Master's Culminating Examination Papers:

Continuously Recognizable Self-Perpetuation and Identity Paradoxes
& Mādhyamaka Ontology and the Problem of Ultimate Analysis
& Daoism and Computation

by Carl Matthew Johnson

Prepared at the University of Hawai‘i at Mānoa, 2006–2008.

Thanks to my wife, Sarah, who has tolerated
my world wandering ways.

“For with much wisdom comes much sorrow;
the more knowledge, the more grief.”
— Ecclesiastes 1:18

⁂

『識得本心本性
　正是宗門大病』

“To know the original Mind, the essential Nature,
This is the great disease of our religion.”
— 佛祖綱目卷第三十八

⁂

『古人の跡を求めず、
　古人の求めしところを求めよ。』

“Seek not after the footsteps of the ancients;
seek what they sought!”
— Bashō

⁂

Life makes no sense; philosophy is a persistent attempt at covering this up.

Table of contents:

Continuously Recognizable Self-Perpetuation and Identity Paradoxes

I. Introduction

In the discussion of vague objects in his *Paradoxes*,[1] R. M. Sainsbury presents a modified version of the sorites paradox directed at the human being: A human being still exists even if he or she loses one molecule from his or her body. Losing one molecule cannot make the difference between existence and non-existence for a human being. Therefore, even if all your molecules are taken away one at a time, you will still exist even if none of your molecules do. However, it is absurd to suppose there could be a human being without a body; therefore, human beings do not exist.

In this same section, Sainsbury presents the familiar example of the ship of Theseus which is slowly replaced plank by plank. The reader is asked, "Did Theseus' ship survive?" If one had made a new ship of the discarded planks, "Does this have a better claim to be the original ship of Theseus?"[2]

In this paper I will show that both paradoxes can be defanged by creating more precise definitions of "human being," "ship," and other relevantly similar composite objects that are based on continuous self-perpetuating processes rather than static substances. The use of a more precise definition for these types of objects does not rely on any particular exotic ontology being the correct underlying metaphysical reality and is quite independent of whatever the particulars that underlie our world turn out to be. Once this more precise definition is in place, I will compare my solution to the problem with similar proposals about the nature of diachronic identity from the perspective of ontology and personal identity. Finally, I will show that paradoxes of the sort given by Sainsbury can be safely sidestepped by denying their assumptions.

---

1. Sainsbury, 49.
2. *Ibid*., 47.

II. Essential features of commonsensical notions of identity over time

A. The problem with using formal identity

Underlying the two paradoxes presented so far is the central question, How can there be things which are not the same as their material composition yet which nevertheless rely on their material composition for their existence? Another pertinent example in this regard is the river. Since at least the time of Heraclitus,[3] philosophers have been familiar with the problem with supposing the material identity of rivers. A river cannot be just what makes it up at a particular time, since if we say that a river is the matter that makes it up, then we must agree with Heraclitus that one cannot step in the same river twice, because a river's material composition is in constant flux. At the same time, we commonly speak of rivers as though they were things that persist for years and years.

To begin, we should note that "same" by itself is an ambiguous term. In our most typical usage of "same," we only mean "same in certain respects." For to say that a river at one time is not the *exact* same in all its properties as the river at another time is tautological. Obviously, a river that exists at a later time has at least one different property than the first river—namely, the property of existing at a different time. Even if one were to discount existing at a particular time as a relevant or meaningful property, there is still the issue that if anything else in the universe has moved between the two times, then the first and second rivers are different with respect to their distance from the objects in the universe that have moved. Of course, such properties as the distance between one river and everything else in the universe are clearly not essential, intrinsic properties for the identity of a river, but the problem at hand is to determine which properties of the river are those that are essential for its being considered the same according to our common usage of the term. Thus, the "sameness" of being the same river (or the same anything else) at two different times must not mean "same in all

3. Note that the famous quotation, "you cannot step twice into the same river" is from Plato's *Cratylus* and only obliquely attributed to Heraclitus. See Kahn 166–167 for a discussion of the origin of the quote and the degree to which it reflects Heraclitus' own attitudes.

respects" but only "same in certain relevant respects," and our task now is to discover what those relevant respects are.

One way to resolve this ambiguity and solve the problem brought up by Heraclitus is to say that ordinary references to sameness in rivers refers to sameness with respect to one of the river's formal physical properties even if its material make up varies. However this quickly runs into difficulties. Imagine that there existed perfect teleporting machines that could duplicate objects down to their most minute physical properties and that we used such a machine to make a perfect physical duplicate of the Earth's Colorado River on Mars. Each water molecule in the teleported duplicate would be positioned as the exact analog of our terrestrial river, and the banks as well would be perfectly copied. Nevertheless, in spite of being in every way like the Colorado River, our duplicate river would not be the Colorado River for the simple fact that it is located on Mars and not on Earth, which is commonsensically essential to the identity of the Colorado River.

Accordingly, we might next propose that geography or physical location is essential to the identity of the river. Unfortunately, this will just as quickly run into problems, since a particular river's banks are constantly changing and being eroded. Returning to our example, the Colorado River's banks were at one time on the surface of a plain, but now they are the bottom of a canyon. Thus, the formal characteristic shared by a river over time cannot be just the shape of its banks, since they are not the same over time. Nor can it be the generalized path of the river, since this is also subject to radical change, as when a river breaks through and finds a new course or when a river alternates between meandering and bypassing oxbow lakes.

One might protest that membership in certain classes allows some flexibility. For example, something can be considered "triangular" for everyday purposes even if it is not precisely composed of three straight lines, but instead it is composed of atoms lined up within a certain degree of tolerance. The river problem though does not allow

recourse to geographic formalism by means of increased tolerances. Consider a case in which two rivers happen to run parallel to one another from source to the sea while separated by only a few meters. In this case, we would say there are two rivers, not one. On the other hand, if there were just one river and it were to shift over by a few meters such that its new course is parallel to its old course, it would still be one river, not two, in spite of having moved over by a distance as large as the distance that separated the other two rivers. Thus, merely adding tolerances is not sufficient to resolve the problems with regarding geography as the essential formal property of rivers. Tolerances are sometimes useful as means of defining classes of objects but insufficient when distinguishing between individual objects like rivers.

A river also cannot even be any such continuous stream of water that shares a common origin, since it is not uncommon for one source of a river to dry up, but for the river to continue its existence by drawing on water from other sources. Even if the Mississippi were to lose the contribution of its own headwaters or of the Ohio or Missouri rivers, we would still call the remaining waters that empty into the Gulf of Mexico the Mississippi River so long as it maintained its integrity as a stream. So long as the main stream continues to flow, a river can still be the same river even if fed by different waters.

Thus, none of the formal properties listed so far—particle properties, geographic location, or head water source—can fully account for our understanding of rivers.

B. The problem with radical nominalism

One response to these many difficulties is to abandon all hope of coming up with a precise definition of river or ship or human being and to switch to a form of radical nominalism about the issue. One might say that a river is whatever people say it is (or worse, whatever people in power say it is) and there's nothing more to it than that. If

we cannot come up with a precise definition of river, then the problem is with the natural imprecision of human languages and nothing can be done to fix it.

This pessimistic account does contain some correct observations, but on the whole it sacrifices too much. Yes, it is difficult to come up with a precise definition of river, and natural language is optimized for properties other than precision, but nevertheless there are cases where it is necessary to provide precise definitions. Even ignoring the paradox that began this discussion, suppose that a river runs through two states and that the two states form a contract that each is allowed to use a certain amount of water from the river. Then suppose that a flood causes a shift in the course of the river, and afterwards one state suddenly begins to take much more than its share of water while arguing, "Who is to say we are in the wrong? The contract we formed was about a certain river, but that river is gone now. The contract can no longer bind us since its object does not exist now if it ever did." Certainly, we would like to have recourse to a precise means by which to settle such a legal dispute. On what grounds can the wronged state claim that the river still exists?

However, even this does not capture the full problem of pure nominalism. Of course words naturally mean just what we as a society want them to mean. Thus the nominalist is correct that it may be the case that our use of some terms is so jumbled that their contradictions and ambiguities cannot be removed. Accordingly, if we wanted to define a river as being the same only if it had the same atoms or the same particle properties or the same banks or same general location or whatever else, we would be entitled to stipulate the use of such a definition. The goal here, however, is to produce a definition that allows us to preserve our commonsensical understanding of terms while still being able to reason rigorously using these terms. The river example is just a test case for a broader class of similar objects that exist as sets without fixed membership. If we cannot define river, then it is likely that we will be even harder pressed to define a number of other terms. The end result will be the inability to make more than

suggestive claims about individual objects in the world. For this reason, it is worth making every effort at capturing an appropriate definition for a term before resigning ourselves to linguistic chaos.

C. The problem with simple co-material identity

One way to solve the river problem might be to propose a theory of co-material identity. In this theory, as long as a certain arbitrarily defined percentage of material from the original composite object at time 1 remains in the object at time 2 then the reconstituted object is the same. Thus for the ship of Theseus example in which there is a "modified ship" that was in continuous service and a "reconstructed ship" made of cast off parts, the modified ship is the same ship so long as at least a particular number of original planks remain. It will then be up to the users of the word "same" to determine whether for their purposes "same" requires just one original plank or at least half of the original ship or some other specified percentage of the ship's composition be preserved. The difference is arbitrary and can be disambiguated as needed, eg. for legal purposes connected to the ship's docking rights one plank might be sufficient, but for the purposes of a museum the "same" ship should be composed of at least three quarters original material.

There are two obvious problems with co-material identity. First, it is not clear whether a ship constructed out of discarded planks would have the same right to the title "ship of Theseus" as the modified ship. It is possible that our definition of sameness will sometimes allow splitting (see below), but we want to avoid this possibility where possible. Another difficulty would come if the modified ship and discarded parts ship both contained fifty percent of the original material. It might seem in this case that which of the two is called the ship of Theseus can be changed by taking a single plank from one ship and putting it on the other and thus changing the balance of parts.

However, the exchange of one plank seems like too small of a change for identity to be passed with it.

If the ship of Theseus example is a difficult one for the co-material theory of identity, then the river example is deadly. It should be clear enough that the Colorado River is rapidly emptying into the ocean. Whenever it would be that we designate the reference set of material in the Colorado River by which to make later comparisons, after a certain amount of time, more of the original river would be in the ocean than contained in what we take to be its banks, and after a little more time none of the original river would be left. Thus simple co-materiality is not enough to give a rigorous definition of commonsensical identity.

D. Continuous recognizability

Organ donation example

Before proposing a new model to solve the various problems listed so far, it might be useful to give some examples that demonstrate what our common sense's requirements for identity are.

Consider for the sake of example a case in which one person, Alice, donates her kidney to her friend, Berta. After the operation, we will naturally want to say that legally the kidney is now Berta's even though historically we can recognize that it was once Alice's. That is we can say that Berta's new kidney is the same as Alice's old kidney. (Also we will say Berta is still Berta in spite of her now containing a new organ. Thus, she is, in one sense of the term, the same person after the operation.) Furthermore the one kidney will continue to have the same two properties (legally being Berta's and historically being Alice's) even if after years of being in Berta's body all of the atoms within the kidney just happen to be replaced by new atoms from the food that Berta eats. Much as Berta retained her identity even with a new kidney, the kidney retains its identity even after its molecular parts are replaced. Thus, we can say that neither property is strictly based on material identity. Perhaps we might then propose that the

identity of the kidney is based on one of its particular formal properties—in this case that it has Alice's DNA patterns even if the atoms making up that pattern are different.

However, as was the case with the river example, formal properties are not a reliable carrier of identity. Suppose that Alice and Berta had been identical twins with identical genomes. In this case the formal property that we are using to distinguish the kidney as being historically Alice's is no longer sufficient to distinguish parts of Berta's body that have the property of being historically hers from those that do not. Thus, if we wanted to say that a person is the same over time because they have the same DNA patterns, we would be unable to distinguish between identical twins and would have to hold that cells within a person with minor DNA mutations are not a part of the same person.

To give a slightly macabre example of the difficulties of formal properties, suppose after removing Charlie's kidney instead of just implanting the organ in Derek mad scientists feed Charlie's kidney to Derek and these atoms from Charlie's kidney just so happened to replace all the atoms in Derek's kidney by moving to places in Derek's body analogous to their former place in Charlie's. Unlike the case of Alice and Berta, we would not say that Derek's new kidney is the same as Charlie's old kidney even if it contains the same material and happens to take on a shape similar to its old shape, because we would hold that at the time of digestion Charlie's kidney was destroyed and any new kidneys that come about cannot inherit the identity of the destroyed kidney. We would only say that the atoms in Derek's new kidney are the same as the atoms in Charlie's old kidney. When those atoms are replaced by Derek's ongoing metabolic processes, we will no longer be able to accurately speak of Derek's current kidney having historically been Charlie's. Thus, the sameness of a kidney over time is not merely its physical shape just as it was not its DNA.

The difference between Berta's new kidney and Derek's new kidney is illustrative of the operant properties that underlie commonsensical identity. Berta's kidney is

the kidney that was Alice's in part just because we could have theoretically watched the kidney continuously existing as the same kidney even after it was moved into place, whereas Charlie's kidney is visually destroyed during digestion. So at least partially, our commonsensical notions of identity fall out of the particular visual object recognition heuristics that are present in the human brain. Accordingly, one model of sameness that we might employ is continuous recognizability. If we could (at least in theory) watch something throughout its transformation and recognize its identity the whole time then it is the same as the original thing. The problem with continuous recognizability as a theory of identity is that it is just a reification of our prejudices. That is, if the question is, "Is this the same object as the one that was here before?" the only advice that continuous recognizability gives is, "Yes, if you continuously thought so. No if not." Thus, it is only useful in the cases where we can already recognize sameness, but it does not tell us how the recognition process actually works. Recognition itself is a black box function, and identity is just the sameness of the function's output over time. What it is that we are recognizing is unknown.

The problem of splitting

The difficult with leaving recognition as a black box is that there are cases where our intuitions seem to clash, and continuous recognizability does little to resolve the problem. Consider an acorn and the oak it becomes. We will say that the acorn is the same as the oak, in part because we could theoretically watch the acorn become the oak, as is done in time lapse photography videos. If we watch further, we might see the oak cut down and turned into a plank of lumber. However, we might be reluctant to say that a desk made from the plank is the "same" as the acorn in a manner analogous to the way that an oak is the "same" as its originating acorn. The sameness of the oak to the desk is purely a material sameness, but the sameness of the oak to the acorn is a different sort of sameness, such that transitivity does not hold between the plank and the acorn. Thus, although continuous recognizability holds from the acorn to the oak to

the plank, "sameness" can be broken into two different sorts, one type of which holds from the acorn to the oak, the other type of which holds from the oak to the desk. Without understanding what it is that we are recognizing, it will not be possible to differentiate these two kinds of sameness.

Similarly, consider a variation on the river example. The river that is the Missouri River in Montana is the same river as the Mississippi River in Louisiana, and the river that is the Ohio River in Pennsylvania is the same as the Mississippi River in Louisiana, but the Missouri River is not the same as the Ohio River in certain relevant respects. If possible, we want our definition to allow for the non-transitive sameness of rivers and tributaries, but doing so on the basis of just continuous recognizability is impossible, since someone walking down from the Missouri River's banks along the Mississippi and on to the Ohio River will recognize what is being seen as the same river throughout the journey. The changes that make the difference between the Missouri and the Ohio (the fact that the streams flow into a common outlet but not each other) will not be picked out by continuous recognizability alone.

E. Self-perpetuating identity

One way to fill in the recognition gap is to say that what is recognized is the self-perpetuation of a particular pattern. Consider the river example. If we define a river as a process by which the gravitationally influenced motion of certain water molecules causes other water molecules to follow them in a similar manner then we can solve many of the problems concerning the sameness of rivers. If we have a river where molecule A is replaced by molecule B, we can recognize the two of them as participating in the same river if we see that it was the gravitationally compelled motion of molecule A that caused the motion of B, which in turn will allow the motion of its successor C. This sameness will hold even if the river changes its shape, so long as the changes do not interfere with the process by which the void left by water molecules

moving due to gravity is filled by other molecules of water that are also gravitationally compelled. Similarly, the river's source changing is fine as long as the water that enters the river from the new source is similarly compelled to run downhill by the opening of space left by the vacancy of its predecessors. The identity of the river does rely partially on co-materiality, but only from moment to moment within the self-perpetuation of the process. Even if all of the river's present molecules are someday gone the river will remain if there is an unbroken chain of causality connecting the current molecules to the molecules that will someday inherit the title of "the same river."

Next consider the ship of Theseus. In this case, we can define a "ship" as the material that inherits the process of being treated as a seagoing vessel by a crew or holding a certain form. Note that the ship need not always be at sea or have a crew, just that it must be either treated as a ship by some community or be a collection of materials that base their shape on the prior shape of the ship. That is, as long as the ship is at sea and being used by its crew, it is the same vessel even if the crew replaces some of its parts, so long as they continue to use the new set of parts as a ship, and when it is uncrewed, it is fine for some of its material to be replaced by new material (eg. through planks rotting), so long as the new material bases its shape on the old within a certain degree of tolerance. Thus, so long as the process by which the parts of the ship are replaced is slow and continuously a part of the process of using the ship as a ship or the ship be weathered by the elements, it is the same ship. Hence taking a plank off the back of a ship and nailing it to a different ship does not make the other ship "the same ship" as the first, since the plank was not being used as a ship during the time between its removal from the first ship and its addition to the second. Similarly, constructing a ship out of junkyard scraps is not sufficient to inherit the title of the same ship.

One consequence of the proposed definition of identity in ships is that if a ship is dry docked, taken completely apart, and left for scrap, but then reassembled using all of and only its original parts by an unaffiliated crew, then it is no longer the same ship,

whereas it may be necessary in the normal service life of a ship to take it apart and reassemble it in a similar manner, but because of the continuity of treatment by its crew, this second ship will be considered the same despite the superficial similarity of its treatment. A crew can reassemble the same ship, but archeologists cannot. This consequence holds whenever the ship is dismantled to the point that it is no longer called a "ship," but rather "a collection of ship parts." It may seem counterintuitive, but I hold that this result is actually a useful feature of my proposed sharpening of the definition of identity for certain composite objects. While it may be somewhat controversial, common sense will not totally reject our saying that a newly constructed ship is a new ship, even if old parts were used. This is clearly a case where our intuitions are weak. Part of the continuity of sameness in retrofitted ships is that they have a continuous treatment as a ship by some community. Lacking that treatment, the identity of the ship parts as parts of a particular ship, rather than parts in general, evaporates. That the sharpening of our definition lets us make a meaningful decision about this potentially controversial choice is a benefit of our proposed definition, not a shortcoming, since we can now make definitive pronouncements about questionable cases.

Consider again the example of the kidneys. Our proposal is that a kidney is the same kidney so long as the biological processes that cause an organism to sustain its existence are still in effect. As with the reassembled ship, we hold that kidney that just so happens to be reconstituted is not the same kidney. Here our intuition more closely aligns with the result of the proposed definition, giving more confidence to the reassembled ship example. Similarly, we allow that the kidney is the same kidney that used to be in someone else even now that it is implanted in its new host. Similarly, the patients can be considered the same people after their surgeries, even if they receive new organs with different atoms, different DNA, etc., so long as their life processes as human beings continue without interruption.

With the example of the oak, we may now distinguish clearly between the biological sameness of the acorn and the tree and the material sameness of the oak and the plank by defining biological sameness according to the self-perpetuating properties of cells, and material sameness as the self-perpetuation of atoms. (Note that since quantum mechanics tells us that atoms contain numerous "virtual particles" that pop in and out of existence every second and help give the nucleus its shape, it is not practical for us to require the material sameness of the plank to hold down to the subatomic level. Instead it must merely hold to it down to the atomic level.)

Finally, we can at last pick out the difference between the Missouri River and the Ohio River. When the water in the lower Mississippi River flows out, it leaves a void that water from both of the rivers rushes in to fill. However, there is no unidirectional causal chain between the Missouri River and the Ohio. Water evacuating from the one river in no way contributes to the evacuation of water in the other river. Thus, we can distinguish between the two rivers, even though they become a part of the same river further downstream.

Note the interesting result that the stipulative definition of identity provided here explicitly does not have the properties of transitivity and symmetrically as the normal definition does.

III. Comparison of proposal to other accounts

The account of commonsensical sameness given here is original but by no means unprecedented. There are two broad categories of antecedent descriptions: those that deal with the ontology of flux and those that deal with the status of personhood in a world of flux. Of course, it is impossible to do full justice to the vast number of prior accounts of identity over time in the space provided here, but giving a few representative explanations will be useful for demonstrating through contrast the strengths and weaknesses of the account I have provided. Of the many accounts dealing with the

ontology of flux, I will restrict myself to Aristotle's theory of substance and the contemporary theory of genidentity. Of the many possible accounts of personhood that could be contrasted here, I will restrict myself to those of Derek Parfit and Mādhyamaka Buddhism. Using these accounts will allow us to challenge the plausibility of the account so far, and in doing so, create a final version which is not only better able to defuse the paradox that initiated our inquiry, but also more suitable for general usage when describing the sameness of changing objects.

A. Aristotle

According to Aristotle in *Metaphysics*, "a substance… is a principle and a cause…"[4] When one inquires about why something is what it is, we are really seeking to learn its cause, "and this cause is the *substance* of the thing."[5] When one examines a composite thing, it "exists in such a way as to be a totality, not like a heap…."[6] To take the example of flesh, it is "not only fire and earth or the hot and the cold but something else besides."[7] The "something else" which gives unity to flesh is the cause of its being flesh. For objects that are substances "formed according to nature or by nature, the *substance* of these would appear to be this nature, which is not an element but a principle."[8] Hence, over and above the material (elemental) make up of composites there are non-elemental principles that serve as causes of unity.

Traditionally, commentators have linked the "causes" mentioned the *Metaphysics* to the causes listed in *Physics*, which they have labeled the formal, the material, the efficient, and the final.[9] Clearly, Aristotle would be in agreement with the part of the analysis at the start of this paper in which it was shown that neither material nor formal definitions of identity are sufficient to track our commonsensical understanding of

4. Aristotle, *Metaphysics* Book Z.17, 1041a10.
5. *Ibid.* 1041b5.
6. *Ibid.* 1041b10.
7. *Ibid.* 1041b15.
8. *Ibid.* 1041b25.
9. Cohen.

sameness for certain kinds of objects. In the cases given above, a substance must take its unity from the conjunction of all four causes. Under Aristotle's scheme, the proposed criteria of continuously recognizable self-perpetuation might be classified as either an efficient cause or final cause depending on where the emphasis placed in self-perpetuation. If self-perpetuation is taken as meaning that the prior form gives rise to the later form, then self-perpetuation would be called an efficient cause. If self-perpetuation is taken as meaning that the arising form will be recognizable as a perpetuation of the prior substance, it might be called the final cause. (This is particularly relevant in the case of biological life, in which the organism may be anthropomorphically conceived of as "wanting" to survive, prosper, and reproduce, which is the "for what?" of the organism's existence.)

However, while continuously recognizable self-perpetuation might be seen as fitting either cause, it is a precise match with neither. It is not enough that to say that one thing acts as the efficient cause of the other for it to count as a part of the thing. For example, in addition to being a cause of its later self, a river is also a cause of erosion, but the silt in the bottom of a river is not a part of the river. At best, it is a part of the riverbed. Mere causal connection is not enough to count as proof of sameness or else everything that gravitationally influences anything else would be indistinguishable from the universe around it. We may metaphorically remark that a person has put something of herself in the house she builds, but we do not mean the house is the same as the person. Efficient causation is not identity.

Similarly, we cannot say that continuously recognizable self-perpetuation is the same as Aristotle's final cause. For Aristotle, the final cause of a biological organism is the prosperity of its soul, but since modern materialists will reject this, we do not wish to include such a requirement in our own account, as doing so would severely limit its usefulness for public discourse. Worse, the hypothesizing of a soul for a river or a ship would limit our account's acceptance to committed animists, whereas our intention is to

gain universal assent to our scheme in order to remove the applicability of the sorites paradoxes from a broad class of objects. Thus, we wish to give a naturalistic account of what it is that self-perpetuation entails without ruling out the possibility of non-naturalistic elements to reality. One might try to do this by simply describing the final causes seen in nature. However, as Roger Ames has quipped, most acorns end up becoming squirrels, not oaks! Hence, any naturalistic account of what it means for something to be the same cannot simply point to final causes in nature as if they were there to be seen on the surface of the thing. Final causes are imputed by interpretation rather than revealed by investigation. As such, there is no guaranty that various individuals will see the same final causes at work in a particular object.

This difference hints at a fundamental difference between the Aristotelian project and the proposal offered in this paper. Aristotle sought to describe the physical nature of reality as it is. The continuously recognizable self-perpetuation model of identity does not attempt to describe the physical nature of reality as it is, but rather seeks to come up with a coherent way of thinking about nature as it is described to us by science and ordinary experience. It aims to be a useful interpretation of the world that provides us with a usefully concrete heuristic for talking about our ordinary intuitions without in any way assuming that the world is obliged to exist at its fundamental level in a manner similar to what we describe.

Thus, both to the problem of distinguishing relevant casual connections from irrelevant ones and to the problem of describing self-perpetuation without presuming the existence of a substantive self-identity, what is proposed here is to defer to the existing recognitional capacities of social situated interpretive communities. In other words, society can already tell that us that though both the empty shell of a chrysalis and the butterfly have direct causal links to the pupa, it is the butterfly not the shell that inherits the mantle of "sameness" with the caterpillar. We do not need a theory to tell us that we think this already. The examples given during the exploratory phase of this

paper would have been meaningless if we did not already possess intuitions about the "correct" meaning of sameness in those cases.

Does this mean then that the theory of sameness presented here is already superfluous? If we have a social capacity to identify what we mean by the "same river" or the "same human being" already, does the use of "continuously recognizable self-perpuation" as a definition of sameness add to our understanding? Yes, it does, because definition given here is meant to sharpen the pre-existing ways of thinking about the world and in so doing aid us in 1) making concrete judgments in situations where we would otherwise be reluctant to make pronouncements (as seen in the dismantling of the ship argument earlier) and 2) removing absurdities that seem to result from trying to interpret the world naïvely as though material identity were the only sort of sameness, namely the paradox presented by Sainsbury.

B. Genidentity

The term "genidentity" was coined by Kurt Lewin in 1922, and since has received broader usage in the field of physics thanks in part to popularization by logical positivists like Rudolph Carnap[10] and logical empiricists like Hans Reichenbach. Reichenbach adopted the term in the posthumously published *The Direction of Time* after espousing similar views in earlier works, such as *Elements of Symbolic Logic* where he writes that Heraclitus

> is right if he intends to say that a river is not a thing in the sense of an enduring substance but an event sequence in which matter does not remain the same; he is wrong, however, if he wants to infer that it is not permissible to consider a river as a thing. The meaning of the phrase 'the same river' is a matter of definition; and with the definition of the word 'river' as denoting a thing that consists in an event sequence it *is* possible to step twice into the same river.[11]

In *The Direction of Time*, Reichenbach adopts the term "genidentity" to describe identity of this sort. As he explains a "thing is series of events succeeding one another in time; any two events of this series are genidentical. The concept of physical identity, of an

10. Sider, 194, footnote 33.
11. Reichenbach (1947), 273–274.

individual thing that remains the same throughout a stretch of time is based on the properties of this relation."[12] Furthermore,

> *[P]hysical identity* of a thing, also called *genidentity*, must be distinguished from logical identity. An event is logically identical with itself; but when we say that events are states of the same thing, we employ a relation of genidentity holding between these events. A physical thing is thus a series of events; any two events belonging to this series are called genidentical. The relation of genidentity is therefore a two-place propositional function which is symmetrical, transitive, and reflexive.[13]

Thus, for events identity is maintained forwards and backwards throughout time as a consequence of their having a single linear chain of existential causation between them. In addition, the relationship of genidentity is applicable to composite, organic entities, such as rivers and human beings. Hence, Reichenbach urges us to discard the dichotomy between things and events and see rather that the two "represent merely different modes of speech," such that the phrase "this tree is old" is translated with equal facility to the phrase, "The first events of the series constituting this tree are separated by a long stretch of time from the present event."[14]

Like the present account, Reichenbach acknowledges that due to the multiplicity of casual connections between events, it is not sufficient to speak of genidentity on the basis of casual connection alone. Accordingly, he defines two kinds of genidentity. "Material genidentity" has three necessary conditions: continuity of change, spatial exclusion, and discernibility of spatial interchange.[15] However, if we examine the world at the quantum level, we find that all of these properties are lacking of subatomic particles, hence "there is no material genidentity at all in the physical world,"[16] though we may continue to speak conventionally as though there were. The only true genidentity in quantum world is "functional genidentity." Unfortunately however, as Bas van Fraassen and Isabelle Peschard note in their paper "Identity Over Time," Reichenbach "does not offer an explicit definition" for functional genidentity.[17] Rather, Reichenbach

12. Reichenbach (1956), 224.
13. *Ibid*., 38.
14. *Ibid*., 224.
15. *Ibid*., 225.
16. *Ibid*., 236.

obliquely remarks that functional genidentity is "a genidentity in a wider sense," in which the latter two qualities of genidentity are jettisoned, "whereas the first is usually adhered to."[18]

Reichenbach is admirably frank about the fact that there is no "true" definition of genidentity, since we "can define genidentity to suit our purposes."[19] However, by defining genidentity as transitive and symmetric, he has made his concept less useful in the case of branching identities. Again, the Ohio and Missouri rivers are distinct from one another in spite of their not being distinct from the lower Mississippi. Similarly, the sameness of the acorn and the oak does not entail the sameness of the acorn and the desk. Thus, the concept of genidentity as defined by Reichenbach is not sufficient to map all of our commonsensical sameness relations.

In their exploration of the salience of identity and genidentity, van Fraassen and Peschard further argue that since there are quantum superpositions in which a pair of particles have no differentiating characteristics the Leibnizian "Principle of the Indiscernibility of Identicals" must be discarded. "[C]onceptual discernibility cannot, even logically, be a foundation for knowledge of numerical distinctness, but is a further stage of knowledge which has to involve a cognitive procedure, of individuation, which requires particular conditions of possibility."[20] Thus, while for Reichenbach functional genidentity is lost in certain interactions, van Fraassen and Peschard insist that there is a distinct concept of identity which is preserved even during entanglement. Hence Reichenbach's genidentity is not the same as the kind of identity over time which subatomic particles have. As for my own proposal, the principle of continuous recognizability is not sufficient to speak about the instantaneous self-sameness of particles in

17. Van Fraassen, 21.
18. Reichenbach (1956), 226.
19. *Ibid*.
20. Van Fraasen, 35. For my part, I wonder why two photons in super-position might not be regarded a single "double-photon." Perhaps they believe it would violate Ockham's razor. In any event, I will refrain from further comment, since I lack the mathematical background to justify such speculation.

super-position, let alone to speak about their sameness before and after entanglement, so it too must be abandoned as an adequate description of the quantum world.

The upshot of this is that neither the concept of genidentity nor the concept of continuously recognizable self-perpetuation are adequate to describe the world in certain quantum configurations. For the purposes of my proposal however, this is not a damning difficulty, since continuously recognizable self-perpetuation is not meant to describe all possible kinds of sameness relations. Rather, it is heuristic that lies over top of whatever ontology an individual may hypothesize for the ultimate or pen-ultimate constituents of reality, be it perdurantism, endurantism, or something wholly different. As for Reichenbach's three necessary marks of material genidentity (continuity of change, spatial exclusion, and discernibility of spatial interchange), while these can serve as a useful formalization of continuous recognizability and self-perpetuation in certain contexts, I believe that ultimately more useful criteria are those that evolve organically from an interpretive community rather than any completely fixed, formal description of what it is to be recognizable or self-perpetuating. Attempting to define these terms prematurely is sure to lead to unintended inclusions and exclusions. If the need arises for a sharpened definition of either term, then we can use Reichenbach's criteria as the starting point for further refinement together with the embodied instincts of the society of inquiry.

C. Parfit

One area in which there is a surfeit of pre-existing communal dialogue about the nature of sameness is that of personal identity. However, as Derek Parfit writes in his paper "Personal Identity," notwithstanding this dialogue,

> We can… describe cases in which, though we know the answer to every other question, we have no idea how to answer a question about personal identity. These cases are not covered by the criteria of personal identity that we actually use.[21]

21. Parfit, 3.

That is, an abundance of dialogue need not produce consensus, even when the facts are completely established. Part of the problem, according to Parfit, is that our intuitions about personal identity are held unusually strongly and in conjunction with another belief about the general decidability of such questions.

> No one thinks this about, say, nations or machines. Our criteria for the identity of these do not cover certain cases. No one thinks that in these cases the questions "Is it the same nation?" or "Is it the same machine?" must have answers.[22]

Thus, our expectation that "personal identity" is a solid relationship leads us to disappointment when we discover that it is no more robust than other conventionally stipulated, provisional definitions. The reason that this expectation is so deeply ingrained in us is that we all experience the world from the viewpoint of our own consciousnesses. Hence the "I" and the question of what I am become deeply familiar to us. This would not be so bad except that we go on from this instinctive attachment to the self to confound other important questions with a presumption that personal identity is always circumscribable. Fortunately, these questions, "can be freed of this presupposition. And when they are, the question about identity has no importance."[23] Thus Parfit's expository purpose is to provide examples that help us to shed our intuitions concerning the solidity of the self.

Considering the possibility of brain transplants and the temporary division and recombination of brain hemispheres, Parfit concludes that, "a person's mental history need not be like a canal, with only one channel. It could be like a river, with islands, and with separate streams."[24] On the basis of this, he argues that we should separate out the concepts of "personal survival" and "identity." One can survive the splitting of one's personality, even if one's numerical identity cannot be maintained. In fact, the reason that we place such great emphasis on personal identity is that maintaining one's

---

22. *Ibid*. Incidentally, one might quibble that people actually do insist (quite violently at times) that there is a single correct concept of national identity. See Greece's continued anger at the use of the name "Macedonia" by its northern neighbor for a blessedly mild illustration of this.
23. *Ibid*., 4.
24. *Ibid*., 7.

identity is indicative of survival, and it is survival that is the real crux of such inquiries. For Parfit, survival is a matter of degrees of psychological continuity rather than an absolute dichotomy. Parfit argues for this conclusion on two grounds. First, through a series of thought experiments about beings that split and fuse their identity at different times and ways, he shows that a gradated definition of personal identity better accords with our intuitions, and second, he argues that from an experiential point of view, seeing one's life from the point of view that identity is a matter of continuity and degree allows one a better outlook on life:

> Egoism, the fear not of near but of distant death, the regret that so much of one's only life should have gone by-these are not, I think, wholly natural or instinctive. They are all strengthened by the beliefs about personal identity which I have been attacking. If we give up these beliefs, they should be weakened.[25]

There are two distinctions between Parfit's account of identity and the account offered in this paper so far. First, for the most part Parfit emphasizes the importance of psychological continuity to the exclusion of bodily continuity. One exception is a footnote in which he explains that psychological continuity is a sufficient condition for survival rather than a necessary condition, since "in the absence of psychological continuity bodily identity might be sufficient."[26] My own inclination is to emphasize the interrelated nature of somatic and psychological survival, since failure to do so can lead to difficulties in characterizing cases of amnesia, etc.

The second and more significant divergence between the accounts is that as described so far continuously recognizable self-perpetuation has been treated as a bivalent truth function. Two things either share the relation of being the same, or they do not. Clearly, however, there are advantages to replacing the concept of "sameness" with "degrees of similarity" as Parfit has done. One would expect it to be easier in general to resolve disputes about whether things are somewhat similar than to resolve disputes about whether two things are different manifestations of the same thing. However, part

25. *Ibid*., 27. Note that the proposal presented here for continuously recognizable self-perpetuation makes no claim about its ability to instill virtue in those who employ it.
26. *Ibid*., page 13, footnote 17.

of the reason that these questions are more easily resolved is that much less hangs on them. Saying, for example, that a person is the same person even after entering a persistent vegetative state is controversial precisely because it brings along with it questions about whether the vegetative individual should be considered to have the same sorts of rights as before their brain was damaged. To just say that one is radically psychologically dissimilar after such brain damage is less controversial precisely because it is less useful of a judgment. While there is not sufficient space to explore the issue here, needless to say, for legal and moral purposes, it is not sufficient to leave the matter behind after resolving that the individual is dissimilar to their past self by whatever percentage. If the next of kin wish to pursue euthanasia, a judgment must be rendered about whether the body that remains counts as the same person or not. Thus, to give an account about "degrees of psychological continuity" is just to restate the issues that are not in dispute, rather than to address other issues that seen to bear on the matter of dispute. Hence, the chief objection that can be leveled at Parfit's account is precisely that there is nothing objectionable in it. What is objected to is what Parfit's account leaves out: a definitive judgment about the matter of survival and identity.

D. Mādhyamaka Buddhism

Parfit's account of personal identity as survival has (like the account given here) many precedents, among them Hume's "bundle of sensations" and the Buddhist notion of self. Since there are many different kinds of Buddhism with different doctrinal emphases, this paper will restrict itself to Mādhyamaka. The conventional level outlook of the Buddhist is that human consciousness is a stream of disconnected mental events arising due to previous mental events. As Paul Williams summarizes, "We are each of us an ever-changing composite of various radically impermanent psycho-physical components extended over space and time."[27] With this background in place, the

27. Williams, 106.

Buddhist concludes that the self is an illusion, as for example in *Bodhicaryāvatāra* 8:101, where Śantideva says,

> The continuum of consciousness, like a series, and the aggregation of constituents, like an army and such, are unreal. Since one who experiences suffering does not exist, to who will that suffering belong?[28]

The basic groundwork behind this argument seems to be remarkably similar to one made by Kant in a footnote to the first edition of the *Critique of Pure Reason*. He writes,

> An elastic ball striking another such ball in a straight direction communicates to that ball (if we take account merely of the positions in space) its entire motion and hence its entire state. Now let us—by analogy with such bodies—assume substances one which imbues the other with presentations along with the consciousness of these. We shall then be able to think an entire series of such substances: the first would communicate its state, along with the consciousness thereof, to the second substance; the second would communicate its own state, along with the state of the previous substance, to the third; and the third substance would similarly communicate to yet another the states of all previous substances, along with its own state and the consciousness of all of them. Hence the last substance would be conscious of all the states of the substances that had changed before it as being its own states, because these states would have been transfered to it together with the consciousness of all of them. **Despite this, however, that substance would not have been the same person in all these states.** [*Emphasis mine*][29]

Thus, for both Kant and Śantideva, merely causal continuation of a consciousness from moment to moment is not enough to speak of a true existence. Interestingly, for Kant, this shows that the self is more real than the person, whereas the Mādhyamika draw the opposite conclusion. For Kant, this serves merely as a theoretical argument against certainty that the self is one with the subject. For Buddhists, however, this serves to show that there cannot be such a thing as a self at all. Both conclusions seem to rest on the supposition that identity of the sort I have been calling "continuously recognizable self-perpetuation" is insufficient to uphold a robust concept of existence. Furthermore, like Parfit, the Buddhists claim that seeing things as mere imputations and aggregates is useful for building virtue and taming the passions. For the Mādhyamika, my attempt to piece back together a meaningful sense of identity for an aggregate being is not just entertaining a pragmatic fiction—it is a pernicious reinforcement of those mindsets which have hampered our enlightenment for uncountable ages! This is not to say that

28. Śantideva, 102.
29. Kant, A364 footnote.

Buddhists must make themselves ignorant of our commonsensical judgments about composite entities. As Williams writes,

> [T]ables, chairs, and mountains seen by cognitions which in everyday life are held to be valid (there is no disfunction in the means of cognition), are correct conventionalities, but still 'fictions.' … [A] great many things (probably all) which we would normally consider to be genuine realities, the 'furniture of our world,' are going to be fictions for Śantideva since they are wholes, composites made out of parts. This includes, of course, the cosmos—the 'totality of things'—itself.[30]

Thus, the basic position of the Mādhyamika is to assent analyses like that of Sainsbury which opened this paper: if no constituent part of a human being is sufficient to give or take away totality from the whole of the human being, then the conclusion is that the totality itself is an illusion. As in *Bodhicaryāvatāra* 9:56–59, the Mādhyamika regularly draw conclusions about the possibility of enlightenment from the fact we find nothing that can withstand repeated analysis,

> If there were something called "I," fear could come from anywhere. If there is not "I," whose fear will there be? Teeth, hair, and nails are not I, nor am I bone, blood, mucus, phlegm, pus, or lymph. Bodily oil is not I, nor are sweat, fat, or entrails. The cavity of the entrails is not I, nor is excrement or urine. Flesh is not I, nor are sinews, heat, or wind. Bodily apertures are not I, nor in any way, are the six consciousnesses.[31]

The Mādhyamika do not restrict these sorts of analyses to human beings, but by means of repeated application of the *reductio*-like prasaṅgika argument attempt to show that all phenomena are *mṛṣā*, fiction, since there is nothing substantial in them that can be identified with a self-existing phenomenon. This is not to say that conventionally perceived objects are utterly non-existent. The Mādhyamika are not nihilists. Rather, as their name suggests, they seek a "middle way" between substantialism and nihilism. As Williams explains a conventional object such as a table,

> will be a fiction because it will not exist the way it appears (it will appear as if it is existing from its own side, as independently self-subsistent, 'inherently' existent, while actually it exists as a conceptual imputation superimposed upon its 'bases of imputation'), but that fiction will nevertheless exist. It will enter perfectly adequately into pragmatic transactional usage and therefore will not be the same as a completely non-existent thing.[32]

---

30. Williams, 108.
31. Śantideva, 122–123.
32. *Ibid*. Note that by page 112 Williams has gone on to argue that—in spite of what Śantideva has claimed, the position of the *Bodhicaryāvatāra* entails the non-existence of objects which are *mṛṣā*. Nevertheless, here we will take Mādhyamaka's self-description as definitive rather than possible logical implications it may have.

It is clear then that the Mādhyamika will permit the use a theory of description such as this paper's continuously recognizable self-perpetuation only so long as the individual using this method of identification is intimately aware that there is nothing independently existing which supports its use. Rather, the use of such a method is possible because of the codependent arising of illusions born of ignorance from the fathomless past. From the perspective of this paper, such a caveat is acceptable if not obligatory.

E. Result of comparison

So far, the proposal presented in this paper has shown itself able to coexist with philosophical systems as diverse as Aristotle's theory of substance, Reichenbach's logical empiricism, Parfit's analytic Anglo philosophy, and Mādhyamaka's non-dualism. If time and space permitted, it might be possible to compare the current proposal to any number of other schools of thought from process philosophy to Daoism. The literature on the Heraclitean problem is quite deep. While the results of such comparisons cannot be known in advance, the evidence suggests the theory here would show itself to be compatible with many such systems. The reason for this wide-ranging compatibility is precisely because it attempts only to describe the surface level of reality, and even that only provisionally and subject to refinement. While it might seem that attempting only to make a surface level description of sameness should result in a "shallow" theory, quite to the contrary, by proposing a relatively modest theory, we are better able to conform our theory to our intuitions where we are sure of those intuitions and to provide guidance where our intuitions leave us without clear guidance. The principle at work here is analogous in the field of ontology to the semi-compatibilist view of determinism that John Fischer endorses in a review article for *Ethics*,

> I want to end by sketching what I take to be a very powerful motivation for embracing "semicompatiblism"—the doctrine that causal determinism is consistent with moral responsibility, even if causal determinism rules out alternative possibilities. I believe that we-you and I and most adult human beings-are morally responsible (at least much of the time) for our behavior. Further, I do not think that this very important and basic belief should be "held hostage" to esoteric scientific doctrines. For example, if I were to wake up tomorrow and read in the *Los Angeles Times* that scientists have decisively proved that

> causal determinism is true, I would not have any inclination to stop thinking of myself, my family and friends, and human beings in general as morally responsible. The precise form of the equations that describe the universe, and whether or not they are or correspond to universal generalizations, are not the sorts of thing that should be relevant to our most basic views of ourselves (as morally responsible agents and thus apt targets of the reactive attitudes).[33]

In the same way that our moral intuitions should not be subject to the vagaries of scientific research, so too our basic intuitions about the sameness over time of composite entities such as human beings, ships, and rivers should not be subject to complete rejection when a new ontology presents itself, though they may, of course, be shown to be merely provisional designations and not reflective of ultimate reality. Any new ontology must show us how to refine the basis of understanding that we already have, but it is impossible for it to completely overturn our existing intuitions, because these intuitions have already shown themselves to be so pragmatically useful for ordinary life.

By comparing to Aristotle, we see that efficient causes are too numerous to reflect our intuitions about identity and final causes are too interpretative to consider as a point of agreement for all observers. From Reichenbach and genidentity, we gain a valuable starting point for further formalization of the concept presented here, and we see the limitations of both genidentity and this concept when dealing with the quantum realm. From Parfit, we see benefits and disadvantages of using degrees of similarity rather than bivalent predicate. From the Mādhyamika, we are warned of the potential dangers of reifying the concept of sameness given here into a self-sufficient mode of being.

IV. Application to paradoxes

Returning to the paradox that opened the paper, we said that, "A human being still exists even if he or she loses one molecule from his or her body." However, in light of our newly refined definition, we cannot assent to that assumption without caveats. If

33. Fischer, 129.

the human being is thought of as a solid, unitary object, then clearly it disappears as soon as a single molecule changes. However, if we think of a human being as a recognizable continuum of self-perpetuation, then it may sometimes be the case that a person persists even after the loss of a molecule, but if this is the case, it is only the case in virtue of the fact that this molecular loss has not disrupted the larger self-sustaining processes within the body. These processes could indeed be disrupted by a too rapid loss of molecules. Thus, the premise must be changed to, "A human being still exists even if they lose one molecule from their body, *so long as the loss of this molecule does not interrupt the biological self-perpetuation of the person as an organism*," since certainly we can imagine cases where the loss of each additional molecule adds ever so slightly to the probability of near term death. With that clause added to the assertion, the paradox collapses, since it no longer allows itself indefinitely repeated application but only strictly limited application, such that the rate of molecular loss is never significantly greater than the rate of molecular replacement. As our intuitions suggest, a person can withstand the loss of any one atom, but not the loss of all or even too many of them.

Similarly, we can answer the question, "Does [the ship built from discarded parts] have a better claim to be the original ship of Theseus?" with a resounding "no," given the stipulation that when the ambiguous term "original ship" is used it refers to sameness according to the criterion of continuous recognizability as described above rather than sameness according to a naïve materialistic criterion. Hence this paradox turns out to be a matter of the ambiguity of the term "same" rather than a matter of vagueness within the term.

With their premises denied, these paradoxes can be safely sidestepped, and we can avoid the recourse to completely radical nominalism and continue to use our commonsensical vocabulary with confidence that if necessary there is recourse to more precise formulations that describe the same basic understanding of the world, even with these vocabulary items do not reflect reality at its deepest level.

Mādhyamaka Ontology and the Problem of Ultimate Analysis

I. Introduction

The Diamond Splinters argument, in which the independent existence of any particular thing is shown to be illusory by means of a *reductio ad absurdum*, is familiar to any scholar or religious inquirer acquainted with Mādhyamaka thought, but what it actually means is less than clear. Can arguments of this sort be applied indefinitely? Even to themselves? Does the fact of their universal applicability imply that nothing exists at all? Or is it just a kind of verbal trick the Mādhyamika perform to help cultivate detachment? Could it be that these arguments only work on one level of reality but not another? Is it necessary for the point that Mādhyamika are trying to make that these arguments always proceed by *reductio*?

In this paper, I will explore different interpretations of Mādhyamaka philosophy in the context of exploring just what sorts of problems any ultimate analysis will encounter. For this exploration, I will first borrow a framework of terminology from Nicholas Rescher concerning principles that any ultimate theory will need to deal with then explicate the logical possibilities they present. With these possibilities in mind, I will briefly explain Kant's ontology and method in order to show how they relate to Murti's analogy of Mādhyamaka philosophy to a Kantian system. Against this, I will present Priest's investigations of the limits of inquiry and objections concerning the difficulties of dualistic systems, as well as his interpretation of Nāgārjuna, before offering a unified interpretation of the Mādhyamaka argument and its consequences, especially with regard to the nature of illusion.

II. Fundamental principles and possibilities

A. Principles

In a chapter entitled "The Price of an Ultimate Theory" in his *Nature and Understanding*, Nicholas Rescher lists four principles, admittedly not originals, which are commonly

implicit in the scientific pursuit of "a Theory of Everything" or TOE. The first principle, which has been formalized since at least Leibniz, is the Principle of Sufficient Reason:

$\forall t\ \exists t'\ t' \Sigma t$

Put into plain English, "for any theory or fact, there exists a higher theory or fact whose explanatory power subsumes the original." We can explain the fact that apples can be seen falling toward the earth at various times by means of a naïve theory of geocentric gravity. With a more sophisticated theory of gravity, we can additionally describe the motions of the sun and the moon as a consequence of the same imputed fact. It is the hope of theoretical physicists today that a further theory will also be able describe gravitational, electrical, and nuclear forces as aspects of the same force. The alternative to the Principle of Sufficient Reason is the possibility that there may be events that are not only without apparent cause but without a hidden cause as well. Naturally, our intuition is distrustful of such a possibility. Some (including Priest) might object that some interpretations of quantum theory violate the Principle of Sufficient Reason, which shows that it is not a necessary truth. However, even in quantum theory, things happen for a reason, it is just that the reason affects things probabilistically rather than deterministically when extrapolating forward from presently know conditions. Thus, everything that happens has a reason for which it happens, but those prior reasons are not sufficient to completely explain why the event happened. Whatever the case, Bell's theorem only precludes *local* hidden variables not non-local ones, so determinism can still be potentially be saved by abandoning locality as is done in, e.g., the Bohm interpretation of quantum mechanics. Of course, the exact parameters of this debate are outside the scope addressed here, so we will leave the issue at that.

Applying the Principle of Sufficient Reason, we naturally suppose that by following the chain of reasoning from an initial fact or theory t to an explaining fact or theory t' then to an explanation of the explanation t'' and on to the limit of the series, we

will at last encounter some ultimate theory or fact, T^*, which explains all prior theories and facts. This intuition is formalized as the Comprehensiveness of T^*:

$\forall t\ T^* \Sigma t$

"For all theories or facts, T^* explains it." So far so good, but it should be apparent to the reader that a conflict is in the offing. If according to the Principle of Sufficient Reason all theories can be further explained, what further explains T^*? To be consistent, we must either say that T^* is only an implicit limit and not actual since the regress is infinite, or T^* is actual, and T^* explains itself. We hereby introduce the principle of Finality:

$\sim\exists t\ (t \Sigma T^*\ \&\ t \neq T^*)$ or equivalently, $\forall t\ (t \Sigma T^* \rightarrow t = T^*)$

"Nothing else can explain T^*. If a fact or theory can explain T^*, then that it is T^* itself." Opposing this is the principle of Non-circularity:

$\sim\exists t\ t \Sigma t$

"No fact or theory can explain itself."

B. Possibilities

At this point, our intuitions about the possibility of an ultimate theory have reached an impasse. On the one hand, the Principle of Sufficient Reason seems to imply an additional principle of Non-circularity, so that no theory is allowed to justify itself, and thus bring the search for more powerful explanations to a halt prematurely. On the other hand, it seems impossible that there are no bedrock necessary facts which ground all other facts. Thus, the logical possibilities can be categorized like so:

1. There exists a finite number of justifications which satisfy Finality.

1a. The ultimate justifications are self-justifying at the cost of circularity.

1b. The ultimate justifications have no further justification and violate of the Principle of Sufficient Reason.

2. There is a regress of justifications which satisfy Finality collectively.

2a. The regress is self-justifying at the cost of circularity.

2b. The regress itself has no further justification and violates of the Principle of Sufficient Reason.

3. There are no ultimate justifications (nothing satisfies Finality).

4. Things are contradictory.

Worth noting here is that so far, we have allowed epistemology and ontology to be considered interchangeable. It is possible that epistemology conforms to one possibility and ontology another. For example, it might be the case that our empirical explanatory process can continue indefinitely, but there are only a small number of bare facts which secretly undergird reality itself. Or perhaps, what we can know is limited and must halt upon reaching certain theories, but those theories are secretly supported by an infinite regress of further facts that are beyond our ability to know.

Historically, many philosophers have been less than clear in distinguishing between these possibilities. Theists generally consider God to be the ultimate reason for everything (or at least, God plus free will), but they have not specified whether God constitutes the reason for Himself, or whether God is without further justification (possibility 1a or 1b). Recently, some string theorists and other scientists seem to think a perfected explanation of the Big Bang itself may constitute a brute fact with no possible further explanations, but others propose a "bubble multiverse" in which our universe is the product of an infinite line of prior universes that split off from one another in Big Bang-like events. The determination of which of these possibilities Mādhyamaka analysis entails is the concern of the remainder of this paper, but first a short digression into Kant's exploration of these possibilties is necessary.

III. Kant

The investigation into the limits of rationality by Immanuel Kant in *The Critique of Pure Reason* is relevant to the explanation of Mādhyamaka philosophy for two reasons. First, Kant's dualistic system of phenomenal reality and noumenal reality bears a striking resemblance to the Mādhyamaka system of conventional truth and ultimate truth. Second, the contradictions by *reductio ad absurdum* that Kant gives in the Antinomies section of the *Critique* bear a resemblance to the prasaṅgika method of *reductio ad absurdum* employed by some Mādhyamika. To begin, we will give an explanation of Kant's ontological system.

A. Ontology

Kant calls his view of fundamental ontology "transcendental idealism" and contrasts it with all preceding views, which he calls "transcendental realism." Explaining the difference on A369–70 he writes,

> By *transcendental idealism* of all appearances I mean the doctrinal system whereby we regard [appearances], one and all, as mere presentations and not as things in themselves… the transcendental realist conceives outer appearance (if their actuality is granted) as things in themselves that exist independently of us and of our sensibility, and that would therefore be *outside* us even according to pure concepts of understanding. It is, in fact, this transcendental realist who afterwards plays the empirical idealist. Having wrongly presupposed that if objects of the senses are to be external then they must have their existence in themselves, ie. even apart from the senses, he then finds from this point of view all of our presentations of the senses are insufficient to make the actuality of these objects certain.
>
> The transcendental idealist, on the other hand, can be an empirical realist or, as he is called, a *dualist*… Hence matter is for him only a kind of presentation (intuition), called external; they are called external not as referring to objects *in themselves external*, but because they refer perceptions to the space… although the space itself is in us.

In other words, for Kant the tragedy of all prior philosophy had been that philosophers supposed that if there were objectively real things, such things would have to exist independently of the mind, however since by definition the mind cannot know anything that is independent from the mind, it cannot know these real objects. Kant reverses this pattern of thinking. For Kant, to be an objectively real thing is to depend on the mind and its a priori categories. As such, through the application of universal

reason, we can give an objectively valid description of the unity which reason is compelled to impute in appearances. Unfortunately, even an objectively real synthesis of appearances cannot be the object in itself, which he also called a noumenon. The synthesis is only an appearance or phenomenon, though no less objective and real for it. Kant is insistent that we can only "think" the noumena which give rise to phenomena, never cognize or know them. (This, as we shall see, presents problems for Kant later.)

B. Method

Kant's argument for his bifurcated system comes in two main parts: the Transcendental Aesthetic and the Transcendental Logic. In the Transcendental Aesthetic, he argues (somewhat unconvincingly, at least in the view of this author) that the possibility of our experience requires an a priori intuition (in Kant's sense of intuition as a kind of non-conceptual sense data) of space and time. The Transcendental Logic breaks down into two further systems, the Transcendental Analytic and the Transcendental Dialectic. The Transcendental Analytic continues the exploration of the a priori concepts behind the a priori intuitions of the first section. The Transcendental Dialectic is more interesting. In it, he explores problems inherent with supposing that one can have a coherent system that does not rely on a phenomena/noumena distinction. In particular, in the Antinomies, he gives *reductio* arguments both for and against the infinity of space, the beginning of time, the infinite divisibility of substance, causation through freedom, and a necessary being. To resolve these contradicting arguments, Kant proposes that the objects under consideration can only be considered to exist as noumena, and hence outside of the a priori categories of our judgments.

In terms of the possibilities listed in the previous section, Kant differentiates between what he calls "mathematical series" and "dynamical series." (The difference need not concern us here. The distinction mostly seems to exist to serve Kant's preexisting religious commitments.) We must consider time infinitely long, space infinitely

large, and substance infinitely divisible, because they are mathematical series, but we should consider causality through freedom and a necessary being to exist as unified absolutes rather than series, since they are dynamical. In all cases, however, if we try to move beyond the possibility 1 or 2 to determine whether the respective a or b subdivision of the possibility applies (that is, whether the absolute or series is self-justifying or without justification) then we move well beyond phenomenal experience and into the realm of the noumenal, in which such answers are not to be found. And this motion into noumenal is what allows the construction of the paradoxical assertions of the Antinomies. The respective absolutes and infinities of each series are mere epistemic imputation based on the a priori requirements of reason but are nevertheless objectively real, since all reasoning individuals will reach the same result through investigation. Going beyond them, since the noumena are not available as objects of inquiry, we cannot say whether the possibility justifies itself or is without justification. In any event, since the answer is beyond all possibility of experience, for Kant the question is unanswerable by our speculative reason.

So we see that Kant hopes to rescue the final basis for being from contradiction by placing it outside the realm of judgment in a noumenal realm to which the categories of judgement do not apply. Whether this attempt is either successful or relevant to an explanation of Mādhyamaka will next be elucidated further.

IV. Murti

In 1955, T. R. V. Murti helped bring Mādhyamaka to the attention of Western philosophy with his seminal *The Central Philosophy of Buddhism*. Unfortunately, as detailed as this work is, it suffers from the author's misinterpretation of some subtle points of Mādhyamaka ontology. Murti states on pages 332–333:

> His [the Mādhyamika] position is akin to that of Kant. … The difference between the two… is that Kant seeks to realize these noumenal realities in a non-intellectual mode—Faith and practical Reason; the Mādhyamika does it in Intellectual Intuition—Prajñāpāramitā. The Mādhyamika is spiritual to the core. His absolute is not void, but *devoid* of finitude and imperfection. It is nothing but Spirit.

In a footnote on 333, he quotes Professor Radhakrishnan approvingly on the topic of śūnyatā, "To call it being is wrong, only concrete things are. To call it non-being is equally wrong. It is best to avoid all descriptions of it."

A. Mādhyamaka ontology compared with Kant

A Kantian interpretation of the Diamond Splinters argument might be that we do not find the independently existing thing in experience, because we are searching for it as a thing in itself (noumenon). Our experience of a phenomenal object cannot possess the requisite totality of properties that one would expect of a true thing in itself. Instead, what we do find in experience is the imputation of an objective totality that arises from the partial experiences of it. For Kant, this phenomenal imputation is *not* a subjective illusion, for all beings possessing reason would create the same imputed object (presentation) given the same non-conceptual sense data (what Kant calls intuition). Furthermore, it is impossible to suppose that we *could* be given the totality of experience regarding a phenomenal object, because our experiences are necessarily rooted in space and time, whereas for things in themselves, space and time are merely ideal.

Next, we need to be careful in order to avoid an internal debate within Mādhyamaka. Depending on whom one asks, conventional truth may or may not also be empty of essence. In order to avoid this dispute, the interpretation given here is based largely on works written before the internal debate in Mādhyamaka arose and particularly on the *Bodhicaryāvatāra* by Śantideva. *Bodhicaryāvatāra* 9:5 claims, "Ordinary people see and imagine things as real and not illusory. It is in this respect that there is disagreement between the contemplatives and the ordinary people." This suggests that Śantideva would not accept the doctrine of empirical realism without first making a distinction about what is being accepted as empirically objective. Notice also that Kant's method in the Transcendental Aesthetic is to ask, "What are the conditions for the very possibility of our experiencing *X*?" Such a question consciously presupposes that the *X* in question

is a valid object of inquiry in some sense. The Mādhyamaka project starts with the opposite assumption. Rather than seeing how it is possible that we have come to have mostly correct views, Mādhyamaka is interested in explaining the false views that keep us bound in suffering. For Kant, universal assent to the existence of objects is a mark of their objective reality. For Mādhyamaka, it is a mark of our collective beginningless ignorance. The difference between the two is their fundamental outlook on the life in this world. For Kant, our reason naturally compels us to accept various arguments, like the cosmological argument for the existence of God, which while not strictly correct, are nevertheless useful for the ordering of our lives according to practical reason. For Mādhyamika, our reason naturally compels us to an eternity of dissatisfaction, because we mistake illusions for real things. Thus, to be an empirical realist would endanger the entire Bodhisattva path of liberation as a uniquely meaningful pursuit. As Śantideva says in verse 9:6 of the *Bodhicaryāvatāra*, "Even objects of direct perception, such as form and the like, are established by consensus and not by verifying cognition. That consensus is false…" Ordinary people using reason in a natural manner see objects of experience as real, external things that can be used to assuage craving and bring happiness, whereas the entire point of Buddhism is that such things can never bring liberation from suffering. Thus whether objects are conventionally without essence or not, Śantideva would still be loath to admit that objectively speaking things are substantially similar the way the appear to us, since to appearance, things satisfy cravings, but this appearance is the cause of samsara.

In this interpretation of Mādhyamaka thought, conventional reality is seen as dangerously deceptive. One might attempt to counter it as an accurate portrayal of Mādhyamaka understanding by presenting Candrakīrti's seemingly conciliatory remarks in *Madhyamakavatara* 6:166-7. Candrakīrti's remarks,

> Things such as jugs, cloth, tents, armies, forests, rosaries, trees, houses, trolleys and guest-houses should be understood to exist in the way they are commonly spoken of by people because the Buddha did not argue with the world over these matters.

> Furthermore by applying the analysis of the cart to part-possessors and their parts, quality-possessors and their qualities, people with attachment and their desires, bases of characteristics and their characteristics and fire and the fuel it burns, one finds that they do not exist in any of the seven ways. But as long as they are not subjected to such analysis, they do exist in another way: namely in terms of their being well known to the world.

At first, it may seem that Candrakīrti is affirming the existence of conventional objects. However, when one carefully studies the sense in which conventional things exist, one sees that it is only in terms of being well known to the world. For Kant, part of what makes things phenomenally real is that they are well known to the world, but phenomenal objects are no less actual things for this, and so long as one does not look for the properties of the thing in itself, one will find the object to be quite real. For Candrakīrti things are well known to the world, but the sage must come to find that they actually possess none of the seven marks of existence, though people may generally assume that they do. Indeed, even the world which "knows" these things to possess existence fails to exist in the truest sense for the Mādhyamika. Kant strongly opposed solipsism, hence such seeming "nihilism" would have seemed to him anathematic to his definition of empirical realist. Thus a Mādhyamika cannot be considered an empirical realist in the Kantian sense.

B. Mādhyamaka method compared with Kant

Murti's real point in bringing up Kant seems to be designed more to emphasize the seeming commonality of their dialectical approach to reasoning. On page 297, he dismisses most of the *Critique* as uselessly in thrall to dogma:

> The Transcendental illusion is the real starting point of his *Critique*. The consciousness of this illusion is engendered by the conflict in Reason as exemplified by totally opposed philosophies. Kant's preoccupation with an explanation of experience serves to confound his readers and cloud the issues. And it is at variance with the anti-speculative tendency of the *Critique*.

For Murti, the importance of the *Critique* is that it is a critique at all. The earlier quote about the similarities of Kant and Mādhyamika not withstanding, the main point that the two share is the tendency to challenge us to find the very limits of reason itself and by doing so, invite us to look beyond the phenomenal realm. Murti sees the exposure of

transcendental illusion as the common feature shared by the Mādhyamika and Kant, and thus feels justified in glossing over differences between transcendental idealism and the two truths system in order to make the commonality of method more clear. Returning to the possibilities outlined earlier, Murti would say that for Mādhyamika, like Kant, we can expose an infinite chain of illusory objects in conventional reality through a dialectical application of the Diamond Splinters argument. In order to go beyond this chain into the noumenal reality or ultimate truth, we cannot employ any ordinary sensory activity, but must use Intellectual Intuition which allows us to have direct access to the absolute totality of Spirit that lies beyond the chain and ultimately justifies it. For Murti, this process must be dialectical, so that we never fall into the error of espousing some view, because doing so would collapse us into either circularity or insufficient reason and hence invalidate the chain of inferences leading to it. Instead, we can only be lead by the dialectical process into directly experiencing the way that this absolute totality of Spirit is outside of our usual categories of classification and hence justifies the chain while remaining immune to the collapse.

For now, we will leave aside the errors Murti has made by interpreting Mādhyamaka in this way and grant his point that the they share with Kant a goal of finding limits to reason. Nevertheless, as we shall see Priest point out, there are serious problems lurking in wait for any attempt at creating a Kantian definition of the limits of reason in this manner.

V. Priest

A. Kant and the problem of expressibility

Graham Priest's *Beyond the Limits of Thought* is an intriguing compendium of various historical situations in which philosophy has run up against the limits of reason, and also, in Priest's view, the existence of true contradictions. Returning to Kant, we explained that he held that the noumenal is beyond phenomenal experience. Indeed, the

noumenal cannot even be spoken of according to the usual a priori categories such as existence or causality. He explains Kant's view on page 81:

> The reason that we cannot have knowledge of noumena is precisely that we cannot even make statements about them: any (meaningful) statement about them would have to apply the categories, and so is impossible.

However, Priest feels that Kant is here contradicting himself. The fact is that Kant *does* apply the categories in speaking of noumena, repeatedly, and any attempt to purge the instances of such from Kant as minor aberrations from his usual method is doomed to fail, because the entire premise of noumena is contradictory. From pages 81–83:

> When Kant says noumena may be supposed to exist (A253=B309) he deploys the Category of existence; when he says they are not in time, he deploys the Category of negation. Even the statement that Categories cannot be applied to noumena deploys the Categories of possibility and negation! … Let me emphasize again: this is not a contradiction of the kind of which one finds so many in the *Critique*: a result of carelessness or of changes of view; it is a contradiction which is occasioned by the very objects of the theory.

Priest's insights here capture the central project of the *Critique*: to offload all of the contradictions entailed in taking the usual objects of experience as things in themselves by moving noumena into their own realm, where, after stripping them of time and space and the categories of judgments, Kant hopes they will become powerless to cause any further problems for Reason. Unfortunately, the project still collides with the dreaded contradiction when one sees that even if noumena could be partitioned off from the world of experience, it nevertheless entails a contradiction, because if it is possible for the noumenal world to influence the phenomenal, then the noumenal is exhibiting causality of a phenomenal nature and thus is pulled down to the same level and subject to the same limitations as phenomenon.

B. Murti and the problems of expressibility and self-reference

Murti is also vulnerable on this point. He does however anticipate the problem, and in the section, "Is śūnyatā a theory?" he protests that it is not, because, "Criticism of theories is no theory" (161). He goes on to explain on 162,

> To analyze a proposition is not to make another proposition. If that were the case, we cannot make any universal statement. For, the statement about the nature of *all* propositions will, on this contention, be itself a proposition… Likewise, the self-conscious awareness of *all* points of view, or Reason as such, cannot itself be a view. Hence the true

> universal cannot be a view-point (a dṛṣṭi); and conversely all points of view are particular, not universal.

One might object that surely criticism is a theory—the theory that the criticized theory is incorrect. Against this, Murti claims that universal propositions are not views, hence neither is the universal criticism of śūnyatā. However, ordinarily, we will call any proposition that can be referred to and predicated meaningfully a view, and surely, it is at least sometimes possible to make reference to universal judgments. For example, we can say without contradiction, "the universal judgment that everything—be it a thing, a word, a concept, this sentence, or anything else—is actually a banana is a false universal judgment." Being universal was no protection for that judgment from being quite demonstrably false. (Even a radical monist would have to admit the properties usually imputed in a banana—being real, yellow, edible, a fruit, and differing from non-material things—are not present in a proposition, though the monist may suppose that the bananas which we actually experience are secretly the same as everything else.) Therefore, if false universal judgments are not made inexpressible by their self-reference, then it seems likely that a true universal judgment is also theoretically possible. But if true universal judgments are possible then they must also constitute a view, since we can say determinately, "this universal judgment is true." Murti might here offer the following line of objection, "the previously given example universal judgment is clearly false, and thus it counts as a view. However, there is another class of universal judgments which, due to epistemic limits, cannot be determined as true or false. These universal judgments are not views." Alas, his last statement sinks him. If Murti's objection is true, then the statement "these universal judgments are not views" is itself epistemically available to us. In other words, if he has a meaningful and valid argument, we should be able to affirm it. But, if his argument is valid, then it is possible to refer to those universal-judgments-which-are-not-views in a meaningful way, and if this is possible, then those universal judgments are views after all, since they still can be

used to predicated meaningfully, even if we lack the ability to predicate as regards the truth or falsehood of their content. "1.) It is true that śūnyatā is not a view. 2.) If an understanding is not a view, we cannot make meaningful assertions about it. 3.) Contradiction from assumptions one and two." Thus, *reductio* has turned against Murti.

If we examine Murti's motivations, we can see that he wants to say, "All views are wrong, and this is not a view." He wants to avoid saying, "All views are wrong except this one," because to allow an exception like this is too ad hoc. Any opponent would of course affirm just the same thing. The only difference is that the meaning of "this view" would refer instead to the opponent's own view. By adding the rider, "and this is not a view," what Murti is really attempting to show is that his view is a *special* view, not subject to the usual limitations, and only his view is able to be treated in this way. The addition makes his view uniquely non-substitutable, since while on final consideration we may not accept that śūnyatā is not a view, the claim is not a completely implausible, as it would be if one were to claim that, say, "string theory is not a view." The same motivation comes into play whenever we use the Principle of Sufficient Reason. There is no reason we could not propose that the Principle of Sufficient Reason is incorrect, and the correct principle is that everything has a higher explanation, except for whatever thing our particular theory considers to be ultimate (be it God or a scientific theory of everything or whatever). However, to do so is blatantly ad hoc and begging the question, because we first determined that the proposed ultimate thing must exist at all on the basis of the Principle of Sufficient Reason, but now we have changed that foundational principle in order to ensure it stops at our ultimate thing. Thus, Murti's attempt to escape this sort of difficulty by finding a special kind of view that is uniquely suited to non-refutation is not misguided, just failed.

It may be objected that Murti is wise to posit that a universal statement cannot refer to itself because allowing indiscriminate self-reference engenders the Curry

paradox and the like. However, the attempt to limit the scope of self-reference to specific areas, as in a type system, is not a viable metaphysical proposition, though it may be a useful technique mathematically. The reason is that any sufficiently complex type system is inherently limited by Gödel's paradox to being either contradictory or incomplete. If Murti were to claim, "it is true that there is no self-reference at the ultimate level," then this is a contradiction for the ultimate level describes itself as being without self-reference. If he were to claim that this can be considered to be the case but it is not explicitly stated at the ultimate level so as to avoid contradiction, then the ultimate level is incomplete. (An incomplete system is one that cannot fully describe itself.) While this may be fine for mathematics, in which the higher system (like the language of the textbook) describes the lower system (the metamathematical system described in the textbook) which in describes still lower systems (ordinary arithmetic), we cannot suppose reality to be incomplete at the highest ontological level without conceding in advance that there are no ultimate explanations, only *penultimate* explanations. In a type system, you can say there is no meaning in asking for the value of a self-referential formulation within a given type, but one can always construct a higher type in order to refer to the lower formula. If there is no self-reference on the ultimate level, then there is always another level beyond the ultimate level which informs us of this truth, and the ultimate is not ultimate after all. Indeed, we find there must be an infinite regress of pseudo-ultimate systems. However, this is precisely what Murti was trying to avoid by positing that the ultimate level is not a view.

The inability of the ultimate theory to be ultimate can be understood on the analogy of the inability of parts of a body to be a body. Just as a physical whole is not the same as a collection of parts, so too a totality of criticism does not emerge from a myriad of criticisms. Each individual criticism may conventionally hold, like the parts of the body, but a totality needs to possess certain properties which it can never get from these parts, namely those of wholeness. An ultimate theory must possess the

property of ultimateness, but if none of the steps making up the ultimate theory possess such a property, it is natural to wonder from what the property can be taken to emerge. This line of reasoning can also be taken in the opposite direction, to show the emptiness of the elements that make up a complete view. If an ultimate view posits that there is no finality, as Mādhyamaka does, then it may be asked if any of the components of their reasoning possess finality. This is why Candrakīrti writes in *Madhyamakavatara* 6:176,

> The logical objections of contiguity and so forth would be applicable to us only if we were to maintain that the nature of the reasoning—that which gives rise to the understanding of what is proved—and the nature of that which is actually understood—that which is proved—were inherently existent. But we assert that they have no inherent existence. Therefore, the objections our critics make against us are utterly in vain.

The ultimate truth is itself empty of inherent existence, hence there is no need for Murti to try defend it as not a view in order to protect it from Mādhyamaka reductive analysis. The Mādhyamika readily acknowledges that the application of their system to itself would destroy it. This does not mean that their system is not a view, just that no view is immune from critique.

C. Priest on Nāgārjuna and the problem of expressibility

This problem of a universal judgment being subject to itself and thus undermining itself is a real problem for both philosophy and logic. Priest's book deal with numerous examples in which a self-referential proposition is created by some means, such as diagonal argument in math or simple demonstratives in the Liar paradox. Priest's conclusion, drawn from the numerous examples in his book, is that only by adopting some form of paraconsistent logic can we avoid the Scylla of paradoxical self-reference and Charybdis of logical explosion (which is the result of arguments in the form $\forall x$ ([A & $\sim A$] $\rightarrow$ x). Limitations of time and space prevent us from detailing the full argument here. Instead we turn to Priest's view of the chief patriarch of the Mādhyamika, Nāgārjuna. (The relevant chapter of *Beyond the Limits of Thought* was co-authored by Jay Garfield, but for convenience, I will continue to refer to the argument as Priest's.)

First, Priest addresses the question of whether Nāgārjuna really did want to say, "all views are wrong," and if so, how he overcame the implicit contradiction in doing so. He considers that perhaps Nāgārjuna only appears to be asserting views but is actually merely negating the views of others without asserting any of his own as Murti suggests. However, the textual evidence is against Murti, as Nāgārjuna can be see to openly claim the fact of dependent origination, etc. Another interpretation might have Nāgārjuna merely uttering words that just so happen to cause us to abandon our views without these words meaning anything to Nāgārjuna himself, but Priest finds no evidence supporting this interpretation. Another interpretation might hold that Nāgārjuna's assertions, where they exist, are merely conventional assertions which he ultimately renounces. This too Priest rejects on the grounds that an ultimate assertion is an assertion about what is left after analysis, and Nāgārjuna and his commentators are in complete agreement that what is left is nothing. This must be an assertion about the ultimate. On page 263, Priest concludes,

> There is, then, no escape. Nāgārjuna's view is contradictory. The contradiction is clearly a paradox of expressibility. Nāgārjuna succeeds in saying the unsayable… As Siderits ((1989), p. 231) has put it, "The ultimate truth is that there is no ultimate truth."

(Note that Siderits does not agree with Priest and Garfield's interpretation of his slogan.)

D. Priest on Nāgārjuna's ontology

Next, examining Nāgārjuna's ontology, Priest rejects Murti's Kantian view of the ultimate/conventional distinction on the grounds that,

> The emptiness of emptiness means that ultimate reality cannot be thought of as a Kantian noumenal realm. For *ultimate* reality is just as empty as *conventional* reality. Ultimate reality is hence only conventionally real! (256)

Kant and Murti want to partition out a realm where the usual rules do not apply to save it from contradiction, but the Mādhyamika expressly allow that the ultimate realm is subject to the same limitations as the conventional. However, just positing the emptiness of everything is not without its difficulties. If emptiness cashes out to meaning,

"without an essential nature," then we may claim that everything lacks a nature. However if everything lacks a nature, then we may say that everything shares a nature, the nature of emptiness. Thus, everything both has and does not have a nature, namely emptiness.

There is an additional problem with positing the emptiness of emptiness. We may inquire about whether things just happen to be without essential properties or if it is impossible for anything to be without essential properties. The many arguments of Nāgārjuna are clearly meant to show it is impossible for things to have essential properties. Priest quotes Candrakīrti's commentary, "Things are not without characteristics through characteristiclessness; to be a thing is to be without a defining characteristic" (266). Clearly this is a contradiction if we take "having an essence" to mean the necessary possession of some property. Nāgārjuna and Candrakīrti are arguing that things both do not have essential characteristics and that things do have the essential characteristic of lacking characteristics.

One way out of the two apparent contradictions might be to argue like Kant and Murti that the properties of the ultimate are such that is possible to both have and to lack characteristics without entailing a contradiction. However, we have already seen the difficulty that this position entails. If we are allowed to add an ad hoc escape clause to our reasoning process, why not do so earlier, before reaching the ultimate level? By saying, "emptiness is empty," the Madhyamika is acknowledging that merely calling everything empty runs into the same sorts of contradictions that damned his substantialist opponents.

E. Nāgārjuna and paradox

Thus, there are multiple contradictions in Priest's interpretation of Nāgārjuna. On the level of expression, he says what cannot be said, and on the level of ontology, everything shares the property of lacking properties. To resolve these conflicts, in a note on

page 250 Priest signals his agreement with Tillemans' 1999 "Is Nāgārjuna Logic Deviant or Classical?" in which it is argued that Nāgārjuna employs classical logic on the conventional level but paraconsistent logic on the ultimate level. Thus, in Priest's interpretation, Nāgārjuna's claim is that our mind naturally imputes external objects as the cause of our experience. This imputation is incorrect and moreover positively harmful, as it mires us in the cycle of existence bound by desire. In Priest's interpretation, to escape this means to see experience as experience and nothing else:

> Penetrating to the depths of being, we find ourselves back on the surface of things, and so discover that there is nothing, after all, beneath those deceptive surfaces. Moreover, what is deceptive about them is simply the fact that we take there to be ontological depths lurking just beneath. (226)

Following from this for Priest is the conclusion that experience if taken in itself is not harmful. Since only the reification of experience into imputed objects is harmful, the ultimate truth is nothing but ordinary experience without such reifying:

> Nāgārjuna demonstrates that the emptiness of emptiness permits the "collapse" of the distinctions between the two truths, revealing the empty to be simply the everyday, and so saves his ontology from a simple-minded dualism. (270)

Thus on Priest's view, in terms of the possibilities of the first section, Nāgārjuna subscribes to possibility 3 ("there are no ultimate justifications") on the conventional and ultimate levels by way of possibility 4 ("contradiction") at the ultimate level. The one potential qualm with Priest's view is that it comes dangerously close to ascribing a kind of empirical realism to Nāgārjuna in emphasizing that the everyday is the ultimate. Calling the everyday the ultimate might be interpreted to suggest that we are safe in relying on everyday senses, when, as was strongly emphasized before, this is far from the case. To the contrary, in the Tibetan version of *Bodhicaryāvatāra* 9:167, Śantideva longs to "reveal emptiness to those ruined by reification." This is because while the ultimate collapses into the conventional level, this does not negate that fact that even on the conventional level, things are not what the consensus of experience seems to suggest. Again, while avoiding taking a position in the long running dispute about the nature of the conventional realm within Mādhyamaka, we note that even

Svatantrika school will claim that the conventional realm can be broken into momentary dharmas and the like. Hence *Bodhicaryāvatāra* 9:8, "There is no fault in the conventional truth of the contemplatives." That is to say, it is those who subscribe to the Buddhist world view who are without fault and not those who subscribe to the conventional truth of ordinary understanding. Nevertheless, outside of this potential misstep, Priest's interpretation of Mādhyamaka is sound.

VI. Conclusions

A. Mādhyamaka and the end of analysis

With this understanding in mind, we are now able to return to the question that originally engaged us: What are the implications of the analysis conducted in the Diamond Splinters argument? First of all, if the Mādhyamika are right, analysis of this sort both has and does not have a stopping point. There is no stopping point in the sense that any object of analysis can be further analyzed into more basic points. This is helpful for the Mādhyamika because if there were such a stopping point, the unanalyzable objects could potentially be objects for craving. There is a stopping point in the sense that it is possible to realize that there is no stopping point, and thus there is no need to conduct further analysis in order to find one. This also serves Mādhyamaka soteriology, since if it were not possible to know that there is no stopping point for analysis, the mind would never be at rest, but instead it would continually analyze objects deeper and deeper in the hopes that a stopping point might later be found. The difficulty is that since there is a stopping point, the mind is tempted to reify emptiness itself into an object worthy of grasping. However, those who understand the emptiness of emptiness will properly refrain from doing so. This is why Śantideva says in *Bodhicaryāvatāra* 9:34, "When neither an entity nor a non-entity remains before the mind, then since there is no other possibility, having no objects, it becomes calm." Thus, the contradiction is that one must cease analysis with the knowledge that there can be no cessation to analysis.

However, have the Mādhyamika avoided the trap the Murti was attempting to avoid when he raised the question, "Is śūnyatā a theory?" The principle of recursive analysis lead the Mādhyamika to renounce all views, but does such a renunciation constitute a view that is any different from, "All views are wrong except mine," which anyone could endorse? Also, if the Mādhyamika answer to this difficulty entails a contradiction, why should analysis cease where the Mādhyamika suppose it does, rather than at some other contradiction, such as those lead to by the theories of substantialists?

The difference between Mādhyamaka views and the views of their opponents is that Mādhyamika follow their principles past the limits of their implications before abandoning them, rather than begging off a contradiction sooner. Any theory could propose an ad hoc explanation before critique is embarked upon. We could just propose that everything that happens is the will of some inscrutably unanalyzable spirit at step one and be done with it. However, a more useful explanation will always go deeper until there is no deeper left to go. Thus, we might attempt to beg off the contradictions inherent in thinking of the experience of objects as proof of objects as things in themselves by saying that since we have no epistemic access to things in themselves and neither can we even theoretically encounter objects in any manner other than through experience, things in themselves are a mystery and leaving it at that, but logical consistency demands that we dig deeper before allowing ourselves the luxury of recourse to the unknowable. Through the *reductio ad absurdum* argument, the Mādhyamaka demonstrates the intellectual laziness of the supposedly logical opponent who has ceased analysis before it turns on itself. Thus, the following exchange takes place in *Bodhicaryāvatāra* 9:109–110:

> 109. *[Objection:]* But if one analyzes by means of analysis which is itself analyzed, then there is an infinite regress, because that analysis can also be analyzed.
>
> 110. *[Mādhyamika:]* When the object of analysis is analyzed, no basis for analysis is left. Since there is no basis, it does not arise, and that is called "*nirvāna*."

The Mādhyamika analyzes even past the point of absurdity, and thus nothing remains to be analyzed, and the mind is pacified.

B. The meaning of illusion and explanation

Next, what is it that Mādhyamaka analysis reveals? Yes, everything is an illusion, but what does that mean? First, it is worth defining illusion itself. Illusion is taking one's experience to imply something that is not in fact the case. In the case of an ordinary illusion, I may, for example, see a train speeding toward the movie screen and react with fear. This is an incorrect reaction, because the train is not actually dangerous to me. My impression of the danger of a train coming towards me is an imputation that causes me to draw incorrect causal inferences. On the other hand, the fact that I am experiencing an image of a train coming toward me is not controverted by the explanation that the "train" is only an image of a train. The common lesson of Kant and the Mādhyamika is that all of experiences are only the experience of appearances, not the direct experience of objects, which is logically impossible. Contra Magritte's *The Treachery Of Images*, even a real pipe outside of a painting is not a pipe. It is only seeing, touching, feeling, and smelling a pipe. Thus, the picture of a pipe and the ordinary pipe are alike in their both not being independently existing pipe-objects. On the other hand, the causal inferences that one can draw from a painting of a pipe and the causal inferences one can draw from an ordinary pipe are quite different. A painting of a pipe cannot allow one to experience smoking, but an ordinary pipe can. Returning to the train example, if from the film of a train coming toward the screen one drew the conclusion that one's life were in danger, this would be a mistake, but to draw the conclusion that the "life" of the girl tied to the railroad tracks in the movie is in danger would be a correct inference. Thus, even illusions have their realm of causal efficacy, it is just that their causal efficacy is not as broad as it may appear to one unaware of the illusion. As Candrakīrti says in

Madhyamakavatara 6:175, "Even though a reflection is not true it can still establish whether or not one's face is clean."

The one difference between Mādhyamaka analysis and the usual explanation of an illusion is that ordinarily proffering explanation is predicated on the real existence of something deeper that explains the illusion. The Mādhyamaka thesis is that analysis can always be applied recursively until there can be nothing deeper that escapes analysis, for the reasons given earlier. Deeper explanations in general come in two types: those that replace the previous explanation and those that merely subsume them. Newton's gravity replaced Aristotle's entirely, only to be completely replaced in turn by Einstein's (though Newton's gravity lives on as a convenient means of estimating). On the other hand, learning that subtraction can be generalized as the addition of negative numbers does not invalidate the use of ordinary subtraction, and the discover of higher realms of non-Euclidean geometry does not preclude relying on the same inferences within an explicitly Euclidean context. For Mādhyamaka, the conventional result of analysis is that which causes us to revise our estimation of the causal efficacy of things in experience. For example, I may think of myself as a independently existing self, but if analysis reveals that I am empty of self then this has direct application on the estimation of my causal efficacy. On the other hand, though on the ultimate level the result of analysis is finding that there is no ultimate ground for being, this has no direct impact on our estimation of the causal efficacy of objects in the world. The only impact of this discovery is on my process of analysis itself. (That is, my analysis can cease, since there is no ultimate being to discover.) Conventional analysis reveals what we do not have here and now; ultimate analysis reveals what we do not have and cannot hope for.

C. Implications

Starting from an entirely different perspective, Rescher comes to similar conclusions about the problem of ultimate analysis from a scientific perspective. Any ultimate analysis must be subject to itself as well, no matter what difficulties this entails:

> We clearly cannot provide a scientific explanation for the whole system of science in terms of something that falls outside: it would not be the whole system if anything fell outside it. Explanatory self-subsumption is infeasible at the level of facts and laws. But at the systemic level it is a conditional necessity: if the system can be explained at all, that explanation must fall within it. As long as we operate on scientific principles, we canot get outside the framework of our completed explanatory system: to explain the system in terms of *X* would simply be to enlarge it to include *X* itself. The quest for a system-external foundation for the scientific rationalization of the system is ultimately senseless. (88–89)

Since we cannot go posit anything outside of the explanatory system itself, the system becomes an interdependent one whose value is judged based on its systematic unity as a whole:

> Thus, we have here a wholly different approach to explanation; one that takes systematization itself as the key, relying not on subsumptive inference but on systemic coordination. (85)
>
> The issue of legitimation is thus settled in terms of a cyclic interdependence and self-supportiveness. The idea of explanatory stratification is misleading: no neat linear order of fundamentality obtains among nature's facts of laws. (90)

Thus even from a scientific perspective something like dependent origination is necessary ("cyclic interdependence"), though of course Rescher's proposal is nowhere near as sweeping as Mādhyamaka, in part because he is considering justification from a scientific basis that takes external objects for granted rather than an epistemic one that questions their very possibility. What is the meaning of Mādhyamaka-style ultimate analysis? In conclusion, we can say that the illusory nature of objects under analysis implies that there are no ultimate justifications, though this comes by way of contradiction on the ultimate level, since the interdependence of things might otherwise be taken as such a justification. In terms of conventional reality, objects must be analyzed in terms of their conventional efficaciousness. Even on this level, however, we will find that the external objects which reason compels us to impute are not ever to be found in actual experience. Thus, Kant and Murti were on the right track with their dialectic

method, but they stopped short of the full implications of their method by withholding the deeper realms they proposed from analysis. Priest's interpretation of Nāgārjuna and his paradoxes is mostly correct, although his emphasis on the identity of the conventional and ultimate realms comes dangerously close to endorsing our commonsensical notions of the world, which Nāgārjuna would not do. Illusions are ordinarily interpreted as real perceptions with real but hidden causes that give rise to the mistaken inference of illusory causes, but in the case of universal illusion, we can say (though we contradict ourselves to say so) that there are no real causes nor can there be, since the perceptions themselves are not real.

Daoism and Computation

I. Introduction

In order to better understand the meaning of *daode*[34] implicit in the *Daodejing* 道德經 (also called *Laozi* 老子), this article will contrast it with a paradigm more familiar to contemporary thought, computation. This comparison is made in light of the numerous proposals that our world is fundamentally computational (as in Stephen Wolfram's *A New Kind of Science* and Gregory Chaitin's *Meta Math!* recently[35] and presaged by John Wheeler's "It from bit" concept and Konrad Zuse's *Calculating Space* and as expressed in pop culture by countless movies and stories) and the seeming similarities between those proposals and the process ontology often associated with contemporary Daoist studies.

Computation and the process interpretation of Daoism do share certain similarities, not the least of which is that in both computation and *daode*, one system serves as the microcosmic simulation or representation of another. Moreover, computation shares with *daode* a blurring of the boundary between theory and practice, since it takes the most theoretical of disciplines—mathematics—and drags it down in the practicality of observable physical transformations, just as Chinese thought emphasizes the importance of concretely utilizing *daode* as a basis for understanding the ineffable.

However, a computational view of *daode*, while helpful as a starting point for understanding *daode*, ultimately overlooks the fundamental differences between the mechanical nature of computation (the possibility of non-determinism within

34. This essay concerns the *daode* 道德 rather than "the Dao," since in Daoist thought, a field (*dao*) is not experienced outside of its particular focii (*de*). Indeed, Daoism in the first instance was first known as the school of "daode," and only later became "Daoism" without reference to *de*. Sima Tan in the Records of the Grand Historian (later completed by his son Sima Qian) was the first to designate the Six Schools, one of which he called the school of "Daode." See Shiji 世紀 (Shanghai: Zhonghua, 1959), p. 3118.

35. Although, *A New Kind of Science* has had relatively little impact outside of the promotion of Wolfram himself and *Meta Math!* is a popularizing turn rather than a scholarly work, I take them as emblematic of a new kind of thinking that is increasingly prevalent. In one sense, the computational model of the world is no different from the once popular Laplacean determinism except in terms of its emphasis on iterative calculation over algebraic continuities. Nevertheless, given the prominence of this kind of thinking, the author considers it to be worth considering the claims of "digital physics" in ontology to be a serious parallel movement to computational functionalism in the theory of mind.

conventional computational system not withstanding) and the creative nature of *daode*, which emerges from the encompassing nature of *dao* and the particularizing nature of *de*.

In order to properly make an evaluation, I will begin with an appropriate definition of computation. With this definition in place, I will use a case study to point out the many correspondences between the computational modeling and the Daoist modeling, before ultimately refining this understanding of *daode* through contrasting *daode*'s inclusive, creative totality of insistent particulars with the privileged perspectives implicit in the computational model.

II. Definition of computation

I begin by noting a shortcoming in much of the existing literature concerning computation. Very often when one examines a seminal work in computational theory—eg. Alan Turing's *On Computable Numbers*—one observes that the author defines computation only in terms of the theoretical capabilities of one particular kind of mechanical procedure and not computation in general. Though mathematicians and computer scientists have shown no reluctance in defining new and different forms of computers and computational procedures in order to explore their limitations, nevertheless, it is difficult to find a definition of computation in general. In part, this may be because it is widely believed that no computer realizable in our world can have capabilities surpassing those of the machine conceived in Turing's paper. Though the definition of a universal Turing machine is quite simple—a machine places a read-write head over an individual square on a tape of unbound length and then moves the read-write head or writes a symbol, depending on the instructions associated with the symbol on the tape and the internal state of the machine—it was found to be theoretically capable of carrying out any determinate algorithm that a human mathematician can and equivalent to Alonzo Church's notion of effective calculability. Thus, work in theoretical

computer science suggests that for practical purposes to be computable is to be computable by a universal Turing machine, and hence it is assumed that computation is best understood as activity of the sort doable by a universal Turing machine. Of course, the thesis that no physical device in our world can surpass the capabilities of a universal Turing machine remains unproven and, in all likelihood, is unprovable, since it is a conjecture about the physical nature of the empirical world.

One interesting observation regarding the question is one made by Richard Feynman. In "Simulating Physics with Computers," he noted that ideally, our world should be perfectly simulable by a classical computer and that "the number of computer elements required to simulate a large physical system [should be] proportional to the space-time volume of the physical system… If doubling the volume of space and time means I'll need an *exponentially* larger computer, I consider that against the rules (I make up the rules [of this thought experiment], I'm allowed to do that)."[36] However, he went on to note that classical computers appear to violate this rule and use exponentially more resources to simulate a quantum interaction than is proportionate. Accordingly, he proposed that physicists and computer scientists work around this apparent difficult by performing simulations using other quantum objects in order to escape this exponential growth in resource requirements. While the issue of the relative power per given resource for quantum versus classical computing systems has not been resolved and furthermore, quantum systems are, in principle, able to be modeled by non-quantum computers (albeit using disproportionately more resources than the system modeled), the presence of this discrepancy suggests that we take seriously the possibility there are physical processes that seem to allow us to compute more than we "should" be able to compute using a given number of resources if we take the universal Turing machine as the ostensive definition of computing.[37]

---

36. Feynman, 134.

Therefore, in order for one's definition of computation to be truly general, it must also accommodate the possibility of various forms of "hypercomputer" such as a computer that can perform calculations using the complete set of real numbers or a computer with access to an "oracle" function capable of predicting whether particular algorithms will halt. Indeed, if the world itself is a computer, then one possibility is that the seeming indeterminacy of quantum physics is merely a byproduct of the difficulty of performing quantum calculations in computers that are only classically Turing equivalent.

One happy exception to the general pattern of ignoring the need to distinguish between specific models of computation and computation in general is B. Jack Copeland's *The Broad Conception Of Computation*. In it, Copeland distinguishes between "Computable" with an uppercase 'C' and "computable" with a lowercase 'c.' The first refers specifically to computability by some Turing equivalent machine; the second refers to computability by any conceivable computing machine. Copeland describe the computing machine in general as follows:

> The general requirements for a computing machine are simple. Part of the machine must be capable of being prepared in configurations that represent the arguments of whatever functions are to be computed, and part of the machine must be capable of coming to assume, as a result of activity within the machine, configurations that represent values of these functions. Sub-devices of the machine—"black boxes"—must make available some number of primitive operations. The machine must have resources sufficient to enable these operations to be sequenced in some predetermined way. …[R]ecursive application, or iteration, of primitives is the essence of computation.[38]

Here we can see that Copeland has laid out a much broader notion of computation than is generally considered. However, there are interpretative hazards when taking this definition to encompass all of computing. First, it can be criticized for requiring "black

---

37. Note that there are two senses in which one computer can be more "powerful" than another. The first is one computer can, if given enough resources, solve a problem that another cannot, even if it is also given as many resources as it might want. In this sense, there is no question that a universal Turing machine is exactly as powerful as a quantum computer if both are allowed to have as much time and memory as needed. "Hypercomputer" is a term for computers more powerful in this first sense. The other sense of powerful is how many resources are needed to perform a certain family of calculations. It is in this sense that quantum computers appear to be more powerful than ordinary computers, although this has not been shown definitively.
38. Copeland, 695.

boxes" as sub-devices of the computer. The term "black box" is of course meant to suggest that the internal implementation details of the computer are abstracted away, so this definition is not restricted to a specific kind of implementation of a computer as the definition of Turing machine is. However, even using the term "black box" at all strongly suggests the existence of internal (that is to say, systemically isolated) mechanical process by which values are calculated. A truly broad concept of computation cannot be so limited if it is to encompass the full range of possible non-digital computations, since unlike a black box, the functioning of some computers may not be so easily divided into internal components and external observers. Instead, the two may be dependent on one another, with information more readily usable at their juncture, but the coordination of both the internal and external necessary for the operation of the device. Thus, the use of black boxes seems to suggest that the computer cannot use knowledge about the world as a whole (including even possible observers) as a part of the process of creating new information for an observer's consumption. Hence, by restricting computing to machines with black boxes, this concept seems to encompass only "complete" computers like universal Turing machines and the von Neumann architectures of contemporary personal computers and to exclude calculating tools like abacuses, in which the operation of the calculating device is partly dependent on the human operator even after the initial programming phase of computation is complete. With the abacus, like a logarithmic table or even the lowly poster of the multiplication table common in elementary schools, a physical artifact is used to help make simple correlations between arithmetic entities. While an abacus may not seem to be "automatic" enough to count as a computer to some, I argue that the difference between the levels of operator involvement in the use of an abacus and an electronic calculator is one of degrees rather than kind. Just as you cannot learn the answer to two plus two using an electronic calculator without pressing 2, +, 2, =, and then interpreting the result on the screen, so too with an abacus, the operator cannot learn the answer to the same

question without first sliding two beads, then sliding two more, and finally interpreting the resulting physical state of the abacus.

Further, the requirement of "primitive operations" being available for the computing machine seems to rule out the possibility of a physical artifact with an unbound number of possible state transitions. For example, it is not clear that a slide rule must in principle have only a finite number of primitive operations. Obviously, any actually existent slide rule will be constructed only within certain physical tolerances, and motions less than what is allowed by its construction tolerances cannot be counted on to give accurate results. In the limiting case, no physical slide rule can be accurate at sizes less than one Planck unit, the smallest usable distance. Hence, moving left or right by a distance equal to the construction tolerances of the slide rule can be considered its primitive operation. However, these limitations are practical ones, rather than theoretical ones. For an ideal slide rule, any motion—no matter how small—will result in a slight change in its output. In such a case, it is not clear that there are any truly primitive operations. Further, we might envision a computer in which the primitive operations of the computer change over time, as the machine "evolves" mechanically. In a science fiction scenario, an intelligent computer/robot may upgrade itself periodically, resulting in an ever shifting basis for its calculating substrata and unfixing the number of its primitive operations. More seriously, the precursors of today's internet have been in continuous operation since the late nineteen-seventies, during which time its basic protocols have been revised substantially. One can easily envision a network of computing devices which is also used without interruption to work on the same calculation for years, during which time its most fundamental computational substructures are completely replaced without halting the overarching activity of the network. Supercomputing networks of this sort are already commonly employed by large corporations like Google.

Finally, I note that caution should be taken when using the language of "computing machine," as it seems to suggest the necessity of human artifice for the construction of the machine, whereas if computation is possibly the fundament of the world, human artifice is ruled out *ex hypothesi*.[39] John Wheeler, a proponent of digital physics, proposes that we reject language of "machine" because it "has to postulate explicitly or implicitly, a supermachine… which will turn out universes in infinite variety and infinite number."[40] Accordingly, our definition of computation would better if it encompassed the possibility of computing without human involvement. Of course, there remains the possibility that there is a Divine Programmer who understands the world program and in reference to Whom the universal computing machine is a machine, but in such a case, it is God rather than the program which is truly fundamental, and the paper is concerned with the possibility that *daode* is a computation-like fundament, not a God-like fundament or a tool used by God. If possible, a definition of computer that does not depend the anthropic perspective would be preferable. (The question of whether such a definition is possible will be explored in great depth later in this paper.)

In order to work around these interpretative difficulties with Copeland's definition of computing machine, the definition of computation that is proposed here takes the activity of simulating to be the ground of all computation. In order to encompass as broad a range of possible computations as possible, we must not mistake numbers, discrete output, or algorithmic processes for what is mathematical in the most fundamental sense. As Heidegger says in "Modern Science, Metaphysics, and Mathematics," "the essence of the mathematical does not lie in number, as purely delimiting the pure 'how much,' but vice versa,"[41] since the mathematical is "the fundamental

39. Excepting the scenario envisaged by Isaac Asimov's science fiction classic "The Last Question."
40. A quote from Wheeler's "World as system self-synthesized by quantum networking," contained within his "The Search For Links" page 314.
41. Heidegger, 277.

presupposition of the knowledge of things."[42] Thus, as a presupposition of knowledge, what is fundamental to the concept of mathematical computation is not the ability to do arithmetic but an ability to predictively simulate some system in advance through the prior comprehension of the system. For example, in a wind tunnel, a model airplane is used to predict the behavior of a real airplane in flight. This device for making predictions connects one physical system to another without stopping to manipulate anything that "represents" numbers or other mathematical entities within its system, but the calculations that it does are no less deterministic and useful within their known tolerance levels for this.

Thus, I agree with the first part of Copeland's proposed definition and take it that the key to any computer is its ability to "represent" various system states and their connection. To put it succinctly, computation is "**the process by which the fact that one system is rule governed is used to make reliable inferences about another rule governed system**." This definition overcomes the interpretive difficulties of Copeland's definition by emphasizing that computation is a systematization of analogs. Thus, in the case of the abacus, the fact that the beads won't merge or divide spontaneously allows us to use them reliably for making finite determinations about the relationships of natural numbers. Similarly, even an idealized slide rule, an evolving computer network, or a wind tunnel can be seen as a rule governed system, even if those rules are not easily enumerated or even fully grasped by an observer of the computing system. Finally, this definition allows for any rule governed systems to be correlated, whether these systems are natural or artificial. Accordingly, the remainder of this essay will be conducted with this definition of computation in mind.

---

42. *Ibid*. 278.

III. Correlations between *daode* and computation

With our new definition in place, certain similarities between *daode* and the computer become clear. In both, the macrocosm is correlated with the microcosm. As a case study to illustrate this connection, in this section I will contrast the process of divining the weather using the traditional "oracle bones" method of ancient China with modern weather predicting computers. During the Shang and early Zhou dynasties, diviners would interpret the cracks that emerge on turtle shells and other bones on the theory that the state of the world as a whole would influence the patterns that emerged on the shell's surface. When a proper ritual rubric is carried out, various inferences about the future can be reliably made. For example, one might ask whether the upcoming harvest season will be auspicious or inauspicious. Though the process by which these predictions were carried out appears completely dissimilar to modern methods of weather prediction, there is an underlying unity of form. In both processes, first proper preparations must be made in order for the process to work correctly. Then a complicated series of correlations are carried out by a physical substratum. Finally, the changes in the physical substratum are interpreted as an answer to the question posed.

The apparent dissimilarity between these two activities is due to an understandable modern bias: the belief that oracle bone predictions do not work. If we bracket the concerns that we may have about whether divination actually works, then we must say that if it did work in the manner that its practitioners believed it to, then the reality that allows it to work is a process of correlations that falls under our definition of computation. One rule governed system (the oracle bone divination equipment or the computer simulation) is used to make reliable inferences about another rule governed system (the world as a whole). The reason that both systems work is that the microcosm, be it of the computer chip or the diviner's tools, is made to reflect the macrocosm, and once the correlation is firmly established a change in one can be read as a change in the other.

Of course, before accepting this analogy between a computer prediction and a divination, we need to reexamine each of the parts of the process—the preparation of the system, the change of the physical substratum, and the interpretation of result—before finally the system as a whole is examined and the overall relevancy of the connection between the two systems is assessed.

A. Preparation

In the preparation of the process, it may be objected that the preparation carried out in a computer simulation is the gathering of data, that is numerical facts, whereas the preparation process in divination consists primarily in the invocation of a question following a long ritual process of cleaning and polishing the bone to be heated. However, two points can be made regarding these dissimilarities. First, the reason that data is not gathered in a divination is that the gathering of data is assumed to be superfluous, since in traditional Chinese cosmography, everything is a correlate of everything else, so the way to learn about what is most distant is to study what is closest at hand.[43] Thus, it is important that our definition of computation allows for the "black box" of the computer to be non-self-contained and interconnected with the system that surrounds it (hence not a "box" at all).

Second, that divining relies on the invocation of human language is not necessarily damning to its computational equivalence or proof that what is being interacted with is primarily a human-like and non-computational spirit. As David Keightley explains in *Sources of Shang History*, "the scapulimantic and plastromantic inscriptions were not simply regarded as prayers or magical letters forwarded to the spirits."[44] Rather, the inscriptions are used to track the accuracy of the divining process and thus to ensure that processes which produce more accurate results are retained and refined and those who master them are properly rewarded. It is the divining process, not the

43. *Cf. Daodejing* chapter 47, "Venture not beyond your doors to know the world…." Translation from Hall and Ames (2003), 150.
44. Keightley, 45.

inscription, which yielded interpretable results. While there is perhaps some anthropomorphic element to Shang era divination, by the time of the emergence of Daoism, *tian* 天 (often translated as "Heaven") had come to be seen as an impersonal force which orders the world.

Furthermore, modern computer programmers regularly use human language to interact computationally with the physical world. In the programming language Python, one makes the computer display the message, "Hello World!" by giving the input `print "Hello World!"`. As Abelson and Sussman write in their classic programming text *Structure and Interpretation of Computer Programs*,

> A computational process is indeed much like a sorcerer's idea of a spirit. It cannot be seen or touched. It is not composed of matter at all. However, it is very real. It can perform intellectual work. It can answer questions. It can affect the world by disbursing money at a bank or by controlling a robot arm in a factory. The programs we use to conjure processes are like a sorcerer's spells.[45]

Hence the difficulty with seeing ancient Chinese divination as computational is not that human language is used. The difficulty is whether it is plausible that their language can really be transformed reliably into other physical processes that in turn can correlate properly with the world as a whole. However, this difficulty can be bracketed for the length of the investigation here, since our concern is with the ancient Chinese understanding of how *daode* operates, not whether that understanding is correct.

B. Physical change

The change of the physical substratum presents another area for critical consideration. In a digital computer, the electronic signals of data recording devices are made to correlate with electromagnetic fields in other calculating devices, then the pattern of these gives rise to new signal patterns which propagate and interact before finally leading to their expression as an image on a monitor or a sheet of paper. In contrast, the changes in the physical substratum of a divination seem more straightforward. Heat causes the material of the bone to expand until it cracks. However, if we ask why the bone cracks

45. Abelson and Sussman, "Building Abstractions with Procedures."

where it cracks, then the process of explanation is much more involved. The bones used for divination represent a "found" computation. It is not a product of human artifice that bones used will crack when heated properly. However in a Chinese cosmology that the bones do crack precisely where they do is a matter of their interrelation in the cosmology of *yin* 陰 and *yang* 陽 leading to a particular physical outcome. Of course, this interrelation is not merely a matter of simple causality in our contemporary sense. As A.C. Graham explains in *Disputers of the Dao*, causal thinking was not present in the sciences of ancient China. For this reason, as he explains,

> It is not just that the explanations of Chinese as of Western Medieval and Renaissance proto-science may impress us as obscure or fallacious like the arguments of the philosophers; the trouble is that for post-Galilean science they are not explanations at all.[46]

As correlative rather than causal systems, the cosmologists of these sytems did not emphasize the ways in which certain isolated states of matter necessarily give rise to one another as contemporary science does. As Graham states, the world of Chinese proto-science is, "a world is which not, like that of Newton, bound by invariable law."[47] Nevertheless, ancient Chinese cosmology is still at least a nondeterministically rule governed system along the lines of a quantum system, rather than a truly lawless anarchy. As Graham goes on to explain about the Five Processes (*wu xing* 五行) of Chinese cosmology,

> It was noticed from an early period that the processes conquer each other in a regular cycle, water quenching fire, fire melting metal, metal cutting wood, wood digging soil, and coming round again with soil damming water…[48]

The *Daodejing* is also presupposed on the indeterministic but rule governed nature of the cosmos. As chapter 16 explains in part,

> In the process of all things emerging together (*wanwu* [萬物])
> We can witness their reversion.
> Things proliferate,
> And each again returns to its roots.[49]

---

46. Graham, 320.
47. *Ibid*. 354.
48. *Ibid*. 326.
49. Hall and Ames (2003), 99.

Thus, all change is based around a common center point from which things reliably emerge and to which things reliably return. Change is neither predictable nor chaotic, but appropriate. Furthermore, it is clear that in the processual nature of *daode*, physical transformations that take place on one level of reality always reflect a micro/macro-cosmic recapitulation these similar transformations at every other level of reality. As the end of chapter 25 of the *Daodejing* states,

> Human beings emulate the earth,
> The earth emulates the heavens [天],
> The heavens emulate way-making [*dao* 道],
> And way-making emulates what is spontaneously so (*ziran* [自然]).[50]

The character translated as "emulates" is 法 (modern Mandarin *fa*), a term that would later by taken by the Buddhists to translate dharma and was also emphasized by the Standardist school. In modern Chinese and Japanese, this same character is used to translate the Western concept of a law. Its translation here as "emulation" correctly emphasizes that changes which arise in this world are rule governed in such a way that each level of activity can be correlated with each other level, and that in order to gain greater insight into one particular level, it is often pragmatically useful to study some other level that is closer at hand instead of looking directly at the level towards which an inquiry is posed. Clearly, such a method of acquiring information accords with the definition of computation employed in this paper.

C. Interpretation

Concerning the issue of interpreting the results of the computation or divination, A.C. Graham seems to suggest that divination in the *Yijing* 易經 (and by extension oracle bone divination) does not really work according to the mechanism accepted by its practitioners, but rather it works by allowing for the tension between the heuristic of the divination and the reality of the situation perceived by the diviner to be worked out through creative interpretation.[51] Thus, divination is an aid to creative thought and not

50. *Ibid*. 115.
51. Graham, 368.

an direct correlate of the physical world. Seemingly at the opposite extreme, François Jullien in *The Propensity of Things* can be interpreted as ascribing to Chinese thinkers a belief that *shi* 勢 (a term for "momentum" that bridges *dao* and *de*) has a correlation with the actual world strong enough that "they could detect in warfare's unfolding a purely internal necessity that could be logically foreseen and, accordingly, perfectly managed."[52] Between these two views of the actuality of the connection between Daoist processes and the world, we must note that even an ordinary computer simulation must be itself interpreted. The numbers on a screen that result after a computer has finished its calculation means nothing in themselves, but rather, they take on meaning only when they are interpreted in light of the mental worlds of the programmer and the user. Although Jullien notes that *shi* inserts itself, "into the distinction between what Westerners have opposed as 'practice' and 'theory,'"[53] we can say the same about computation as well. Computation takes the most theoretical of disciplines—mathematics—and drags it down in the practicality of observable physical transformations.

D. Process as a whole

Finally, let us consider a computer program as a whole. If we have the text of a program like the aforementioned `print "Hello World!"`, by itself we have nothing, since a program cannot be run without a system. Furthermore, in one important sense, the text of the program is not the program itself. Neither is even the machine code (the "ones and zeros") to which the text program is converted the program itself. The program as executed is a real collection of electrons and magnetic fields here, a real absence of electrons and magnetic fields there, and the real potential for various interactions between the particles and fields. The existence of a program as we think of it is an abstraction—and thus in one sense "unreal." When we define a computation as the reliable

52. Jullien, 25. Note that Jullien's own view of the role of *shi* does not seem to be this mechanized. However, since his language allows for this interpretation, I will use it here as a convenient foil.
53. *Ibid*. 38.

inference about one rule governed system from another, the system that is being inferred about using the physical system is a logical system where the text of the program is true. Thus, in computing the program, we bring it out of the theoretical by tying it with some actual—the actual physical make up of the computer running the program. What Hall and Ames say of the *dao*, we can also say of the relation of the text of the program to the physical substratum of the program: it is merely a convenient collection of "thises and thats" that allow us to make meaningful abstractions about the real world. This convenient collection must necessary smooth over some important physical possibilities. Most notably, since a computer is a "reliable" means of correlating systems, thinking computationally means ignoring the possibility that the computer could malfunction or be turned off. By willfully ignoring these possibilities, however, the programmer gains the reward of taking what is beyond our comprehension (the swirling mass of countless swarms of subatomic particles) and bringing it to within our comprehension (the program text) in service of finding some result that we have not yet brought out from our comprehension of its basis into reality (the result of running the program).

The same relationship that holds between a program and a computation can be said to hold within *daode*. Hall and Ames give a "bare-bones reading" of *Daodejing* chapter 42 thus,

> *Dao* engenders one,
> One [engenders] two,
> Two three,
> And three, the myriad things.[54]

They also give a fuller translation,

> Way-making (*dao*) gives rise to continuity,
> Continuity gives rise to difference,
> Difference gives rise to plurality,
> and plurality gives rise to the manifold of everything that is happening (*wanwu*) [55]

54. Hall and Ames (2003), 143.
55. *Ibid*. 142.

We can interpret the passage in the following way. *Dao* is the condition necessary for transformation to occur: the system as a whole. As the *Yijing* says,

> Thus, that which goes beyond form is called *dao*; those things that have form are called phenomena. The transforming and tailoring of things is called flux. The extending and applying of things is called continuity. To take up this understanding and bring it into the lives of the common people is called the grand undertaking.[56]

Within the system when a particular foci or *de* arises, this naturally begins a process of change and counter-change that results in the complete structuring of the ten-thousand things. Under this interpretation, *dao* is the computer; *de* is a program; the generation of the ten thousand things happens as a result of their iterative interaction. For humans to understand the world, we must understand the system that gives rise to the world, then apply this understanding to specific cases through the grasping of their particular foci, then turn that understanding to our own advantage as we use the iterative nature of the world to our own advantage. As chapter 64 of the *Daodejing* states, we must "Deal with a situation before it happens."[57]. Our ability to do so is in turn dependent on our understanding of *daode* giving rise to actions that best conform to the eventual happenstances of the evolution of *daode*.

IV. Differences between *daode* and computation

Having made the case for interpreting *daode* as computation, I will now refine this interpretation by noting the areas in which our previous interpretation fails to properly characterize Daoist thinking. There are several categories into which these failures fall. First, there are problems with attempting to give a single characterization to the *dao*. Next, there is the importance of *de* to realizing *daode*. Finally, there are general problems with taking computation as primordial.

A. Difficulties with characterizing *dao*

To begin, let us turn to the very first words of the *Daodejing*,

> Waying-making (*dao*) that can be put into words is not really way-making,
> And naming (*ming* [名]) that can assign fixed reference to things is not really naming.[58]

56. Translation from *Yijing*, Xici 1, 12 courtesy of Roger Ames.
57. Hall and Ames (2003), 177.

With these words in mind, it is clear that any concept of *daode* as computation will run into serious difficulties if it naïvely assumes that *dao*, like a computer program, can be given a single, fixed explanation. There is no single "source code" that serves as a privileged interpretation of *dao* as there is for a computer program. I have already argued against the conception of *daode* as either a simple, causally determinate system or an absolute chaos. As chapter 34 of the *Daodejing* states, "Way making (*dao*) is an easy-flowing stream/ Which can run in any direction."[59] Literally, the passage states that the *dao* "can go left and right." This strongly suggests that the spontaneity of *daode* cannot be pre-emptively circumscribed by human beings. The difficulty then is in determining to what degree *daode* can be considered to be rule-governed, when the rules that govern the unfolding of *daode* are not extrinsic to it, but intrinsic yet unfixed.

Simple indeterminism is not enough to dissuade us from a computational interpretation. Chaitin's constant Omega, for example, is a mathematical entity that is essentially random, in the sense that it cannot be understood in terms of anything simpler than itself, yet it can be formally defined in a relatively simple manner as the halting probability of any algorithm. Thus, by building on the work of Turing and Gödel, Chaitin has demonstrated a "random" number that can be constructed within the rule governed system of mathematics. The trick of Chaitin's constant and related constructs is that they are only random in the sense that their value is externally inexplicable. It is perfectly determined internally, but working out precisely what it is determined by the simple definition of the constant is an intractable problem that would, in principle, require more than an infinite amount of conventional computations effort because of its incredible information density.

Whether the unpredictability of *daode* is like Omega or not depends upon whether its ineffability is epistemic or ontological. Algorithmic randomness relies on

58. *Ibid*. 77.
59. *Ibid*. 130.

epistemic ineffability to conceal its ontological determinacy. Whether *daode* also relies on epistemic limitations in order to conceal ontological unity cuts at the heart of a central question of contemporary Daoist studies: Is *daode* transcendent or emergent? If *daode* is in fact transcendent, then there is in fact a constant *dao* that exists independently of our knowing. If *daode* is emergent, then *dao* and *de* are self-arising and without an external source of structuring. This paper cannot hope to fully settle the question of the transcendence or emergence of *daode* within the limits available here, but some remarks will be made.

In *Thinking from the Han*, Hall and Ames define strict transcendence as "*A* is transcendent with respect to *B* if the existence, meaning, or import of *B* cannot be fully accounted for without recourse to *A*, but the reverse is not true."[60] In the case of computation, transcendence is encountered whenever one of the rule governed systems being related exceeds in its actuality the representational capacities of the other system. Thus, for an ordinary computation, we can say that the physical system transcends the computational interpretation imposed upon it, since there is a loss of information as one goes from the messy reality to the smoothed over interpretation of that reality. (That is, there are always a myriad of physical states that have the same computational effect, since, as previously mentioned, while we talk about the current in the wire not being a "one" or "zero," in reality, there are small, insignificant fluctuations in voltage that are covered over by the interaction of logic gates, and are thus unknowable on the basis of the output of the computation alone.) In a computational world, this order of transcendence would be reversed as the physical world would ultimately depend on the epistemically inaccessible Great Computer. Like the Kantian noumena or the physicist's Theory of Everything, the Great Computer as the reality behind the illusion would only be the subject of speculation, not knowledge, for us physical beings. If *daode* is like the

60. Hall and Ames (1998), 190.

Great Computer, then it too should transcend the physical world. This view, however, is not as well supported by classical Daoist texts in the opinion of this author.

For example, chapter 62 of the *Daodejing* ("Way making (*dao*) is the flowing together of all things (*wanwu*)."[61]) strongly suggests the emergence rather than transcendence of *daode*. While I earlier explained how chapter 42 relates the emergence of all things from *dao* as the one gives birth to the two, etc., nevertheless, other evidence suggests that this process should not been seen as statically unidirectional. It is just as proper to say that the myriad things give birth to the three, which gives birth to the two, etc. The reason for the viability of this reversible comprehension is that the origin of the world is not seen as one time event, but an on-going process contained in each moment.

An example of this atemporal structure is seen in *Daodejing* chapter 40, in which it is explained that,

> "Returning" is how way-making (*dao*) moves,
> And "weakening" is how it functions.
> The events of the world arise from the determinate (*you* [有]),
> And the determinate arises from the indeterminate (*wu* [無]).[62]

The last two lines might also be rendered as, "Things under heaven arise from what there is, and what there is arises from what there is not." Thus, neither being nor non-being is primordial in the working out of *dao*, and neither rule governed determinism nor chaotic indeterminism is more basic to its nature. Rather, the *dao* works by undermining such oppositions, erasing their incommensurability in its "returning" and "weakening."

A concrete illustration of this reversibility can be seen in *The Great One Gives Birth to the Waters*, a text that was apparently once a part of the *Daodejing*, only to be lost until it was found again in an archeological dig in 1993.[63] *The Great One* begins by explaining how the Great One gave birth to the heavens in collaboration with the waters, which in turn leads to other births culminating in the creation of the yearly cycle. Having

61. Hall and Ames (2003), 173.
62. *Ibid*. 40.
63. *Ibid*. 225.

explained all this, the text then traces its way backwards from the yearly cycle back up to the Great One.[64] The purpose of this reversal is to show that while there is a unidirectional order of birth, with the Great One at the top and the yearly cycle at the bottom, there is not a unidirectional order of priority. While the year cannot be said to give birth to the Great One, nevertheless, the Great One would not be the Great One without the existence of the yearly cycle. Thus, the Great One does not transcend its descendants, nor is it transcended by them.

If *daode* is similarly non-transcendent, then it cannot be perfectly simulated from without, because any external simulation would need to rely on the existence of complete structural similarities between the calculating system and *daode* in order to create a system of reliable inferences between the two systems, and such complete mirroring of structure cannot exist if mirror needs to also mirror itself. Thus, the mysteriousness of *daode* arises not from its distance from human life, but its closeness to it. As chapter 38 of the *Daodejing* states, "'Foreknowledge' is tinsel decorating the way [*dao* 道],/ And is the first sign of ignorance."[65] That is, thinking that we know the way before we actually do is the first step on a false path. Truly grasping the way means being aware of its ineffability and resolutely able to accept whatever the way may bring.

Thus, the first major blow to the computational interpretation of *daode* is that, unlike a computer system, while *daode* is rule governed, it neither transcends its effects nor is it transcended by them.

B. Difficulties accounting for the role of *de*

One reason that we first had such facility at analogizing *daode* to computation, only to later encounter difficulties with computational transcendence is that so far our exposition has accounted for *dao* to the near exclusion of *de*. Loosely, one can say that if *dao* is the "potential" of the system, then *de* is "actuality" of it. (Of course, this is only loose

64. *Ibid*. 229 – 230.
65. *Ibid*. 136.

talk, since both *dao* and *de* contain measures of each.) In the focus-field model of Hall and Ames, *de* is the focus in which the "insistent particularity" of the whole comes to the fore. *De* are those concrete realities which keep *daode* from becoming a swamp of abstract generalities, but at the same time, *de* and *dao* are structured in such a way that each is contained within the other. Within the *Daodejing*, *de* is repeatedly stressed as the key to proper understanding of how to make one's way in the world. For example, chapter 65 explains that in the area of statecraft,

> Those who really know the distinction between using knowledge or a lack of it in governing the state will moreover become its model.
> And those who really know this model are said to be profoundly efficacious (*de* [德])[66]

Thus, *de* is the profound efficacy that comes from knowing one's limitations. Along the same lines, chapter 38 extols the effortless success that comes from understanding particulars:

> It is because the most excellent (*de* [德]) do not strive to excel (*de*)
> That they are of the highest efficacy (*de*).
> And it is because the least excellent do not leave off striving to excel
> That they have no efficacy.
> Persons of the highest efficacy neither do things coercively
> Nor would they have any motivation for doing so.
> … Thus, only when we have lost sight of way making (*dao*) is there excellence[67]

Hence, again we see that true excellence comes from an understanding that does not seek to control or rationalize—a major theme of the *Daodejing* that is difficult to account for in the computational model of *daode*. Time and again, the *Daodejing* points its readers to the importance of what is not (無), the obscure (玄), the mysterious (妙), and the feminine (牝). To understand why these aspects of reality are so emphasized, we must examine of the importance of *de* in *daode*.

The alternative to a Daoist strategy of non-coercively coping with ignorance is one almost invariably taken in the software development field—planning out in advance every eventuality and pre-programming a response for each occasion. For example, one famous problem for conventional, contemporary computers is that

66. *Ibid*. 179.
67. *Ibid*. 135 – 136.

attempting to divide by zero may cause a software crash. For this reason, conscientious developers need to study every part of their code where two numbers may be divided, think about the possible inputs at that stage of the code, devise an appropriate response to a request to divide by zero, and insert a test that ensures that an appropriate response is carried out in the event of a zero division. Of course, division by zero is only one of the best known of a large number of potentially destabilizing software operations. There are innumerable others, such as buffer overflows, improperly terminated strings, memory leaks, and so on. Worse still, even minor hardware malfunctions, such as a bit being changed by cosmic radiation, may cause a cascade of further errors if the malfunction occurs in a critical juncture. In order to avoid these problems, the developer must use their foresight at every step of the way to identify any and all potentially problematic areas and write an appropriate response. Unfortunately, finding these "corner cases" through routine planning or testing can be very difficult, since it is precisely the unanticipated nature of these bugs which makes them so dangerous. Of course, it is easy to make a computer program that has a response to all situations. One could, for example, program one's computer to always resume operation after a division by zero as though no request to divide had been made. Unfortunately, such a response would not be appropriate in the majority of cases and would lead to undesired results, which is the reason that it is not already the procedure followed in the event of a division by zero. While it is easy to have a response for any situation, it is much more difficult to have an appropriate response for all situations. The measurement of appropriateness of response is too fine a calculation to make in general on the basis of anything less than everything all together.

Computation is not the only field where the difficulty of crafting appropriate response rears its head. To give just one more brief example, in the field of law, it is normal for a law to be written only after an offense has occurred once and it is recognized as something to be prevented. Since even simple laws may be worded in such a

way as to lead to unexpected consequences, the law must be revised periodically in order to deal with unforeseen situations that arise. The difficulty for both the computer software and the law is that programmatic rules exist to smooth over differences between situations and create a uniform protocol of response. Unfortunately, in the real world, there are too many unexpected occurrences and implications to create an abstraction that can smooth over all distinctions without also losing fine details. These lost details have a habit of accumulating over time and leading to crises. The Daoist emphasis on the *de* of *daode* is precisely a reminder of the dangers and opportunities that arise out of the accumulation of lost details. For the Daoist sage, these unexpected *de* occurrences are not a problem to be either avoided or mastered in advance through careful pre-programming of responses, but rather, it is the emergence of these occurrences which is to be encouraged and coped with as they arise. They are the source of novelty and originality in *daode* and what separate the spontaneous yet appropriate flow of *daode* from the inexorable grinding of a single, mechanistic computation on the one hand and the sheer randomness of cosmic dice on the other.

Among the qualities lost when one focuses on *dao* to the exclusion of *de* is human creativity. The point of mastery of *dao* is the expansion of human creativity to such a degree that even formidable, unforeseeable situations can be easily coped with. Pablo Picasso is credited (most likely erroneously) with the aphorism, "Computers are useless. They can only give you answers." This precisely captures the difference between Daoist thinking and computational thinking. In order to treat a system as a reliable means of correlating with another system, it is necessary to abstract away the irregularities of each system in order to emphasize the reliability of their connection. In order to understand the systems correlated, it is first necessary to remove extraneous details. However, the *Daodejing* repeatedly warns of us of the dangers that come from this willful ignorance of the seemingly insignificant. As chapter 71 informs us, "Knowing that one does not know is knowing at its best."[68] In order to know our own

limitations, it is necessary to emphasize the mysterious and hidden so that we are challenged to ask new questions and create new, uncalculated responses. These responses optimize the situation without seeking to dominate it. They gain awareness without becoming limited by the narrow scope of their self-supposed knowledge.

C. Difficulties with postulating the ontological priority of computation

In addition to the difficulties already listed with supposing that *daode* can be conceived of as computational, there are also reasons to question whether the world itself could be computational. I earlier defined computation as "the process by which … is used to …." Here, the passive voice has elided an important distinction. Put into the active voice, the definition becomes, "the process by which we, conscious observers, use … to …." In other words, what is computation and what is not computation is a matter of interpretation

This in turn creates two possible definitions for "computer." A computer may be anything that *could* be used for computation. In this case, everything physical is a computer, since if nothing else, it could be used to assist in counting by assigning it a number. On the other hand, a computer may be anything that *is* being used for computation. In this case, a new laptop still in the box is not yet a computer, since its potential for use in computation has not yet become actual. Applying this to the cosmos as a whole, we find that either the cosmos must be a computer, no matter what its underlying ontological makeup may be, or the cosmos was not a computer until conscious observers came into being, but now it is. Neither result is particularly appealing or enlightening, and both point to the unresolved difficult first hinted at when I began to define computation. Though it is possible to create a definition of computation that removes the implication of human artifice in the creation of the computer, it is not

68. *Ibid*. 189.

possible to create a meaningful definition of computation that omits the role of the human observer

Regarding the idea that computation can exist as computation independently of its recognition by an observer, Jaron Lanier proposes an interesting thought experiment that shows the danger of accepting this premise along with the computational theory of mind:

> Some people might remember the "rain drops" argument. Sometimes it was a hailstorm, actually. The notion was to start with one of Daniel C. Dennett's thought experiments, where you replace all of your neurons one by one with software components until there are no neurons left to convert. At the end you have a computer program that has your whole brain recorded, and that's supposed to be the equivalent of you. Then, I proposed, why don't we just measure the trajectories of all of the rain drops in a rain storm, using some wonderful laser technology, and fill up a data base until we have as much data as it took to represent your brain. Then, conjure a gargantuan electronics shopping mall that has on hand every possible microprocessor up to some large number of gates. You start searching through them until you find all the chips that happen to accept the rain drop data as a legal running program of one sort or another. Then you go through all the chips which match up with the raindrop data as a program and look at the programs they run until you find one that just happens to be equivalent to the program that was derived from your brain. Have I made the raindrops conscious? That was my counter thought experiment. Both thought experiments relied on absurd excesses of scale. The chip store would be too large to fit in the universe and the brain would have taken a cosmologically long time to break down. The point I was trying to get across was that there's an epistemological problem.[69]

While it is true that not all who examine this thought experiment will see the panpsychism it seems to imply as an unacceptable result, nevertheless, the danger that Lanier's thought experiment appears to demonstrate is that unmoored from an actually operational frame of reference, anything can be made to represent anything else. Thus, we cannot distinguish between computational and non-computational entities on an objective basis. To say something is "a computer" is akin to saying something is "useful." Any object can be useful given the right agent in the right circumstances. Things may be commonly construed by us as "useful objects," but this is merely a shorthand way for us to say that we believe that is more likely that these objects will get a chance to manifest their potential for usefulness than other objects. So too, calling something a

69. Lanier.

computer means merely that we believe it to be more likely than usual that object so described will be used for making truly complicated computations.

If being a computer is not an objective aspect of entities in the world, then the notion that the world itself is at its most basic level a computer must be rejected if our subjectivities are presumed to rest on the objectivity of the outer world. At best, we can say that viewed from a certain perspective, the rules that govern the behavior of sub-atomic particles may be similar to recursively executed computer algorithms, à la Wolfram, but even if that is the case, it would not be strictly correct to call the world a computer, since "being a computer" is something that depends on the construing of an observer. There are then two possible ontologies to work out on this basis. In the first, the fundamental parts of our world are what they are, but they are most easily described by computational formulae. Such an ontology is only different in specifics from the ordinary scientific materialist ontology. In the other possible ontology, the world is a true computation, but unlike other computations, there is no physical basis for the computation which produces the physical world. Rather, a particular transcendent subjective perspective on the world gives rise to our own subjectivities through its iterative development. This ontology seems to be no more than a newly refined sub-category of theism, rather than a radical new ontology as computationalism promised to usher in. In either case, the fact of computationalism is not fundamental, but at one remove from what is most basic. All of this further separates the computationalist perspective from that of Daoism, in which *dao* and *de* are at once fundamental constituents of the cosmos, and at the same time emergent members of it.

V. Conclusion

In the course of this paper, I have shown that while Daoism shares many attributes in common with the computational view, in the end, there remain significant difference between them. Specifically, the computational view of the world fails to account for the

true comprehensiveness and thus emergence of *dao* and the insistent particularity and thus mysteriousness of *de*. Nevertheless, recognizing the shared background of processual analogizing shared between the two outlooks enables us to deepen our understanding of both, and further, it allows us to recognize the degree to which computational thinking pervades our ordinary thought already at this early juncture in the information revolution. Ultimately, however, the computational view is a merely subjective imposition on the world, and, unlike *daode*, it is not capable of sustaining an ontology in itself. By bolstering our view of computation with insights from Daoism, we may see future advancements in our understanding of both.

Bibliography

Abelson, Harold and Gerald Jay Sussman. *Structure and Interpretation of Computer Programs*. MIT Press, 1984. Available at `http://plato.stanford.edu/archives/win2003/entries/aristotle-metaphysics/`. Last accessed April, 2008.

Aristotle. *Metaphysics*. Trans. Hippocrates G. Apostle. Indiana University Press, Bloomington and London, 1966.

Asimov, Isaac. "The Last Question" in *Isaac Asimov: The Complete Stories, Vol. 1*. Broadway, 1990.

Candrakīrti. *A Guide to the Middle Way*. (*Madhyamakavatara*.) Translated by Stephen Batchelor from Geshe Rabten's *Echoes of Voidness*. Wisdom Publications, pp. 82 and 84, 1984.

Chaitin, Gregory. *Meta Math!*. New York: Vintage Books, 2005.

Cohen, S. Marc, "Aristotle's Metaphysics", *The Stanford Encyclopedia of Philosophy (Winter 2003 Edition)*, Edward N. Zalta (ed.), `http://plato.stanford.edu/archives/win2003/entries/aristotle-metaphysics/`. Accessed April, 2008.

Copeland, B. Jack. "The Broad Conception of Computation." *American Behavioral Scientist*, May 1997; vol. 40: pp. 690-716.

Feynman, Richard. "Simulating Physics with Computers" in *Feynman and computation: Exploring the limits of computers*. Ed. Anthony Hey. Perseus Books, Cambridge, MA, 1999: pp 133 – 153.

Fischer, John Martin. "Recent Work on Moral Responsibility." *Ethics*, Vol. 110, No. 1. (Oct., 1999), pp. 93-139.

Graham, A.C. *Disputers of the Dao*. La Salle, IL: Open Court, 1989.

Hall, David L. and Roger T. Ames (2003). *Daodejing, "Making This Life Significant": A Philosophical Translation*. New York: Ballantine Books.

——— (1998). *Thinking from the Han: Self, Truth, and Transcendence in Chinese and Western Culture*. Albany: SUNY.

Heidegger, Martin. "Modern Science, Metaphysics, and Mathematics" from *What is a Thing?* in *Basic Writings*, revised and expanded edition. Trans. David Krell. Harper San Francisco, 1993.

Jullien, François. *The Propensity of Things: Toward A History of Efficacy in China*. Trans. Janet Lloyd. New York: Urzone, 1995.

Kahn, Charles H. *The Art And Thought of Heraclitus*. Cambridge University Press, 1979.

Kant, Immanuel. *Critique of Pure Reason*. Translated by Werner S. Pluhar. Hackett Publishing Co., Indianapolis. Originals 1781, 1787; translation 1996.

Keightley, David. *Sources of Shang History: The Oracle-Bone Inscriptions of Bronze Age China*. Berkley: University of California, 1978.

Lanier, Jaron. *Why Gordian Software Has Convinced Me To Believe In The Reality Of Cats And Apples: A Talk With Jaron Lanier*. Interview with the Edge Foundation Nov. 19, 2003. Available at `http://www.edge.org/3rd_culture/lanier03/lanier_index.html`. Accessed Nov. 18, 2007.

Murti, T. R. V. *The Central Philosophy of Buddhism*. Second edition. Bradford and Dickens, London, 1955, 1960.

Parfit, Derek. "Personal Identity." *The Philosophical Review*, Vol. 80, No. 1 (Jan., 1971), pp. 3-27.

Priest, Graham. *Beyond the Limits of Thought*. Second edition. Oxford University Press, 1995, 2002.

Reichenbach, Hans (1947). *Elements of Symbolic Logic*. MacMillan Co., NY.

——— (1956). Ed. Maria Reichenbach. *The Direction of Time*. University of California Press, Berkley and Los Angeles, 1956.

Rescher, Nicholas. *Nature and Understanding: The Metaphysics and Methods of Science*. Oxford University Press, 2001.

Sainsbury, R. M. *Paradoxes*. Second ed. Cambridge University Press, 1995.

Śantideva. *A Guide to the Bodhisattva Way of Life*. (*Bodhicaryāvatāra*.) Translated by Vesena Wallace and B. Allan Wallace. Snow Lion Publications, Ithica, 1997.

Sider, Theodore. *Four-Dimensionalism: An Ontology of Persistence and Time*. Clarendon Press, Oxford, 2001.

Turing, Alan. "On computable numbers, with an application to the *Entscheidungsproblem*." *Proceedings of the London Mathematical Society*, 1936; S.2, 42: pp 230 – 265.

Van Fraassen, Bas C. and Isabelle Peschard. "Identity Over Time: Objectively, Subjectively." The Philosophical Quarterly, Volume 58, Number 230, January 2008 , pp. 15-35(21).

Wheeler, John A. "Information, Physics, Quantum: The Search for Links" in *Feynman and computation: Exploring the limits of computers*. Ed. Anthony Hey. Perseus Books, Cambridge, MA, 1999: pp 309 – 336.

Williams, Paul. *Studies in the Philosophy of the Bodhicaryāvatāra: Altruism and Reality*. Motilal Banarsidass Publishers, Dehli, 2000.

Wolfram, Stephen. *A New Kind of Science*. Champaign: Wolfram Media, 2002.

Zuse, Konrad. *Calculating Space*. MIT Technical Translation, 1970.