna appears to be an eliminativist, insofar as he refutes the ultimate existence of the self but does not bother to provide any alternative account of the self after denying its existence. In conclusion, rather than providing a reductionist or non-reductionist account of the self, Nagarjuna's *Mulamadhyamakakarika* provides a compelling eliminativist account of the self.[3]

[3] Even though we might conclude that Nagarjuna is an eliminativist because he is neither a non-reductionist nor a reductionist about personal identity, we should be careful in ascribing the metaphysical doctrine of eliminativism to Nagarjuna, insofar as he is, on most interpretations, a philosophical skeptic. This is to say that Nagarjuna really wants to

Why does any of this matter? Commonly, many people see philosophical reflection as being too esoteric or abstract to warrant any sort of practical utility. Whether this is the case with regards to other metaphysical issues, I cannot say, but in this specific case, with regards to the question of whether the self exists, there *do* seem to be real, practical benefits, or at least implications, involved in rejecting the existence of the self in the manner that Nagarjuna does. One practically relevant implication that appears to follow from the rejection of the selfis the ethical value of this change in perspective. This is because the *acceptance* of the existence of the self "distorts our sense of agency, and distorts our attribution of agency to others," because this false view "leads us to see ourselves and others as free agents instead of embedded agents, leading to reactive attitudes that fail to reflect reality" (Garfield *Losing Ourselves* 171). When we see ourselves as being *selves*, we conceive of ourselves as being radically free, as being fully responsible for our actions, while if we reject the notion of the self, we can no longer see ourselves in this way. While perhaps responsibility might make us seem more reflective and rational, it is arguably also alienating, insofar as each of us is wholly responsible for the creation of ourselves, in which case our failure to create ourselves in some way, or our failure to be the sort of person we want to be, is solely and completely our own fault, and not the fault of anyone else. If we want to reject this view of responsibility, if we want to cut ourselves some slack, as it were, when it might come to our own personal failure in being the sort of person we want to be, then we can reject the self to see ourselves as selfless beings who are intimately related to the world at large and to one another.

Some people might see this as being a negative implication of the view that the self does not exist, but there are others—myself included—who see this implication as liberating. It reduces the amount of responsibility we place on ourselves and others when we choose—or, perhaps more accurately, when we are *caused* to choose—to act in a certain way. On this view, rather than being free agents who are causally unconstrained by the world at large, we are instead organisms that exist only in relation to the causal nexus of the world. More importantly, on this view, we are not beings who are fundamentally separate from the *world*, nor is the world itself fundamentally separate from *us*. We and the world exist in an intimate embrace, a cosmic dance, of interdependence. In sum, then, by viewing ourselves this way, we can understand our actions and ourselves as being historically embedded, environmentally determined, and socially influenced. Moreover, we can recognize the true extent to which we interdependently exist in relation to one another and to the world—we can recognize ourselves as interdependent, not as *independent,* beings.

reject the very project of metaphysics *per se* because he wants to abandon the very notion of ultimate truth, or ultimate reality, in itself (Siderits *Buddhism as Philosophy* 191). Given Nagarjuna's claim that ultimately true metaphysical facts do not exist, it is difficult to say that Nagarjuna would agree that *ultimately* there is no self, or that *ultimately* eliminativism is true. This is because, if he agreed that eliminativism is ultimately true, or that ultimately there is no self, then either of these claims would imply the existence of a metaphysical fact and hence the falsity of the claim that there is no ultimate truth. Consequently, when I say that Nagarjuna is an eliminativist, what I mean is that Nagarjuna rejects the ultimate existence of the self, but I do not mean that Nagarjuna would agree with the statement that it is ultimately true that there is no self. In fact, Nagarjuna would agree that neither the claim that the self ultimately exists, nor the claim that the self does not ultimately exist, is ultimately true. However, Nagarjuna would likely agree that one of these claims is a *better* way of thinking about the world. That is, Nagarjuna would likely agree that it is better to think that the self does not exist because we can live better lives if we think of the world in this way, but even in thinking this way, we must be careful to not also think that it is ultimately true that the self does not exist (Siderits 206).

References

Bhikku Pesala, editor. *The Debate of King Milinda: An Abridgment of the Milinda Panha.*Revised edition, Penang, Inward Path, 2001.

Cho, Francisca. "Buddhist Mind and Matter." *Religions*, vol. 5, no. 2, 2014, pp. 422–434.

Ganeri, Jonardon. "Conceptions of Self: An Analytical Taxonomy." *The Self: Naturalism, Consciousness, & the First-Person Stance*. Oxford University Press, 2012, pp. 35–49.

Garfield, Jay. *Losing Ourselves: Learning to Live Without a Self*. Princeton University Press, 2022.

Garfield, Jay, translator. *Mulamadhyamakakarika: The Fundamental Wisdom of the Middle Way*. By Nagarjuna, Oxford University Press, 1995.

Garfield, Jay and Graham Priest. "Nagarjuna and the Limits of Thought." *Philosophy East and West*, vol. 53, no. 1, 2003, pp. 1–21.

Hanner, Oren. "Buddhism as Reductionism: Personal Identity and Ethics in Parfitian Readings of Buddhist Philosophy; from Steven Collins to the Present." *Sophia*, vol. 57, no. 2, 2018, pp. 211–231.

Jiang, Tao. "Incommensurability of Two Conceptions of Reality: Dependent Origination and Emptiness in Nagarjuna's MMK." *Philosophy East and West,* vol. 64, no. 1, 2014, pp. 25–48.

Radhakrishnan, S. "The Ethical Idealism of Early Buddhism." *Indian Philosophy*, Vol. 1, 2nd ed., Oxford University Press, 2008, pp. 286–402.

Rahula, Walpola. *What The Buddha Taught*. 2nd ed., Grove Press, 1974.

Siderits, Mark. *Buddhism as Philosophy: An Introduction*. Hackett Publishing, 2007.

———————. *Personal Identity and Buddhist Philosophy*. Routledge, 2015.

Williams, Bernard. "The Self and the Future." *The Philosophical Review*, vol. 79, no. 2, 1970, pp. 161–180.

Yaffe, Gideon. "Locke on Consciousness, Personal Identity and the Idea of Duration." *Noûs,* vol. 45, no. 3, 2011, pp. 387–408.