The Cornell Undergraduate Journal of Philosophy

Logos

λόγος

Volume VI • 2010

Logos : λόγος

Volume VI • 2010

An independent student publication

Logos: The Undergraduate Journal of Philosophy, an independent student organization located at Cornell University, produced and is responsible for the content of this publication. This publication was not reviewed or approved by, nor does it necessarily express or reflect the policies or opinions of, Cornell University or its designated representatives.

Mail:
Logos: The Undergraduate Journal of Philosophy
c/o The Sage School of Philosophy
218 Goldwin Smith Hall
Cornell University
Ithaca, NY 14853

Email:
logos-mailbox@cornell.edu

Contents

Editors' Introduction

The staff of *Logos* is proud to present the sixth volume of Cornell University's undergraduate journal of philosophy. After carefully considering the many submissions we have received over the past year, we have selected an eclectic set of five articles, chosen for their creativity, cogency, and depth of philosophical inquiry.

Every year, the *Logos* staff is committed to showcasing some of the finest philosophical work from undergraduate students across the nation and the globe, and our sixth volume is no exception. In "Limits of Matter: An Analysis of the Epistemological and Ontological Status of 'Limit Entities' in Stoic Metaphysics," Tony Mills offers an interesting investigation of the role of limit entities within the Stoic metaphysical system. In his article, "Naturalized Perspectivism: Where Cognitive Science Meets Phenomenology," Brian Wermcrantz explores the embodiment theory of cognition with respect to the work of Descartes, Merleau-Ponty, Wittgenstein, and others, arguing for a phenomenological approach to the theory. Patrick Decker provides an analysis of two methodological approaches to providing an account of 'explanation' in "Giving An Account of Explanation: The Language User Versus the Technical Approach," while "A Case for Stage View Presentism," by Jorgen Hansen, is an original and rigorous examination of several ontological accounts of objects and events. Our final contributor, James Hodson, defends the computationalist theory of mind and cognition in his piece, "Consciousness as the Domain of a Computation."

In addition to the five authors included in the present volume, the *Logos* editors would like to thank all those who submitted their work to our journal. We also extend our gratitude, as always, to our dedicated undergraduate staff, for their hard work and year-long contributions to the production of *Logos*. In closing, we would like to acknowledge our advisor, Professor Andrew Chignell, for his unwavering and invaluable guidance and support.

Ted Hamilton
Ariana Marmora
Editors-in-Chief, *Logos*
Cornell University

Limits of Matter:

An Analysis of the Epistemological and Ontological Status of 'Limit Entities' in Stoic Metaphysics

Tony Mills
Northwestern University

Debates concerning the divisibility of space, time, matter, and motion can be traced back at least to Zeno of Elea, and their elaboration in the atomism of Democritus and its Aristotelian critique. A question underlying these debates concerns the status of the limit: is there a point at which division is no longer possible? If so, there must be a limit, epistemologically or ontologically fundamental, which prevents further division. This question is perhaps most urgent, as a point of divergence among various Ancient Greek schools of thought, in the case of the divisibility of matter. For a given solution to the problem of matter's divisibility – whether in the form of postulating indivisible atoms underlying phenomena, or differentiating between actual and potential infinities – sheds light on that particular philosophical school's metaphysical commitment with regard to the nature of what constitutes the physical world.

Stoic natural philosophy presents an interesting moment in this history, affirming with Aristotle, and against the atomists, the infinite divisibility of matter. It does so, however, without committing to the metaphysical system that permitted the student of Plato to defend this position. Indeed, it seems as though much of Stoic thought is as opposed to the Peripatetic tradition as it is to the atomistic one, and thus the status of "limit entities" – as Michael J. White calls them – presupposed in the Stoic affirmation of infinite divisibility, cannot be understood according to the Aristotelian potentiality/actuality dichotomy. On the contrary, the Stoic position regarding "infinite divisibility" is a result of their commitment to an altogether different dichotomy: between what exists (corporeals) and what does not exist (incorporeals). As we shall see, however, despite the strict nature of this metaphysical schema, the epistemological and ontological status of limit entities remains deeply ambiguous – at least according to the fragmentary accounts we possess which address the topic – within Stoic metaphysics.

In what follows, I will outline the ontological features of the Stoic metaphysical system in the hopes of ascertaining the place of limit entities therein. It will prove illustrative to focus on the Stoic doctrine of "total blending" in relation to their commitment to the infinite divisibility of matter, which will enable us to grasp, or at least to speculate about, the Stoic conception of matter more generally. My concern is neither to reconstruct the textual evidence in favor of one position or another, nor to gather the various arguments according to a principle of systematic coherence, Stoic or other. My intention, on the contrary, is to grasp the locus of an ambiguity in the Stoic system that has important implications for the Stoic system itself, as well as conceptions of matter in general. I will argue, with White, that limits can be understood most coherently within Stoic metaphysics "beyond the corporal/incorporeal" dichotomy, as "geometers' fictions" or "analytical constructs." Against White, I will try to demonstrate how limits nonetheless remain *necessary* both analytically and perceptually without, however, constituting ontological features of the natural world. To begin, it will be helpful to enumerate the broader features of the Stoic metaphysical position.

I. GENERAL FEATURES OF STOIC METAPHYSICS

The Stoics adhere to a rigorous materialism that abandons the Platonic dualism of Form and appearance, as well as the Peripatetic hylomorphism which sought to take its place. Like the atomists, the Stoics affirm that only material, that is to say *corporeal* bodies exist, and that only such an ontology can explain the causal nexus discernible in nature and account for the rationality of the cosmos. As Long and Sedley observe, the Stoics make corporeality "the hallmark of existence."[1] But Stoic philosophy is no less opposed to the atomistic materialism of Democritus and Epicurus, in which limit entities are accorded an irreducible ontological status. According to Democritean atomism, from which Epicurus' scientific account is taken, "the basic units of the world are everlasting, unchanging indivisible units or atoms."[2] It is with these tiny "bodies with sizes and shapes" that "divisibility comes to an end…"[3] As we shall see, the materialism of the Stoics is in many ways broader than that of its predecessors and competitors. It is distinguished by its emphasis on the cohesive character of nature as a totality comprised not of indivisible corpuscles, but of an ontological continuum. The broadness of Stoic ontology is at once identifiable in the conception of being on which it hinges, and in the category of materiality itself.

The Stoics famously divide the natural world into two categories: that which *exists* and that which *does not exist*. These two categories, composed of corporeals and incorporeals respectively, are derived from a prior ontological category of "somethings." Long and Sedley write that, "…it is not with the existent but with the prior notion of 'something' that the Stoic ontological scheme starts."[4] In short, the Stoics abandon the identification of 'reality' with existence by affording non-existents a peculiar, but nonetheless ontological, status. We shall begin, however, with the corporeals. These existents are bodies that are subject to causal interaction. That is to say, corporeals either act upon, or are acted upon by, other corporeals: "…a cause is a body…it is a fundamental Stoic tenet that only bodies can act or be acted upon."[5] Broadly speaking, corporeals can be divided

[1] A.A. Long and D.N. Sedley, *The Hellenistic Philosophers: Volume 1* (Cambridge: Cambridge UP, 1987), 163.
[2] Anthony Kenny, *Ancient Philosophy* (New York: Oxford UP, 2004), 95.
[3] Ibid., 27.
[4] Long and Sedley, 163.
[5] Kenny, 164.

into two types: active and passive.* Active matter, or *pneuma*, is the animating or "active principle...designated 'creative fire'" that pervades everything; it is indestructible and ontologically fundamental.[6] The Stoics identify this active matter with god, even though it does not exist in a realm that is transcendent of finite beings. On the contrary, *pneuma* subtends and encompasses all beings – inanimate objects, plants, rational and non-rational animals – constituting the totality of nature, in which all things participate coherently.

Passive matter, the second type of corporeality, is derivative, or "less ontologically fundamental."[7] That is to say, passive matter is composed of elements – fire, air, earth, and water – and is both finite and formless.** *Pneuma* acts on the passive matter, giving it form. Unlike Peripatetic hylomorphism, however, the "formal cause" (in this case the active principle) is not immaterial – indeed, it is a corporeal – and is therefore not a true Aristotelian cause. On the contrary, the active and the passive, although they are separate types of corporeality, exist simultaneously within the same objects – something that is forbidden in both the Peripatetic and atomistic schemata. Indeed, for the Stoics, *pneuma* is constantly acting on the passive matter, and thus the latter depends on the former for its continued existence. Long and Sedley characterize the relationship of active to passive matter thus:

> According to this scheme any object, or the world as a whole, can be analyzed as a composite of matter and god...Matter is what you would get if (*per impossibile*) you could remove all the characteristics of an object which make it something particular. God accounts for those characteristics...Matter needs god in order to be a particular entity, and god needs matter in order that there shall be some entity for god to characterize...[8]

---

* This division is somewhat heuristic. While it is by no means a misrepresentation of the Stoic position, it oversimplifies it significantly, ignoring the fourfold division of corporeality into substrate, qualified, disposed, relatively disposed – and the subsequent two-part category of "qualified" (common and peculiar). It is beyond the limits of the present analysis to undertake a sufficient explication of these categories, each of which presents considerable textual and philosophic ambiguities. For such an elaboration, see Long and Sedley, 162-79.

[6] Long and Sedley, 163.

[7] Michael J. White, "Stoic Natural Philosophy (Physics and Cosmology)," in *The Cambridge Companion to the Stoics* (Cambridge: CUP, 2003), 133.

** As Michael J. White observes, these elements can be further divided into active and passive: "air and fire are active, earth and water are passive" (White, 135). There is some debate surrounding the status of these elements, specifically fire and air, and their relationship to the active principle, as well as to the Aristotelian theory of elements. *Pneuma* is generally translated as "breath," and is often employed accordingly by Stoic thinkers; but fire appears to occupy the position of animating principle in early Stoic thought, and it continues to play an important role in the doctrine of cosmological conflagration in later Stoic writings. Provisionally at least, *pneuma* can be understood as "fiery breath" constituted by the elements of fire and air. See White for a lengthy discussion of these problems.

[8] Long and Sedley, 271.

Thus, while "god" or the "active principle" and passive matter exist simultaneously in objects, both are corporeal and have a causal relationship. In other words, god acts as the "sustaining," complete, or principle cause of the passive matter – that cause which will "...sustain its effect as long as it itself lasts."[9] Yet, this broad and dynamic conception of materiality – encompassing the active and all-pervasive principle as well as the formless and inert matter, in addition to the material elements that compose them – comprises only one aspect of the Stoic ontological schema. If, as our discussion has demonstrated, causal relationships are constitutive of what it means to "exist," the entities comprising the second ontological category – the incorporeals, or non-existents – subsist outside of the causal nexus.

That is to say, incorporeals are not acted upon, nor do they act upon anything else. But incorporeals are not, for all that, epistemologically or even ontologically irrelevant. On the contrary, incorporeals – which include *lekta* (or sayables), time, place, and void – are constitutive of the reality we inhabit. Long and Sedley describe them as those entities which, "although not bodies...are felt to be an ineliminable part of the objective furniture of the world."[10] In short, for the Stoics, incorporeals are necessary aspects of our reality, requisite for our knowledge about the world and of the causal relationships that govern it, even though incorporeals do not "exist" or enjoy causal efficacy within it. Thus, for example, *lekta* are foundational for Stoic epistemology: they permit us to signify objects of thought, and to understand and evaluate the truth-conditions of that which is signified.[11] *Lekta* also play a central role in causal relationships, functioning as the incorporeal effects generated by the interactions of corporeal causes.[12*]

Meanwhile, time, another incorporeal, is doubtless important for certain forms of causation – specifically, causal succession – but it is also the law of our finitude, as well as a central aspect of the cosmic cycle of eternal recurrence.[**] Place is "what is occupied through and through by an existent," and therefore

[9] Ibid., 341.

[10] Long and Sedley, 164.

[11] Sextus Empiricus, *The Stoics Reader*, ed. Brad Inwood and L. P. Gerson (Cambridge and Indianapolis: Hackett, 2008), 89-90.

[12] Ibid., 90-91.

[*] It should be noted that it was in logic — a category in which the Stoics included what we call epistemology — and aetiology that the Stoics offered some of their greatest insights. Indeed, modern logic owes much to Stoic advancements in the field. It was Chrysippus, for example, who invented the *modus tollens* and the *modus ponens*. Moreover, while the Stoics abandon the Peripatetic conception of causality, they develop a sophisticated theory of causation in which a given cause can be understood only in relation to a web or nexus of causes. In this sense as well, one could hardly underestimate the importance of *lekta* for Stoic philosophy.

[**] In addition, various Stoic philosophers, including Chrysippus and Posidonius, develop elaborate theories of time. Of particular interest, for example, is the apparently paradoxical Stoic belief, reported by Plutarch, that there "neither remains nor is left in the 'now' any part of present time." Beyond being a "proper subject of thought and discourse," time — like *lekta*, although to a lesser extent — seems to be a rather important topic in the Stoic metaphysical tradition (Long and Sedley). See Inwood and Gerson, 87-88.

permits us to have an ordered understanding of objects in the world, and to map out their movements and relationships.[13] Void is perhaps less fundamental for Stoic epistemology than *lekta*, time, and place; it is, however, important ontologically, as the "infinite…outside the world," or beyond the cosmos.[14*] The striking importance of a category understood to incorporate things that do not exist is one of the salient features of Stoic metaphysics, and it is illustrative of the inclusiveness of their "materialism." Indeed, one might say that incorporeals have an irreducible epistemological and ontological status that in some cases is equal to or even surpasses the status of corporeals (especially the passive matter) in furnishing us with knowledge of the world. To use an anachronistic phrase, one might say that incorporeals are necessary as the conditions for the possibility of experience. In any case, a prerequisite for grasping the breadth of Stoic materialism is recognizing the ontological import of the category of non-existents.

II. THE NATURE OF LIMIT ENTITIES IN TWO STOIC DOCTRINES: "TOTAL BLENDING" AND THE INFINITE DIVISIBILITY OF MATTER

What, then, are limits? That is to say, how do the Stoics understand them according to their metaphysical schema? Plutarch reports that, according to Chrysippus, "…the limit is not a body…"[15] In addition to accounts such as this, it is fairly certain that for the Stoics limit entities are not corporeal. The contrary belief would contradict a large body of their doctrine (including "total blending" and infinite divisibility) for which there is ample secondary and primary textual support. But while the Stoics hold that limit entities are not corporeal, it is not immediately clear – and indeed, as we shall see, it is perhaps fundamentally unclear – whether or not they can be adequately construed as incorporeals. If limits are to be included in the category of "non-existents," they must be derived from the prior ontological category of "somethings." That is to say, while they would subsist outside of the causal nexus, they would nonetheless enjoy an ontological status on par with time, place, void and *lekta*. If, on the other hand, limit entities are understood as "…purely mental constructs [that] fall… outside the corporeal-incorporeal dichotomy," as more than one commentator has suggested, they would be denied the ontological privilege of constituting

[13] Long and Sedley, 294.
[14] Ibid.
* It is true that the infinite nature of the void is a matter of debate among the Stoics. According to Aëtius, for example, Posidonius believed that the void was "outside the cosmos if not infinite…" On the other hand, Chrysippus offers a proof for the unlimited nature of the void. See Stobaeus *Anthology*, in Inwood and Gerson, *The Stoics Reader*.
[15] Long and Sedley, 299.

the physical dimensions of our world.[16*] Indeed, they would, to use White's terminology, yield "a degree of falsification or misrepresentation of the way the physical world really is."[17] In order to address this ambiguity, we must examine more carefully the Stoic conception of matter. To begin, we turn to their doctrine of "total blending."

In explaining this doctrine, it will become clear how for the Stoics two bodies can occupy the same space without paradox. Alexander of Aphrodisias – who, along with Plutarch, criticizes the Stoic doctrine of mixture – reports in his *De mixtione* that Chrysippus distinguishes three types of mixture. The first, as White describes it:

> ...is simply the juxtaposition (*parthesis*) of sizeable 'chunks' of different stuffs... which Chrysippus says occurs by 'fitting together' or juncture (*hamrê*) and in which each constituent preserves its proper nature and quality.[18]

This kind of mixture is perfectly conceivable within atomistic physics – indeed, as White notes, this is the definition of corpuscular mixture or change.[19] The second type of mixture consists in:

> total combination (*sunkrisis di' holôn*), in which the constituent substances and the qualities they contain are altogether destroyed in order to produce something qualitatively distinct from the constituents.[20]

Alexander does not dwell on the latter type of mixture, but White glosses it as a "total replacement, in a given material substratum, of one set of qualitative determinations by another..."[21*] Finally, the third and most controversial type of mixture is "total blending (*krasis di' holôn*) or blending proper."[22] This is what Alexander describes as the:

---

[16] Long and Sedley, 301.

* Cf. Michael J. White, *The Continuous and the Discrete* (Oxford UP, 1992), and Jacques Brunschwig, "Stoic Metaphysics", *The Cambridge Companion to the Stoics*, 206. White and Bunschwig (among others) observe that there is textual evidence supporting both positions. There is, however, much debate about which of the various conflicting accounts can be trusted, but my concern in the present analysis is more "philosophical" than historical. Hence, I am more interested in the theoretical coherence of each position within Stoic ontology, than in the veracity of specific accounts of Stoic doctrine.

[17] Michael J. White, *The Continuous and the Discrete* (New York: Oxford UP, 1992), 286.

[18] White, "Stoic Natural Philosophy (Physics and Cosmology)," 147.

[19] Ibid.

[20] Ibid.

[21] Ibid.

* See also Robert B. Todd, *Alexander of Aphrodisias on Stoic Physics* (Leiden: Brill, 1976), especially pages 49-65.

[22] White, "Stoic Natural Philosophy (Physics and Cosmology)," 147.

> ...interpenetration of two or more bodies in such a way that each preserves its own proper nature and own qualities in the mixture...the commingled bodies go through one another (*chôrountôn di' allêlon*) in such a way that there is no part of them that does not partake of everything in such a blended mixture. If this were not the case, then the result would no longer be a blending (*krasin*) but rather a juxtaposition (*parathesin*).[23]

Unlike the first two forms of mixture recounted by Alexander, total blending requires an absolute mixture, in which the properties of each substance remain distinct from that of others, while occupying the same space in such a way that the two substances can be separated out again.[24] The apparently paradoxical nature of the Stoic doctrine of total blending – that is to say, its disregard for some of the fundamental ontological claims of more widely accepted natural philosophies – invited significant criticism, and in some cases, outright mockery and dismissal from contemporaneous and later commentators.

The distinguishing feature of Stoic ontology is the conception of matter that it employs and presupposes. We have already encountered the phenomenon of total blending in our explication of the simultaneous cause that relates the active principle to passive matter. Given the importance of this relationship in Stoic philosophy in general, the doctrine according to which two bodies can simultaneously occupy the same place cannot be dismissed as incoherent sophistry. For the Stoics, and apparently for Chrysippus in particular, total blending is not only a central doctrine, but a rather unexceptional phenomenon. Indeed, in his *Physics*, Chrysippus is reported – by Diogenes Laertius, among others – to have provided a fairly banal example:

> And blends occur through and through [totally]...and not by surface contact and juxtaposition; for a small amount of wine thrown into the sea will be extended [through it] to a certain degree and then blended with it.[25]

It is immediately clear that to maintain this doctrine, the Stoics must abandon important features of the atomistic understanding of matter.

Indeed, as Long and Sedley observe, the conception of matter employed

[23] White, "Stoic Natural Philosophy (Physics and Cosmology)," 147.
[24] Long and Sedley, 293.
[25] Robert B. Todd, *Alexander of Aphrodisias on Stoic Physics* (Leiden: Brill, 1976), 31.

by the Stoics is precisely not a corpuscularian one. Instead, it is a conception according to which the two things occupying the same place are not "two determinate and independently existing bodies, but...two bodily functions (breath and matter) which jointly constitute every determinate and independently existing body."[26] In other words, the Stoic doctrine of total blending requires that the limit or surface of a given body be not a determining feature of that body; on the contrary, it is clear that for the Stoics such a limit could not even be said to exist in any physical sense. For it is precisely in lacking such limits or surfaces that two bodies can be *totally* blended, and not merely juxtaposed.

In short, the blending in question does not have a spatial character – that is to say, the two substances are not spatially juxtaposed, but *qualitatively* distinguished from one another even while they occupy the same spatial area. Thus, the "limits" of these bodies must not possess corporeal existence. The Stoic conception of matter can thus be said to be explicitly "anti-corpuscularian" or "anti-atomistic." But while the Stoics and Aristotle may in this sense share a common enemy, the doctrine of total blending would be no more acceptable according to the principles of Aristotelian physics. Indeed, it is precisely in order to avoid such an absurdity that Aristotle, while rejecting Plato's Forms, nonetheless maintains that the formal cause of a body is not itself bodily.[*]

In any case, it is clear that if Stoic doctrine is to make any sense, limit entities cannot enjoy a corporeal status. Indeed, we have seen that it is precisely the incorporeality of limits (or surfaces) that permits two bodies to occupy the same place, and to function not as independent and indivisible corpuscles – which possess physical limits or surfaces – but as relatively indeterminate functions. But can limit entities be said to belong to the category of incorporeals? Before attempting to answer this question, we must turn to a second doctrine of Stoic ontology.

With Aristotle, the Stoics argue against another feature of the corpuscularian philosophy, namely, the indivisibility of matter.[**] Indeed, according to many sources, including Diogenes Laertius, Chrysippus affirms infinite divisibility: "(1) Division is to infinity, (2) or 'infinite' according to Chrysippus (for there is not some infinity which the division reaches, it is just unceasing)."[27] Aristotle maintains

[26] Long and Sedley, 294.

[*] Hylomorphism is a complicated doctrine to say the least, but it is in no way necessary to broach this topic in any detail to recognize that the Stoics do not share it with Aristotle. Cf., Aristotle's *Physics* as well as his *De Anima*.

[**] Of course, the Stoics affirm the divisibility of motion, space, and time in addition to matter. A more exhaustive analysis of the status of limit entities would doubtless require an exploration of the divisibility of these as well. An analysis of that sort, with a particular emphasis on motion and time, would be crucial to understanding the Stoic view of matter in terms of a "field of force" – as Sambursky has suggested – or a mathematical function. While discussions of the concept of time abound in Stoic scholarship, the relationship of time to infinite divisibility and, more specifically, the Stoic affirmation of motion *kata athroun*, is a topic that seems somewhat neglected.

[27] Long and Sedley, 298.

this position by appeal to his metaphysics of potentialities and actualities.† It is unnecessary to recount Aristotle's argument for potential indivisibility in order to recognize that the Stoics do not appeal to such a metaphysic in their own affirmation. On the one hand, we can see that while Stoic natural philosophy shares some features of the Peripatetic school (e.g., the affirmation of infinite divisibility), one of its central ontological doctrines – total blending – is explicitly opposed to it.*

Once again, the coherence of the Stoic position of the infinite divisibility of matter depends on the assertion that, as Proclus reports, "limits of this sort – ...[limits] of bodies – subsist in mere thought."[28] For if limits or surfaces of bodies *existed*, there would be a point at which division becomes impossible (without an appeal to an Aristotelian notion of "potential" divisibility). Hence it is clear that the Stoics hold limits – as in the example of total blending – to *subsist*, and not to *exist*. Yet, given the Stoic category of incorporeals, the ontological status of the subsistence of limits is not immediately obvious. The question remains, in other words, whether or not limit entities – in this case, limits of bodies – should be understood as incorporeals subsisting in, or mere constructs of, thought.

Proclus' use of the phrase, "subsisting in mere thought" might imply that limits do not simply lack existence, but ontological status *tout court*. This inference is complicated, however, by the fact that the Stoics accord incorporeals the same mode of being, namely, subsistence. Indeed, if it is the case, as Long and Sedley propose, that "to be something" is for the Stoics "to be a proper subject of thought and discourse," then limits could very easily be included in the category of "somethings."[29] Yet, as Long and Sedley note, while limit entities,

> ...are often assumed to be incorporeals...they are not listed as such in the sources...and it may be more correct to classify them...as *neither* corporeal nor incorporeal...If something is a purely mental construct – an invention or idealization – the question of its corporeality or incorporeality might be held [by the Stoics] not to arise.[30]

Long and Sedley dismiss the rare textual evidence to the contrary as "too polemical to carry much weight."[31] But if the categorization of limit entities is to be undertaken on the basis of something other than textual speculation,

---

† Cf., Aristotle's *Physics*.

* Cf. Long and Sedley, 301, for a discussion of the Stoic championship of infinite indivisibility – and the continuum in general – in relation to Aristotle's position in his *Physics*.

28 White, *The Continuous and the Discrete*, 285.

29 Long and Sedley, 164.

30 Ibid., 165.

31 Ibid., 165.

there must be something specific about their mode of being that precludes their relegation to the "highest genus" of "something."

III. LIMIT ENTITITES BEYOND THE EXISTENT/NON-EXISTENT DICHOTOMY

While it cannot be said definitively, based on our examples – which we have limited to the case of matter – it certainly *seems* as though limit entities cannot possess the ontological status enjoyed by *lekta*, time, place, and void. For if matter is to be understood as capable of both total blending – a phenomenon in which bodies are perhaps best characterized as functions rather than spatially distinct objects – and infinite division, the limit cannot enjoy an ontological status in any physical sense. Moreover, limits appear to occupy a very minimal place in the Stoic picture of the universe – and in Stoic metaphysics in general – save for their epistemological value in mathematics (specifically geometry).

Whereas nearly every incorporeal enjoys a central role in some Stoic doctrine or other, limit entities must be abandoned in order to make sense of two important doctrines of Stoic ontology, namely, total blending and infinite divisibility.* In other words, like "fictional entities," limits do not appear to occupy space, according to Stoic metaphysics, in any substantive sense. Pursuing the thesis that limit entities are mere "geometers' fictions" – as White does – leads to some very interesting speculations about the nature of matter. White writes:

> Suppose that we regard all surfaces, edges, points, etc. as elements of a mental grid imposed on spatial extension and time. What do we have when we remove the grid? According to the picture I am considering we have infinitely divisible spatial magnitudes. But an infinite sequence of such nested intervals (a monotonic non-increasing sequence) does not converge to some limit entity or boundary – a surface line, or point. Rather, convergence is to some interval of indeterminacy or fuzziness.[32]

While White goes on to develop this position rigorously using fuzzy set theory, we need not recount these technical aspects of his speculations. It suffices to recognize that understanding limit entities beyond the "corporeal-incorporeal dichotomy" yields a radical view of matter that shares little to nothing with atomistic or Aristotelian ontology.

* It is true, as we have already noted, that void is less fundamental for Stoic epistemology than *lekta*, time, and place. Void is important ontologically, however, and can thus be said to subsist in a rather strong sense for the Stoics – even if it is precisely non-existent.

[32] White, *The Continuous and the Discrete*, 287.

According to such a view, while the "interval of indeterminacy" characteristic of relations between bodies is "theoretically divisible *ad infinitum*," in keeping with the Stoic position, "there is a sense in which it represents an ontologically foundational level of physical reality." That is to say, the "...indeterminacy or fuzziness is not merely an *epistemic* feature of our interaction with the world."[33] The "fundamental" nature of matter, on this radicalized Stoic view, is indeterminacy. Such an ontology results in "...a sort of...'pre-analytical continuity' of *to holon*" in which "one physical object is so topologically interconnected with its environment that there are no joints, so to speak, between them; they insensibly blend into one another..."[34] Thus, construing limits as fictions or analytical constructs in the Stoic schema reveals an ontology that is by no means foreign to traditional interpretations of Stoic metaphysics, but nonetheless radical in its affirmation of a "pre-analytic continuity." For White, this conception of matter has appeal in itself, beyond even fidelity to Stoic metaphysics. But he seems to allow only two possible interpretations of Stoic ontology with regard to limit-entities. There is, on the one hand, the view that limit entities are mental constructs that are nonetheless "essential to our understanding of the physical reality and...which cannot correctly be said to misrepresent, falsify, or distort that physical reality."[35] On the other hand, there is the interpretation we have been following, according to which such limit entities "are geometers' fictions and, hence, involve a degree of *mis*representation."[36]

But I would like to propose an intermediary position, according to which limits have no ontological status, but have an important and *unavoidable* epistemic one. On this view, limits, construed as idealizations of the natural world, would be necessary for various kinds of knowledge (e.g., geometry), and even our perceptions, without themselves being ontological features of matter. This position maintains a conception of matter in which indeterminacy, and not corpuscularity, is "ontologically fundamental," while preserving the usefulness of the limit as a mathematical necessity, and an apparently commonsensical notion. Indeed, according to commonsense experience, objects surely appear to possess spatial limits or surfaces.

To pursue even this more moderate speculation, however, it would be necessary to provide a rigorous and in-depth analysis of the status of place, and especially motion and time, in relation to the Stoic affirmation of the incorporeality of limit entities. Moreover, recognizing limits as more or less universal, both epistemically and perceptually, would require revisiting the Stoic

---

33 White, *The Continuous and the Discrete*, 287.
34 Ibid., 324-25.
35 Ibid., 286.
36 Ibid.

doctrine of universal or natural conceptions. For if the ontological character of matter contradicts our commonsense experience, it is unclear how the Stoics, who find "consensus" to be indicative of naturalness, could reject even their corporeality.* Thus what remains for White a neglected philosophical question – how can limit entities be understood as pure constructs when they appear rather unavoidable in everyday experience? – becomes a textual one as soon as limit entities are accorded the status of being commonsensical, or in the Stoics' own vocabulary, "universal by consensus." That is to say, it would be necessary to demonstrate how the Stoics can coherently affirm the non-ontological and constructed nature of limits, while recognizing their commonsensical, and even "universal" status.

Such a project, doubtless speculative in nature, falls well beyond the purview of the present analysis, which has attempted merely to elucidate the basis upon which it might be developed. One could say that while the ambiguity regarding the status of limit entities in Stoic ontology is in some sense irresolvable – at least given the available sources – it is a productive ambiguity insofar as it points to a possible understanding of matter that avoids many of the problems engendered by ancient natural philosophies. Furthermore, this ambiguity offers an alternative to both Aristotelian and atomistic physics – and an attractive one in view of contemporary physics.

* In this sense, therefore, it is hardly our speculation that produces – even if it highlights – this apparent contradiction regarding spatial limits and consensus. Indeed, we have already seen that for the Stoics, even if limit entities could be construed as "somethings," they would nonetheless remain without corporeal being. For a discussion of natural notions, cf., Long and Sedley. See also John Sellars, *Stoicism* (Berkeley and Los Angeles: UC Press, 2006), 76, for an argument against innateness in Stoic epistemology.

BIBLIOGRAPHY

Aristotle. *The Basic Works of Aristotle*. New York: Modern Library, 2001.

Gerson, L. P., and Brad Inwood. *Hellenistic Philosophy*. Cambridge and Indianapolis: Hackett Publishing, 1997.

Gerson, L.P., and Brad Inwood. *The Stoics Reader*. Cambridge and Indianapolis: Hackett Publishing, 2008.

Kenny, Sir Anthony. *Ancient Philosophy*. New York: Oxford University Press, 2004.

Long, A.A., and D.N. Sedley, *The Hellenistic Philosophers: Volume 1*. Cambridge: Cambridge University Press, 1987.

Sellars, John. *Stoicism*. Berkeley and Los Angeles: California University Press, 2006.

Todd, Robert B. *Alexander of Aphrodisias on Stoic Physics*. Leiden: Brill, 1976.

White, Michael J. "Stoic Natural Philosophy (Physics and Cosmology)." *In The Cambridge Companion to the Stoics*. Edited by Brad Inwood. Cambridge: Cambridge University Press, 2003.

White, Michael J. *The Continuous and the Discrete*. New York: Oxford University Press, 1992.

Naturalized Perspectivism:

Where Cognitive Science Meets Phenomenology

Brian Wermcrantz
Grinnell College

This paper draws together interdisciplinary support for a recently popular trend in cognitive science known as embodiment theory. In addition, it forwards an argument for a phenomenological approach to embodiment theory, a hypothesis I call "naturalized perspectivism." I critically investigate embodiment theory as it pertains to Cartesian dualism, behavioral psychology, artificial intelligence, Merleau-Pontian phenomenology, a Wittgensteinian critique of rule-following and sensation language, and, finally, contemporary experimental neuroscience. My analysis argues that embodied cognitive science should adopt the stance of naturalized perspectivism, which understands the perceiving subject as essentially embodied and extended into its immediate environment, and thus calls for the development of a more holistic methodology for the scientific study of cognition.

As opposed to a representationalist approach to cognition, the theory of naturalized perspectivism rejects the conception of the perceiving subject as an organism that acts by following internalized rules and perceives using mental representations of an objective, external world. Instead, by taking this stance towards embodiment theory, I posit that human organisms have evolved to experience the "natural world" in similar, objective ways only insofar as we share similar bodies and particular evolutionary-friendly embodied goals. That is, during the process of perception and learning, an organism's cognitive system cuts across brain-body-world divisions by constantly reconstituting the phenomenologically-experienced world as it offloads its dispositions into sensory objects for its own evolutionary purposes. To ward off charges of relativism, I argue that naturalized perspectivism is fully grounded in biology and evolution insofar as it understands the body and the similarities between bodies to be the fundamental source of the behavior, communication, and development of an organism's neurological system.

To introduce embodiment theory, it is useful to explain how a proponent of the view would respond to three popular, yet problematic theories of the past, the first being Cartesian dualism. Descartes divides the world into two distinct kinds of metaphysical substances, minds and bodies, which have categorically different essences, thought and extension, respectively. Descartes recognizes, however, that mind and body appear to be causally connected; when I think in my mind, "raise my arm," my body begins to move. One way Descartes attempts to explain mind-body interaction is with his representational theory of perception, which holds that the mind makes internal *mental representations* of the external world. To support this representational model of perception, Descartes needs to address the question concerning precisely how the mind and body, as metaphysically separated, interact to form these mental representations.

In an attempt to solve this problem, Descartes argues for the existence of

God, who would ensure that external objects are represented in the mind in a non-deceptive, non-transformative manner. He claims that this process occurs in the brain's pineal gland, a place where the mind and body somehow interact. Operating under the dualist conception of a "thinking self" faced with the task of establishing the existence and authenticity of "external, bodily objects," Descartes is forced to invoke a non-deceiving God who guarantees that our minds (when functioning properly) accurately represent bodies. Such a God thereby warrants the legitimate causal connection between mental representations and extended substance. Thus, unless invoking an omniscient God can be considered a sufficient explanation, Cartesian dualism cannot explain in concrete terms how the mind, as thinking substance, and the body, as an extended "space occupier," actually interact to form mental representations. Descartes' dilemma faces any representationalist view of cognition, such as emergence theories and other contemporary forms of dualism.

Embodiment theory avoids the Cartesian problem of mind-body interaction by understanding the mind as extended or essentially "embodied" in its immediate environment. From the start, embodiment theorists examine the mind not as composed of metaphysically distinct thoughts, but instead as the mental operations consisting entirely of bodies affording particular actions in particular situations. As opposed to Cartesianism, embodiment theory holds that knowledge of the "external world" cannot be warranted exclusively by God. If that were the case, then *subject*-sensitive influences, such as action goals and embodied skills, could not affect perception, when in fact they do in a very important way.

As a demonstration of this fact, Andy Clark provides an insightful example of how a baby learns about slopes. He cites a longitudinal study that investigates an infant's knowledge regarding maneuvering about inclines during its transition from crawling to walking. Understandably, as the crawlers increase in experience, they gradually begin to learn how to maneuver about or avoid the steeper slopes. Remarkably, even after receiving extensive training crawling up an incline, once infants learn how to *walk* they must entirely re-learn information about steep slopes.[1] Essentially, the first time infants walk up the very same slope they once so cleverly navigated on all fours, they demonstrate a lack of information as if they were exploring the area for the first time. Thus, it appears that knowledge and perception are action-specific on a foundational level. If stored mental representations were the medium for knowledge and perception, then the infants would be able to use at least a portion of what they learned from crawling when they started to walk. That is, an infant would

[1] Andy Clark, *Being There: Putting Brain, Body and World Together Again* (London: MIT Press, 1997), 37.

be internally representing features of the slope *itself*, independent of its current embodied projects in the "external world." Instead, the subject learns about slopes only by virtue of how he or she can respond in a given situation, or, as Clark puts it, "how slopes figure in specific contexts involving action."[2] There is no internalized bank of knowledge from which we act.

Whereas Cartesianism holds that we acquire knowledge of, and perform actions in, the world of bodily substance via constant recourse to mental representations, embodiment theory recognizes that our bodily engagement with our current environment shapes our knowledge and perception to begin with. Accordingly, our sensory perceptions consist exclusively of action-laden content, because, from the start, they depend on how our bodies are currently behaving towards particular sensory objects in a way that affords skillful, coordinated action towards them. Thus, without any mention of mental representations, embodiment theory recognizes that mental knowledge and "external" bodily features of the world interact coextensively.

A viable model for the specific mechanism underlying this embodied approach to cognition and knowledge acquisition can be found in the "reflex circuit" proposed by John Dewey. As an early proponent of embodiment theory, Dewey explains how an embodied approach addresses shortcomings associated with behavioral psychology, a second problematic theory of perception. Dewey describes the stimulus-response model of behaviorism as "sensation-followed-by-idea-followed-by-movement."[3] This is to say that, according to the behaviorist, stimuli are first given as self-contained objects in the environment, which are then passively collected by sensory organs, and then responded to according to learned stimulus-response patterns. As the basis of this model, behaviorism holds that only observable bodily action (i.e., not mental states) should be the focus of objective, scientific inquiry. But, if sensory objects are self-contained, external objects, and if we do not have mental states, how then does the perceiving subject determine relevant similarities between their current sensory experience and their past stimulus-response pairings? How are the stimulus-response histories stored without memories (i.e., mental states)? Dewey avoids this problem by replacing the linear, mechanistic stimulus-response model with a multi-directional circuit.

According to Dewey, at the "first" stage of perception, *sensation*, behaviorism falsely posits that stimulus and response operate as independent, self-constituted entities. Alternatively, Dewey recognizes that stimulus and response are members of one fluid, coordinated exercise, what he calls the "reflex circuit." To explain this model, Dewey discusses the famous example

[2] Clark, 37.

[3] John Dewey, "The Reflex Arc Concept in Psychology," *Psychological Review* 3:4 (1896): 358.

of the child and the flame. While a child reaches for a flame, the act of seeing constantly interacts with the act of reaching. According to Dewey, in an account of action, stimulus and response have determinate content only insofar as they coordinate with each other to carry out a unified function. Dewey describes this when he writes, "[i]f the light did not inhibit as well as excite the reaching, the latter would be purely indeterminate, it would be for anything or nothing, not for the particular object seen."[4] Here, Dewey makes clear that for the flame to be a stimulus – for it to enter the reflex circuit as the beginning of an action – it must *already* be engaged with the motor system. In this sense, stimulus and response feed into each other during the process of both perception and learning.

Clearly, then, sensory stimulus and motor response cannot be distinct, self-contained entities, because, if that were the case, stimuli would be inconsequential for the respective motor action. Actually, during the act of reaching, the multi-directional feedback dynamics between eye, hand, body, and world guide the motor response and, at the same time, the processing of sensory stimuli. By recognizing that the feedback loop of the sensory-motor system *itself* guides action, without the need for constant recourse to mental representations, embodiment theory endorses what Clark calls "soft assembly," where a multitude of local interactions between motor and sensory systems leads to emergent features such as coordinated movement. In this way, embodiment theory provides a much more efficient and adaptable stimulus-response model: as Clark articulates, "multi-factor, decentralized approaches [...] yield robust, contextual adaptation as a cost-free side effect."[5] Stimulus-response dynamics constantly guide and adjust the organism's processing of sensory stimuli – a readily adaptable and highly efficient model that does not require mental states and can better account for immediate learning, unlike the behaviorist model.

Dewey also challenges the behaviorist model in terms of the "second" stage of action, *the response*. When the child is burned, he genuinely experiences the response sensation only by virtue of the previous eye-arm-hand coordination, *not* as an entirely new experience. Learning occurs "only because the heat-pain quale enters into the same circuit of experience with the optical-ocular and muscular quales."[6] As the response occurs, the stimulus is reconstituted to be "seeing-of-a-light-that-means-pain-when-contact occurs."[7] Behaviorists would claim that the response should be treated as a different experience, arguing that the continual pairing of two distinct experiences (e.g., touching and burning)

[4] Dewey, 358.
[5] Clark, 43.
[6] Dewey, 358.
[7] Ibid., 359.

would mediate learning and response via classical conditioning.

However, Dewey recognizes that, because stimulus and response are part of one fluid exercise, real learning (i.e., learning that has consequences for future action) occurs precisely at the point when the sensory stimuli are reconstituted and reorganized within the overall sensory-motor system. Hence, for a *response* to have real consequences for learning, it must be integrated with the stimulus: "the so-called response is not merely to the stimulus; it is, so to speak, *into* it. The burn is the original seeing."[8] What Dewey means is that we learn to view the sensory object in a way that immediately elicits and affords an intelligent response (e.g., avoidance behavior by seeing the flame as "pain-when-touched"). In learning, the subject constantly offloads his or her "memory" into the sensory stimuli themselves, and thus uses the world as its own best representation. In other words, the embodied subject feeds back what he or she has learned into the way the world shows up in the next circumstance. As the "mind" extends into the sensory world in this way, there is no need for mental representation or internal world maps.

Understanding learning as stimulus reconstitution affords embodiment theory an evolution-friendly account for how our sensory-motor system has evolved to use *instincts* as a cost-effective tool for quickly avoiding harmful stimuli. For example, when we view a pile of spoiled food, we instinctually and immediately begin to feel nauseated; the stimulus itself is reconstituted to be "stomach-nausea-when-viewed." Without the need for recourse to repeated stimulus-response conditioning, embodiment theory ensures that as few as one response can reconstitute the sensory object to entail immediate and long-term leaning. Furthermore, because sensory stimuli are constantly changing in this way, embodiment theory accounts for how each stimulus-response experience is radically unique as it is phenomenologically experienced.

Whereas behaviorism struggles to account for how the subject notices similarities between his current context-situated stimulus perception and abstracted stimulus-response pairs from his past (without recourse to internally stored memory), embodiment theory realizes that the organism's phenomenological engagement with the world immediately provides these connections and similarities. That is, by offloading "memory" into the phenomenal objects themselves, these objects elicit an intelligent response as soon as they enter the reflex circuit. The perceiving subject constantly uses the world as a backdrop for storing (i.e., embodying) his or her dispositions and past experiences. Importantly, this process is very similar between organisms due to similar biological makeup and shared evolutionary incentives – not to mention

---

[8] Dewey, 359.

language, which will be discussed later. Generally speaking, the organism uses the world itself as its map without the mediation of internal representation.

Thus, the reflex circuit of embodiment theory takes stimuli and responses as not self-contained items, but instead as mutually-constituted components of their overall circuit from the onset of their involvement in action and perception. For a stimulus to register – the very fact *that* it is a stimulus for a subsequent response – requires it to be interpreted as something that has importance for coordinated action. The "Gestalt effect" has particular relevance here. When a group of individuals view a Gestalt image, such a necker cube, an "identical" picture will be perceived differently between subjects as one of two forms (e.g., a cube facing inwards *or* outwards, a young lady *or* an old lady; see Figure 1). This fact demonstrates that we do not view objects atomistically, but instead, from the start, as action-relevant, holistic forms; to borrow a phrase from Wittgenstein, all seeing is seeing *as*. It is not that we interpret or infer form and meaning from the "identical set of lines that comprise the picture," but rather the reverse: we are presented with a meaning-laden, already-formed Gestalt image and only then do we decompose the image into smaller, formless segments from which we infer its atomistic makeup.

Figure 1

For example, when viewing the old-young lady image, the horizontal line towards the center gains its meaning as it functions in the overall image: as either the mouth of the old woman's drooping face, or the necklace of a young lady. In this way, the Gestalt effect also reveals how embodiment theory understands stimulus and response as components that obtain their respective meanings top-down via the overall function of the reflex circuit to which they belong. In one reflex circuit, the stimulus of a flame's light can excite neurons in the brain that elicit the arm as a response action, whereas simultaneously, in another reflex circuit, pain in the very same sensory-motor areas of the arm can be the stimulus for the response of withdrawing the arm. Such an account provides the organism with an adaptable and efficient mechanism for learning in a dynamic environment without the need for mental representation.

In addition to its theoretical and practical virtues compared to behaviorism, embodiment theory also provides a viable alternative to a third problematic theory of perception, computation-based artificial intelligence (AI). This view of cognition, known as "computationalism," attempts to model the human mind using computations within a computer's internal representations, symbol circuits, and world maps. Researchers in this field designed the CYC computer, one of the largest AI projects to date, which consists of a vast bank of internally stored "information units" that serve as a detailed encyclopedia of explicit facts, maps, and rules about the world. To operate in the world, the computer uses these explicit facts and internal world models as a knowledge base from which to calculate an intelligent response.

As a result of this set-up, embodiment theory finds computationalism to be "dualistic" in two regards. First, although the CYC computer's symbol manipulation circuits are physical, they are not embodied. That is, even though they do not belong to a distinct, non-physical substance (i.e., Descartes' thinking substance), central pre-processing still operates via abstract, encoded symbols, which are *disembodied* in that they are not comprised of features of actual bodies in actual situations in the world. Instead, they are comprised of abstract variables and symbols used for a multitude of sensory inputs. The second way computationalism is "dualistic" is that it *abstracts* away from the local environment. That is to say, the CYC computer gathers sensory input and then computes abstract, representational models of the world for centralized symbolic processing. For example, when directing a robot arm, the program directs movement by adjusting the position via constant reference to an internally represented world model and intended position for the arm. On the spot, local dynamics between the device's motor system and the sensory environment do not, to borrow Clark's words, "soft-assemble" coordination, nor can they re-organize or change the CYC perceptual inputs, as with stimulus reconstitution in the reflex circuit. Instead, sensory stimuli are processed via permanent physical circuits.

These two "dualisms" give rise to the *frame problem*. According to Daniel Dennett, an early commentator on the problem's relevance to AI, computation-based AI must face the insurmountable problem of how to successfully infer the effects of an action without *explicitly* calculating every one of its action's non-effects. Dennett explains this predicament with a clever example of a robot designed to remove an object resting on a wagon from a room.[9] As with any centralized, disembodied device, the robot determines how to act based on explicitly calculating what it should do via symbol manipulation and internal world models.

In Dennett's example, the first robot fails in its task because it cannot predict the consequences of its actions: pulling the wagon out also turns out to have brought with it a time bomb located in the wagon beside the object. Although it knew the bomb was in the wagon, this brave artificial creature did

[9] Daniel Dennett, "Cognitive Wheels: The Frame Problem of Artificial Intelligence," in *Minds, Machines and Evolution*, ed. C. Hookaway (Cambridge: CUP, 1984), 2.

not *explicitly* predict that the action of pulling out the wagon would also bring with it the bomb. Hence, a second version, the "robot-deducer," is designed to calculate *all* the outcomes of its actions before doing anything. This unfortunate robot gets blown up as it sits deducing one-by-one every single consequence of pulling the wagon out of the room no matter how irrelevant (e.g., will removing the wagon change the color of the walls?). Thus, the designers program a third robot capable of differentiating between relevant and irrelevant implications of its actions. Still, however, the time bomb eventually explodes as the robot sits explicitly ignoring the thousands of implications it has determined to be irrelevant.

In total, Dennett's example demonstrates that in order to have the most basic form of adaptive intelligence, computation-based AI must explicitly rule out *every* obvious irrelevant consequence of its actions before acting in the world. In the robot's language of "mathematical logic" it is necessary to make explicit not just the changes brought about by its actions but also all those features of the environment that do not change. Before carrying out any response, the CYC computer must process all the minor features about the world that we as humans assume but wouldn't bother to overtly say or predict. Any and all commonsense facts about the world must be either previously stored or calculated in this way. Adding more and more data to the CYC's knowledge bank will surely not make the robot better at learning or quicker to adapt. Thus, computationalism is a terribly inefficient model of cognition and a poor model of mental functioning. Without any potential for real adaptation, when faced by a dynamic environment, computationalist artificial creatures are fatally flawed.

Embodied AI uses a decentralized, layer-based architecture to provide a better chance of avoiding the frame problem. Clark cites research done by Rodney Brooks at MIT's AI lab, where a robot is built with several activity-based "layers."[10] These robots are composed of a collection of *embodied*, non-centralized subsystems, each of which consists of a complete behavior-determining, input-to-action description. For example, one layer could be "stop if an object is directly ahead." As learning takes place (and as the robot "evolves"), more and more layers are added incrementally, each stage creating a functional whole. One of Brook's layer-based robots, Herbert, can successfully navigate a dynamic environment and collect soda cans. By virtue of his layered architecture, the robot can readily adapt and respond to obstacles in its path. There is no central processing or internal mapping, but simply a collection of competing layers that are selected according to sensory inputs. In this way, the environment, via the robot's sensory-motor mechanisms, guides the creature

[10] Clark, 12.

along by activating layers that automatically focus on the particular relevant features (to the layer), as well as naturally assume that certain other unexamined features remain present in the background.

To a certain extent, these layers can be understood as Hebert's possible percepts. As with the reflex circuit, these embodied layers process stimuli, from the start, as components of coordinated action. For example, if we are to apply this model to the child maneuvering about a slope, one could imagine the infant adding, deactivating, and switching between cognitive "layers" as they are elicited by its moment-to-moment bodily engagement with the slope. Particular features of the slope are not processed as mental representations of self-contained, external "bits of data," but rather as components that afford particular actions according to the overall dynamics of the current reflex circuit (or layer) that is activated. In this way, Hebert's intelligent behavior is soft-assembled via constant local, dynamic interaction between the environment and a multitude of competing and subsuming layers. If one layer is damaged, for example, the system compensates automatically and carries on acting – what Clark refers to as a "robust" solution.[11] Because of this adaptive strategy, an embodied approach to AI enables the subject to act and survive in a dynamic world by replacing computation-based map or rule-following with de-centralized engagement with the real world.

Thus, when compared to Cartesian dualism, behaviorism, and computationalism, embodiment theory, as a scientific account, presents itself as a model better prepared to avoid theoretical problems and to provide practical solutions. In total, an embodied approach asserts that perception and learning are a matter of reconstituting elements of the natural world according to subject-sensitive embodied projects (e.g., crawling, walking, reaching for a flame, picking up cans, etc.). As Clark writes, "intelligence and understanding are rooted [...] in something more earthly: the tuning of basic responses to a real world that enables an embodied organism to sense, act and survive."[12] On this view, it seems that embodiment theory posits that an organism's sensory-motor reflex circuits *afford* a world that "enables [it] to sense, act and survive."[13] To this effect, embodiment theory insists that stimulus reconstitution does not occur via mental representations (or "interpretations"), but *in* the world itself as it is lived and phenomenologically experienced by the organism.

Embodiment theory urges cognitive science to revise traditional dualistic conceptions in the field of philosophy that understand the natural world as an external and pre-given substrate from which we contemplate, conceptualize,

---

[11] Clark, 43.
[12] Clark, 4.
[13] Ibid.

or infer conscious representations. Alternatively, according to an embodied approach, the world is to be understood (under a monist and perspectivist framework) as enacted and afforded continuously by the subject's embodied engagement with it. In order for cognitive science to provide the most accurate, objective understanding of the organism's cognitive system, it must take into account the fact that the world exists as phenomenologically distinct and perspective-laden for each embodied subject. That is, as the organism learns and perceives, he or she does *not* reconstitute a pre-established world (i.e., start from scratch, so to speak), but instead reorganizes his or her most recent enacted perception of it. At the same time, our perception of the world is essentially naturalized and evolves as we learn new ways of acting with our bodies.

According to this "naturalized perspectivist" stance, the world always exists as an action-laden perspective for an embodied subject. Learning and the very process of perception itself are matters of continuously offloading past experiences into the phenomenological experience itself. Without a separate realm of mental representation, naturalized perspectivism is committed to the notion that a subject can have influences on their experience of the world (qua phenomenology). Critics might caution that such an argument is circular insofar as our understanding of the body is itself only our perspective. However, such a critique assumes a dualistic framework and can be displaced by the simple, yet important matter of fact that we all experience a world where humans have similar bodies and similar evolutionary incentives – and indeed are defined by this very fact. As Merleau-Ponty puts it, "the subject does not live in a world of states of consciousness or representations from which he would believe himself able to act on and know external things by a sort of miracle."[14] Thus, for cognitive science to remain progressive and accurate, it must include first-person phenomenological reports within the data of science. Maurice Merleau-Ponty provides an analysis of such data.

At the beginning of his book, *Phenomenology of Perception*, Merleau-Ponty asks us to reexamine our understanding of the term "sensation." For Merleau-Ponty, the Cartesian notion that we experience the world as internal representations of the external, bodily world in fact "corresponds to nothing in our experience."[15] Whereas Descartes investigates how a "thinking subject" relates to his "external world," Merleau-Ponty starts his investigation by taking the relation to be direct, immediate, and embodied. He rejects mind-body and subject-object dualisms in favor of understanding the mind and perception as essentially embodied from the start. Merleau-Ponty argues that ordinary

[14] Maurice Merleau-Ponty, *The Structure of Behavior*, trans. A. Fischer (Boston: Beacon Press, 1964), 189.

[15] Maurice Merleau-Ponty, *Phenomenology of Perception*, trans. Colin Smith (New York: Routledge, 2002), 3.

phenomenal experience presents us *immediately* with a world of "external objects" – not separate, self-contained internal representations or sensations from which we infer the object.

Moreover, and central to Merleau-Ponty's theory of perception, we experience the object *through our bodies*; our physicality influences what we see and is the necessary and permanent condition of experience. In this way, sensory perceptions are not mental states or intentional representations, but instead ways that our conscious intentional body comports itself to objects in our situated experience. Merleau-Ponty references various Gestalt images and perceptual illusions, which demonstrate how we always view objects as meaningful, holistic forms with distinctive foregrounds and backgrounds. If our phenomenal scene was atomistically built, "[w]e ought, then, to perceive a segment of the world precisely delimited, surrounded by a zone of blackness."[16]

According to Merleau-Ponty, in the process of perception, the human organism is always faced with the task of orienting its body so as to achieve its best relation with perceptual objects – what he describes as its "maximal grip." In this way, an accurate sensory perception is nothing more than a skilful bodily engagement with an object, or as he writes, "an optimal body-environment relationship that relieves the 'tension'."[17] In other words, the coherence of images emerges when the body properly adjusts itself so that it can act on the object. For an individual walking about in a museum, for example, Merleau-Ponty asserts that it is just a natural fact that our body, even without our conscious, explicit violation, will want to move to the ideal viewing angle to see the paintings with maximal clarity of detail and overall form. The process of perception is a process of oscillation towards equilibrium. Thus, the dynamic relationship between objects and our bodies is normative in that it adjusts to arrive at better or more correct bodily engagement with the object. There is no universal, abstract world, but only a host of perceptions that are possible by virtue of our body and determined by our natural drive towards maximal grip. Between individuals, these perceptions (i.e., phenomenological experiences) are highly comparable due to our similar bodies and evolutionary drives.

As an extension of his theory of maximal grip, Merleau-Ponty develops an account of learning that has many similarities with Dewey's reflex circuit. According to Merleau-Ponty, at the first stage of learning, we act like centralized computers: we deal with the world (poorly) by decomposing it into context-free, often quantitative discriminations (e.g., a driver noting his speed by consistently looking at the speedometer), and then act according to a list of rules that address

[16] Merleau-Ponty, *Phenomenology of Perception*, 5.

[17] Hubert Dreyfus, "Merleau-Ponty and Recent Cognitive Science," in *The Cambridge Companion to Merleau-Ponty*, ed. Taylor Carman and Mark BN Hansen (Cambridge: CUP, 2005), 138.

these features (e.g., shift once every 15 mph). However, as the beginner learns, he or she starts to notice additional aspects of the situation and becomes more closely engaged with it. For example, whereas a novice driver approaching a dangerous curve at a high speed will proceed to explicitly consider the angle, speed, and outside conditions, the expert driver will *feel* that he is going too fast and then decide to ease the break. The expert driver has developed a vast repertoire of situational discriminations that allow him to make subtle, refined perceptions. Merleau-Ponty argues that it is precisely *not* a matter of collecting more and more "stored" rules via mental representations for one to follow. Instead, skill acquisition occurs when *new dispositions* emerge in response to the situation. Thus, intelligent behavior is when a subject's dispositions ensure that in the phenomenal experience the relevant similarities show up that elicit a natural coordinated response.

As a way of addressing the frame problem head-on, Merleau-Ponty recognizes that the ability to restrict perceptional intake to the relevant features for action and response is an essential part of learning. In his words, "a person's projects polarize the world, bring magically to view a host of signs which guide action."[18] These "signs which guide action," that show up in the objects themselves are very similar to the offloaded "memories" inserted during stimulus reconstitution in Dewey's model and in the theory of naturalized perpsectivism. Situation-focused dispositions, just like layers and reflex circuits, warrant that the embodied subject naturally focuses on relevant features and inexplicitly assumes that certain other non-effects remain in the background. Thus, learning for Merleau-Ponty and Dewey is the gradual replacement of reasoned, rule-following responses with intuitive reactions, where the subject uses the world itself as a means to elicit a response without having to mentally represent the world. A model example of Merleau-Ponty's phenomenological account, where an individual skillfully performs tasks without the intermediary conscious reference to rules, is muscle memory. As with the reflex circuit, Merleau-Ponty's phenomenological account posits that the organism uses the world itself (as it is experienced) as its own map and best "representation."

The later philosophy of Ludwig Wittgenstein also supports the viability of naturalized perspectivism. In particular, Merleau-Ponty's account of perception and skill acquisition lends itself closely Wittgenstein's insights on rule-following. Wittgenstein argues that "no course of action could be determined by a rule, because every course of action can be made out to accord with the rule."[19] To explain this, he provides the example of an individual (presumably) following the rule "plus 2" while actually adding increments of four after he or she passes

---

[18] Dreyfus, 132.

[19] Ludwig Wittgenstein, *Philosophical Investigations* (Upper Saddle Diver, NJ: Prentice-Hall, 1958), 201.

1000.[20] Wittgenstein's point is that we can, and in fact do, successfully operate not by explicitly following a rule. To understand correctness of use, there is no recourse to some internal authority beyond the actual application in use of the rule. This is because there is no fact of the matter that determines the correct application of some internal rule. Rather, skillful and intelligent behavior is revealed in use via community assent. Wittgenstein describes our participation in a community with pre-established criteria for correctness as the "form of life" that we adopt in language and action. Thus, this general characterization of Wittgenstein supports embodiment theory insofar as it understands human action not to be a process of following and applying internal rules, but rather as a means of engaging in the world.

Wittgenstein extends his discussion on forms of life in his critique of a private language, which suggests the presence and viability of naturalized perspectivism in language. On Wittgenstein's account, we learn phenomenological (i.e., sensation) language as children by participating in a pre-established form of life, or "system of reference by means of which we interpret an unknown language."[21] That is, when a child is hurt, he is trained to use *words* to express his pain instead of crying. In this way, the child learns psychological vocabulary through public training in which non-verbal pain behavior is gradually replaced by verbal pain behavior. In Wittgenstein's words, "the verbal expression of pain replaces crying and does not describe it."[22] Hence, when a child articulates pain behavior, he or she does not introspect or "describe it" (i.e., concentrate on what is occurring in some 'inner realm' and then ostensibly name it), but instead chooses to express pain behavior in another form (i.e., verbal pain behavior). The stage is already set, in that, as we learn language, we enter into a pre-established relationship between verbal pain-expressions, and pain-behavior expressed in the bodily actions of other speakers.

As children, we learn sensation words, and are understood communally, insofar as we have the "right," not the "justification," to use certain expressions.[23] Wittgenstein stresses this distinction to illustrate how a child learns the proper use of words: by participating in the sensation language game and being positively reinforced if other participants understand, not according to conditioning based on whether our "mental representations" match our verbal expressions. In total, Wittgenstein's critique demonstrates that phenomenological language has no purchase on the subject's radically singular phenomenological experience. Because language is naturalized by the body and the community,

20 Wittgenstein, 185.
21 Ibid., 206.
22 Ibid., 244.
23 Ibid., 289.

we can meaningfully operate within the sensation language game, even with an essentially unique phenomenal moment or perspective.

Remarkably, Wittgenstein's private language argument also explicitly advocates for an embodied approach to the linguistic conception of the mind and body. He argues that if we conceptually divide the psychological and physical realms, as dualism does, then certain absurd examples of language in everyday use seem possible. The fact that "I can have your pain" is semantic nonsense, and not some empirical inaccuracy, shows that such a claim is grammatically illegitimate in our sensation talk. Commenting on the referential conception of language, Wittgenstein writes, "this picture with its ramifications stands in the way of our seeing the use of the word as it is."[24] Thus, meaning exists insofar as we decide to adopt the established grammar – and with it the proper conceptual distinctions – of a particular language game. The proper conceptual distinction acquired by children is not between inner and outer, but instead between two types of outer. That is, the only way to communally understand the use of sensation language rests on how the child learns the distinction between two types of *body*: those resembling living things and those that do not. This connection arises from how a child is conditioned to use the sensation language game where sensation terms are tied to their expression through their parent's *living* body.

Under this framework, statements like "that rock has pain" properly do not hold meaning. A child verbalizes behavior by replacing bodily actions (i.e. crying) with the articulation, "I am in pain." Generally put, we are trained not to identify mental events with sensation terms, but to participate in an established way that living things act. Thus, Wittgenstein provides an account of language use and acquisition that is against representaionalist rule-following and in favor of an embodied approach. Moreover, he provides a theory of phenomenological language that allows each individual to have a radically unique, phenomenological sensation, while at the same time to be able to successfully communicate his or her experience in a meaningful way.

Another form of support for embodiment theory can be found in contemporary neuroscience. Recent research supports a non-representationalist picture of mental states, where the brain is considered primarily as a device for guiding action. For example, as Hubert Dreyfus points out, recent neural net research suggests that a layer-like brain model could explain learning. Simulated neural networks consist of sets of input nodes coupled with a system of output nodes, mediated in between by "hidden" connections.[25] When these neural networks are trained via stimulation of input nodes with output nodes, the

---

[24] Wittgenstein, 305.

[25] Dreyfus, 133.

weights on the connecting intermediate nodes (via "neurons that fire together wire together") are adjusted to ensure a more appropriate response next time. The current weights on the connections between input and output nodes essentially entail learning by storing the history of past input-output sequences without elaborate memories or explicit rules. Importantly, these interconnections do not serve as internal, mental representations, but as *implicit* ways of influencing future actions based on past experience. Thus, once a subject gets exposed over and over to a number of situations, the world (as it stimulates certain input nodes) begins itself to solicit the correct response.

As the hidden weights adjust with skill acquisition, new responses and situational discriminations are possible. Learning is a matter of conditioning our neural nets so that an intelligent response is automatically triggered by the currently stimulated input node. In this way, the response reorganizes the stimulus in a top-down manner, as with Dewey's reflex circuit. Furthermore, as intentional bodies moving towards maximal grip, learning should be understood as analogous with perception. Learning occurs when an individual adjusts himself so that relevant features naturally *present themselves* in their environment. In Merleau-Ponty's language, our past naturally projects into the current phenomenal moment by creating a "milieu" of possible affordances for further action.[26] With this cognitive architecture, the subject avoids the frame problem, as he does not need to explicitly determine relevant rules to apply; it happens naturally as his or her hidden nodes will afford the proper response. Again, because of the simple fact that our bodies are built in similar ways with similar evolutionary incentives, these nodes develop between individuals to the extent that we share very comparable phenomenologically-lived worlds. As with our previous examples, these behavior-determining neural nets can carry out adaptive intelligence without mental representations or internal models.

In addition, there is a significant amount of other evidence in experimental neuroscience that supports a non-representationalist view of the brain. Two prominent figures in contemporary cognitive neuroscience, Gerald Edeleman and Bernard Baars, provide models that take the brain as primarily a device for aiding in the selection of intelligent behavior. Both of their models hold that the top-down neural dynamics between sets of neurons in the brain and the sensory stimuli serve to select physical (i.e., neural) arrangements towards the world that elicit an automatic response.

Specifically, Edelman's Theory of Neuronal Group Selection asserts that a cluster of highly resonating, interconnected neurons – what he refers to as a "dynamic core" – contributes to consciousness.[27] As its name suggests, a dynamic

[26] Dreyfus, 158.

27 G.M. Edelman, *Second Nature: Brain Science and Human Knowledge* (London: Yale University Press, 2006), 27.

core is temporally synchronized via electrophysiological firing rates, while at the same time continually rearranging its own anatomic composition.[28] Borrowing insights from Darwin, Edelman holds that competition serves as the primary drive for neuronal group selection. As the theory of naturalized perspectivism would accept, in Edelman's theory, brain-world dynamics provide a pool for the selection of a particular set of neurons that comport towards the world. These neural networks should be conceived not as representational, but simply as bodily changes that entail certain input-output circuits that move the organism towards maximal grip. Those neurons neighboring the dynamic core compete for inclusion via electrophysics, while those that are irrelevant for the current action goals proceed to leave the dynamic core. The constant need for re-organization of the core may reflect Merleau-Ponty's notion of the natural drive for equilibrium.

Similarly, Bernard Baars' notion of a global neuronal workspace asserts that conscious awareness arises not from a fixed part of the brain, but rather from the coordinated electrical activity of widely distributed network neurons.[29] Like the dynamic core, the global workspace is a select coalition of neurons that are resonating throughout the brain. Again, these neural networks should be thought of not as internal representations of the outside world, but instead as physical arrangements embedded in one overall brain-body-world system. As neighboring sets of neurons compete for inclusion, sensory input automatically elicits the workspace. Top-down electrophysical dynamics of the workspace can affect which input nodes are activated in the workspace. In this way, Baars views conscious awareness as a problem-solving "workspace" for selecting the proper input and output nodes (i.e., stimuli and responses) that efficiently carry out evolution-friendly actions. This model of the brain is remarkably similar to Merleau-Ponty's description: "a field of forces which express the influence of external agents [...] balance themselves according to modes of preferred distribution."[30] As Merleau-Ponty would agree, the workspace will process novel stimuli and unanticipated events only if they meet the top-down constraints of the neural network – a method for automatically restricting perceptual intake so that the organism moves towards maximal grip. To this effect, what counts as meaningful noise is mediated by top-down, normative constraints of the brain-body-world system as a whole.

All in all, with the support of these models, embodiment theory takes a naturalized perspectivist stance on the mind. Because the embodied subject

[28] G.M. Edelman, *Second Nature: Brain Science and Human Knowledge* (London: Yale University Press, 2006), 28.
[29] J. Newman, B. Baars, and S. Cho, "A Neural Global Workspace Model for Conscious Attention," *Neural Networks* 10:7 (1997), 1199.
[30] Merleau-Ponty, *Phenomenology of Perception*, 4.

enacts the sensory world relative to his or her embodied projects and histories, the perceptual experience is necessarily different for each subject. By taking this perspectivist stance on the study of cognition, embodiment theory opens itself to charges of "anything-goes" relativism. Such accusations, however, ignore the fact that although subject-sensitive influences affect perception, they do so in a very naturalistic, evolutionarily-determined way. That is, human organisms have evolved as a species to share a bodily structure and evolution-friendly drives. We perceive a similar world insofar as we share certain bodily features that afford particular bodily comportments to sensory objects. Neural arrangements, as one example, are organized by top-down electrophysical forces. Thus, naturalized perspectivism, properly conceived, defends itself from charges of relativism by anchoring the plurality of our phenomenological experiences (e.g., different stimulus reconstitution between individuals) to our similar biology and thus our similar sensory experiences.

Summarizing, it becomes clear that naturalized perspectivism, with the support of cognitive science, phenomenology, linguistics, and neuroscience, presents itself as a viable model of perception. An embodied approach to cognition argues that the motor system engages with the world via a multitude of competing and subsuming reflex arcs. Learning occurs as we offload "memory" into the world and in turn create new neural arrangements and new situation-specific discriminations which exist in each organism's phenomenal moment. Such a model lends itself nicely to evolution's tendency to build intelligent creatures circuit-by-circuit (or layer-by-layer) in functional wholes that can most efficiently and adaptively respond to the dangers and rewards in one's dynamic environment.

All in all, the theory of naturalized perspectivism provides a powerful explanatory model of mental functioning on both theoretical and practical levels. On the theoretical level, it avoids conceptual problems common to other approaches: the problem of mind-body interaction for Cartesianism, the problem of learning without mental states for behavioral psychology, and the frame problem of computation-based artificial intelligence. Furthermore, an embodied approach to cognition starts by viewing the brain as a device for helping to determine actions that allow the organism to survive in challenging and rapidly changing environments over the course of its life. In this way, it is a theory that foregrounds the practical dimensions of mind in such a way as to make them more open to evolutionary forms of explanation. Such an account supports how the evolutionary process has constrained the development and function of the nervous system. In this regard, embodiment theory argues that the mechanisms of perception have developed according to one of nature's overriding rules: that natural selection promotes brain-body-world interactions that constitute the most efficient and adaptable ways of responding to threats and rewards in the organism's environment.

By realizing that our cognitive architecture implicitly restricts perceptual intake to relevant features, phenomenology appears to be an influential aspect of perception, and thus an important subject for scientific studies of cognition. The theory of naturalized perspectivism asserts that a subject's embodied dispositions *create* for it a world where it can most efficiently operate and socially interact as a biological organism. If it is to incorporate these insights, cognitive science must take the world not as an objective, universal substrate, but as importantly different according to features of each organism's biology and phenomenological history. Overall, the interdisciplinary support gathered in this paper encourages researchers to begin to replace the traditional dualistic approach to the mind-body problem with a more holistic methodology.

BIBLIOGRAPHY

Clark, Andy. *Being There: Putting Brain, Body and World Together Again.* London: The MIT Press, 1997.

Dennett, Daniel. "Cognitive Wheels: The Frame Problem of Artificial Intelligence." In *Minds, Machines and Evolution*. Edited by C. Hookaway. Cambridge: CUP, 1984.

Dewey, John. "The Reflex Arc Concept in Psychology." *Psychological Review* 3:4 (1986).

Dreyfus, Hubert. "Merleau-Ponty and Recent Cognitive Science." In *The Cambridge Companion to Merleau-Ponty*. Edited by Taylor Carman and Mark BN Hansen. Cambridge: CUP, 2005.

Edelman, G.M. *Second Nature: Brain Science and Human Knowledge*. London: Yale University Press, 2006.

Merleau-Ponty, Maurice. *Phenomenology of Perception*. Translated by Colin Smith. New York: Routledge, 2002.

Merleau-Ponty, Maurice. The Structure of Behavior. Translated by A. Fischer. Boston: Beacon, 1964.

Newman, J., Baars, B., and S. Cho. "A Neural Global Workspace Model for Conscious Attention." *Neural Networks* 10:7 (1997).

Wittgenstein, Ludwig. *Philosophical Investigations*. Upper Saddle Diver, NJ: Prentice-Hall, Inc., 1958.

ACKNOWLEDGMENTS

I would like to thank my academic advisor, Professor John Fennell, for his constant diligence, patience, and excellent teaching skills over the course of this senior year Mentored Advanced Project and as an academic mentor for the past two and half years.

Giving An Account Of Explanation:

The Language User Versus The Technical Approach

Patrick Decker
University of Colorado, Boulder

One of the main divides between various approaches to an account of explanation centers on the 'language user' versus the 'technical' approach.[1]* In the following paper I will explore this divide, arguing that the technical approach should be preferred.

In section 1.1 I will outline the two approaches as they are *typically* characterized, touching on some of the advantages and disadvantages of each approach. I will touch on how the divide manifests itself in practice. In section 1.2 I will argue that the technical approach is often mischaracterized. I will attempt to clarify the technical approach, and what I take to be the most basic difference between the two approaches. In section 2 I will introduce a case of scientific reduction as an analogy to the technical approach to help show that the consequences of the technical approach are acceptable. In section 3 I will propose three possible objections the language user might make, and I will respond to each.

I will conclude that once the technical approach is clarified, and its consequences likened to those of scientific reductions, it should be considered a promising approach. Further, when one of the possible drawbacks of the language user approach is considered, it should seem plausible that the technical approach is preferable to the language user approach.

SECTION 1.1

The standard way of characterizing the contrast between the technical approach and the language user approach is as follows:

The technical approach does not utilize actual applications of the term 'explanation,' whether in everyday life or science, as a starting point for an analysis of explanation. Therefore, someone friendly to the technical approach will not reject an account if it entails that no real explanation has ever been given. Instead, it is possible 'explanation' is misused in everyday** applications, and what we called "explanations" are not really explanations.[2]

The language user approach, however, finds the possibility that no real explanation has ever been given unacceptable. Rather, a satisfactory account must take into consideration the ways in which 'explanation' is actually used in everyday life. The language user believes that everyday applications of

---

[1] Ruben, David-Hillel. *Explaining Explanation.* (New York: Routledge, 1990). 12.

* This is a divide which can develop while giving an analysis of any concept, and constitutes a much larger methodological divide. However I think if there actually is a 'correct' way to approach a given analysis, it might be determined on a case-by-case basis. As such what follows will be framed in the context of 'explanation' but *might* also be applied to the more general methodological question.

** I intend the term 'everyday explanation' to include all 'explanations' given in practice, whether in science, history, or everyday life.

[2] Ibid.

'explanation' can act as empirical data, constituting the starting point from which an analysis of explanation can be given. A survey of the contexts in which 'explanation' is applied can lend insight into what constitutes an explanation. Therefore to completely disregard the way 'explanation' is actually used, as the technical approach is willing to do, is to miss the point of giving an account of explanation. The language user believes that the technical approach can merely give an account of some concept arbitrarily called 'explanation.'

Although I presented the two approaches as ideals, it should be realized that some combination between the two approaches is possible. In practice Carl Hempel has most famously embodied the technical approach – to the extent that in outlining essentially the same divide* in accounts of explanation, Rudolph Weingartner calls the divide "Hempelians" versus "anti-Hempelians."[3] David Lewis, and David-Hillel Ruben's accounts are also on the technical side of the divide.[4] In practice, the technical approach tends** to generate accounts of explanation that involve metaphysical facts or relations – such as laws of nature, causation, or other dependency relations. In the case of Hempel, the explanans must logically entail the explanandum, but this might not be a fundamental feature of the technical approach, despite the fact that Hempel has most famously embodied the technical approach.

Under the language user approach, metaphysical facts or relations and laws are typically abandoned. The explanans need not guarantee the explanandum, and under some accounts, it need not even make it more likely to have occurred. Often an explanation is thought to be context dependant, and some, such as Van Fraassen, say that explanation is a three-place relation.[5]

Although the methodological divide manifests itself as a messy divide in practice, I will refer to the two approaches only as ideals. I will not attempt to sort through the various directions the two approaches have lead different philosophers.

SECTION 1.2

The technical approach, as stated above, is mischaracterized. Specifically, the claim that the technical approach does not use *any* 'evidence' from the ways that we normally use 'explanation,' making any account of explanation given by

* I think it can be considered the same divide only in so much as the divide in practice is a manifestation of the methodological divide.

[3] Weingartner, Rudolph. "The Quarrel about Historical Explanation." *The Journal of Philosophy* 58:2 (1961): 29-45. 30.

[4] Lewis, David. *Philosophical Papers, Volume II: Casual Explanations*. (Oxford: Oxford University Press, 1986).

** It might be that the technical approach could be used *without* generating an account of explanation that relied on metaphysical facts or relations, *or* it might be that this is actually a necessary feature of the technical approach.

[5] Van Fraassen, Bas C. "The Pragmatic Theory of Explanation." *The Scientific Image*. (New York: Oxford University Press, 1980).

a strictly technical approach merely an account of something *called* explanation.

I think that this is mistaken. Although the technical approach is willing to conclude that no actual explanations are ever given in practice, it cannot actually disregard all existing *intuitions* about 'explanation.' The technical approach must instead be guided by *some* intuitions about explanation — intuitions which in some way must be linked to, or generated by, the ways in which we use 'explanation' in practice. These intuitions can be teased out by the careful application of thought experiments designed to inform us about what we really think an explanation is.

Even the most strictly technical approach cannot completely reject all intuitions about 'explanation,' since if this were actually so, I agree that all one could do is make something up and call it 'explanation.' What the technical approach recognizes is that we might be *confused* about what an explanation is. The technical approach seeks to discover some feature(s), which is *intuitively* essential to explanation.* This feature could be a feature which we *think* exists in everyday explanations, but which is in fact missing, and we are simply deluded into thinking it is present. This would account for why on the one hand a real explanation has never been given, but why intuitively we think that we have given many real explanations. Further, if it could be made *perfectly clear* to anyone — even a language user — that this feature really is lacking from everyday explanation, he or she would concede that everyday 'explanations' are *not* explanations.** The trouble lies in the fact that it might be quite hard to show that this feature *really* is lacking, perhaps because we have become so accustomed to everyday explanations that we unconsciously build this feature into the explanation.†

Therefore, the technical approach separates itself from the language user not in virtue of the complete rejection of 'evidence,' but instead by the thought that although *some* of our intuitions about explanation *must* be correct, we must be careful in choosing which ones to trust, since many of them might be incorrect. The approach recognizes that we have epistemic limitations that may require careful thought to overcome. As such it may turn out that no actual explanations have ever been given in practice.

SECTION 2

* I assume, based upon both my own intuitions and the accounts given in practice, that this feature will be some type of metaphysical fact or relation — specifically, it might be that explanation depends on the fundamental laws of physics and the fundamental conditions of the universe.

** In practice the issue of partial versus complete explanation would arise at this stage, however I will ignore the issue in this paper.

† Velleman's distinction between 'emotional understanding' and 'intellectual understanding' might be employed here as a way of accounting for why we mistake everyday explanations for real explanations. It might be true that everyday explanations give emotional understanding, but it is a 'projection error' to think that these give intellectual understanding.

Scientific reductions of everyday concepts might be analogous to the consequences of the technical approach. Such cases might help show that the consequences of the technical approach are acceptable.* Consider the following example: scientific discoveries about what it is to be 'solid' can force us to revise our concept of 'solid.'

Specifically, in the everyday use of 'solid,' it might be part of the concept that the space a given solid object occupies is filled with matter. However, scientific inquiry has shown that in fact a solid is actually mostly empty space. This scientific discovery ought to make one revise his or her concept of solid. It is possible for scientific discoveries to undermine every aspect of the everyday – scientifically naïve – concept of 'solid.' If such a replacement occurs, even if the replacement is only partial, there are two possible choices we can make.

We can accommodate for these scientific discoveries in one of two ways. The first way would be to change our everyday concept of 'solid' to accommodate the new scientific fact about solid matter; for example, we can no longer say that something solid is something that a hand can *never* pass through.** The second way would be to split the concept; we can say that there is 'solid' in the everyday sense, and 'solid' in the scientific sense. However, if we choose to split the concept, then every instance of 'solid' in the everyday sense is, strictly speaking, a misuse. I think that such scientific reductions are fairly clear-cut and ought to be considered acceptable.

Similarly, the technical approach is open to the discovery that some aspects of the everyday concept of 'explanation' are mistaken, in which case the concept should be amended to accommodate this. It should be an acceptable consequence if the amended concept entails that every application of the term 'explanation' in practice is a misapplication, in the same way that this is an acceptable† consequence of scientific reductions.

SECTION 3

In this section I will propose three possible objections the language user might make and attempt to respond to each in turn.

The first objection a language user might make is to simply claim that the technical approach cannot give us the sense of explanation we are interested in. Whatever explanation is, it *must* conform to the way in which we *actually* use

* Although some philosophers will surely reject scientific reductions, I will assume that such reductions are acceptable, and liken the technical approach to them because I consider them to be a clearer case.

** My understanding is that current physics says that it is possible, although highly improbable, to pass through a brick wall.

† Of course not everyone will agree that the consequences of scientific reductions are acceptable.

'explanation'; anything else is wrong, or at best, uninteresting.

However, this objection fails to realize a possible problem with the language user approach. The technical approach assumes that due to epistemic limitations, it is possible that we have been mistakenly applying the term 'explanation.' To deny this as at least a plausible possibility is a mistake. Historical instances of this can even be cited; consider the term 'witch.' If 'witch' is to mean a person with magical powers or something of the sort, then presumably any actual instance of the term 'witch' is a misapplication of the term. Given this, it would be of little use to survey instances of the application of 'witch.' Further, if one should insist on using the actual applications of the term as indications of its concept, then in the case of 'witch,' they would not come away with an account that included a person with magical powers, since in no instance of its use was this the case. Therefore, if 'explanation' turns out to be like 'witch' the language user approach would be of little use.

The second objection the language user might make is to argue that cases of scientific reduction, such as 'solid,' are not analogous to the technical approach. It might be pointed out that advances in science have not shown that we have been incorrectly applying the 'term' solid. Everything we called 'solid' before various scientific discoveries, we continue to call 'solid,' even while conforming to the amended scientific concept of 'solid.' Therefore, the language user might claim that the technical approach is not analogous to cases of scientific reduction, since the technical approach might have the consequence that nothing we called 'explanation' prior to the technical approach can be properly called 'explanation' after we accept the technical account.

The language user might concede the following response to this objection: in fact, what scientific reductions do *is* exactly what the technical approach is willing to do. Although we continue to point to the same objects and call them 'solid' even after a scientific reduction, we could not correctly do this if we insisted on keeping the pre-scientific conception of solid. If we accept the scientific reduction, it follows that no object is, or was ever, 'solid' conceived in the pre-scientific-reduction sense.*

The third possible objection the language user might make is as follows: I have argued that the technical approach seeks to discover some intuitively fundamental feature of explanation, but I have not given an example of such a feature, nor, do I think, has any philosopher thus far in the philosophy of

* Instead one might respond to my claim here by saying that in actuality what we do is create two senses of 'solid,' an everyday sense and a scientific sense. Frankly I am unable to take a firm stand on what in fact occurs in the case of a scientific reduction. Whether we completely replace the old concept and simply keep the term, or if we spilt the concept into two senses. However it does seem that if we split the concept we are mistakenly applying 'solid' when used in the pre-scientific-reduction sense. If the concept is split, and it turns out that this is an acceptable result in scientific cases, then the analogy is not lost. It might be that splitting the term 'explanation' is also an acceptable result.

explanation done so. Therefore, the language user might simply claim that, considering no such feature has been found, it is unlikely such a feature actually exists. They would object that I make only a claim about the *possibilities* of the technical approach. The language user would likely concede that they cannot show that it is *impossible* for the technical approach to accomplish what I have argued it could, but nevertheless the approach has thus far failed, and it might be reasonably claimed that it is time to look elsewhere.

Although I think this is a good objection, I would respond that if the feature explanation relies on is a metaphysical one,* it is plausible to suppose that the universe is made up of fundamental particles governed by fundamental laws of nature that can in principle be discovered. It might also be possible for the technical approach to demonstrate that such laws and conditions are required by explanation without actually knowing the laws and conditions.

CONCLUSION

The fact that the technical approach is willing to conclude that no explanation has ever been given in practice should not constitute sufficient grounds to abandon the approach. The technical approach does not entail a complete rejection of intuitions about explanation and is therefore not in danger of giving an arbitrary account. Scientific reductions result in analogous consequences which are often considered acceptable. Furthermore, the language user approach has problems of its own. If it is plausible that we misuse the term explanation in practice, as we do other words, then the language user approach will fail to generate a correct account. Given these points, I think it is at least plausible that the technical approach should be favored over the language user approach. Whether an acceptance of the technical approach should lead one to split the term explanation into two senses is a question that deserves further attention.

* This might seem a contentious claim in itself, but I would argue that since this seems to be the direction the technical approach has taken philosophers, it is therefore reasonable to assume that this is the type of feature the technical approach might discover.

BIBLIOGRAPHY

Lewis, David. *Philosophical Papers*, Volume II. *Casual Explanations*. (Oxford: Oxford University Press): 1986.

Ruben, David-Hillel. *Explaining Explanation*. (New York: Routledge): 1990.

Van Fraassen, Bas C. "The Pragmatic Theory of Explanation." *The Scientific Image*. (New York: Oxford University Press): 1980.

Weingartner, Rudolph. "The Quarrel about Historical Explanation." *The Journal of Philosophy* 58:2 (1961): 29-45.

A Case for Stage View Presentism

Jorgen Hansen
Utah Valley University

1. PRESENTISM, ETERNALISM, AND PERDURANTISM

In her paper entitled "Presentist Four-Dimensionalism," Berit Brogaard argues for a presentist theory of perdurantism.[1] *Perdurantism* is the view that objects are 'spacetime worms' which persist through time in virtue of having temporal parts at any particular time. Objects, then, are aggregates of these temporal parts (which I will call *stages*). Each stage is not temporally extended *per se*, but each spacetime worm (or aggregate of stages) extends temporally by having stages fixed at many successive times. Thus, perdurantists argue that stages exist at a time, whereas worms exist over time. *Presentism*, in contrast, is the view that "the only things that exist are things that presently exist."[2] Presentism is often regarded as being incompatible with perdurantism. If it is the case that only present things exist, then it seems that temporally extended worms cannot exist.

Brogaard counters this claim by arguing that only *stages* (instantaneous temporal parts) of objects *exist*, but that "objects have four dimensions in the sense that they have an unfolding temporal dimension in addition to the three spatial ones."[3] So at any (present) time, every stage is wholly present, but stages themselves never persist from time to time – i.e. stages only exist at the (present) time, but never at multiple times. Thus, Brogaard's view can so far be summarized as follows: only the present exists, and only stages exist at any present time; but stages are merely *parts* of worms, so it seems that things like persons, cups, and so forth never actually exist. So, what we seem to have here is a three-dimensional view that has been "smuggled… in the back door."[4]

However, Brogaard also counters this objection by pointing out that *events* are temporally extended, and yet these never exist wholly at one time. Brogaard argues that, similar to events, "objects, such as you and me, may have extended temporal parts even though these are parts which exist always only in the sense that they unfold themselves, incrementally, through their successive stages."[5] Accordingly, although objects never wholly exist, they can temporally extend (or persist) in virtue of their successive stages.

Thus, Brogaard argues that *we are worms*, but I, the worm, only partially exist at any given (present) time. This is similar to the eternalist version of the worm theory. The eternalist (or non-presentist) ontology is one wherein the past, present, and future all really exist. So, the present enjoys no privileged existence. The worm extends temporally by having parts at every successive

---

[1] Brogaard, Berit. "Presentist Four-Dimensionalism." *The Monist* 83:3 (2000): 341-356.
[2] Ibid., 3.
[3] Ibid., 4.
[4] Ibid., 7.
[5] Ibid.

time, just like a spatial object (such as an arm) extends spatially by having parts at every successive place. The eternalist worm view differs from the presentist view, however, in that the spacetime worm itself, as well as all of its parts, exists (tenselessly) over time – just like an arm exists over space. Whereas on Brogaard's view, the analogy between space and time seems to disappear, in that the spacetime worm never actually exists *per se*, due to the fact that only present things exist.

This point is pressed by Jiri Benovsky in an attempt to refute Brogaard's thesis. Benovsky argues that the presentist version of perdurantism faces a problem that the eternalist version avoids: the problem of *objects having non-existent parts*. Benovsky asks "how is it possible to claim that material objects have temporal parts at other times than the present if these parts don't exist?"[6] The problem for the presentist perdurance theory, then, is that on this view objects never wholly exist, but only partially exist; so at any time at which a part exists, all other parts of that so-called object do not. So where the eternalist worm theorist is committed to the view that objects only partially exist at any given time, such a view allows that one can take a tenseless perspective, where *all other parts* also exist, however just not at *this particular time*. But the presentist is committed to the former claim, where other such parts simply do not exist, since tense is taken seriously. Thus, on Brogaard's view, the vast majority of any object's parts are absent at any present time. Benovsky cites this problem and asks why we should regard these stages as *parts* of any object — given that no such object exists?[*]

I find Benovsky's argument against Brogaard's view very strong – strong enough, at least, to warrant a revision of her view. Presentists usually avoid perdurantist claims of persistence *for the very reason* that saying an object only *partially* exists at any given (present) time is an unattractive consequence to have. Thus, when Brogaard states

> According to presentist four-dimensionalism, then, to say that W. V. O. Quine exists (with tensed 'exists') must be characterized as loose talk. What we really mean is that an instantaneous stage of W. V. O. Quine exists. One might feel uneasy about [this] idea... [b]ut there is nothing strange here. We also speak of time as if it existed, although strictly speaking, at any given moment, only one of its instants… exists.[7]

[6] Benovsky, Jiri. "On Presentist Perdurantism." *Nordic Journal of Philosophy* 8:2 (2007): 79-88. 83.

[*] Benovsky also argues that Brogaard's view does not avoid the problem of change, a claim which Brogaard argues makes her view more attractive than the eternalist view.

[7] Broogard, 4.

I argue that this statement yields two unattractive results. First, Brogaard states that only a *part* of Quine exists, a problem presentism not only usually avoids, but a problem that most presentists *want* to avoid. Second, Brogaard seems to characterize 'time' in such a way that each present instant exists only as *part of* some larger thing that is *time itself*. This, to me, seems to present presentism in a way which is contrary to the standard view. Ned Markosian (who argues in favor of presentism) states: "Presentism seems to entail that there is no time except the present time."[8] This claim seems much more accurate to me. When Brogaard states that only one of *its* (i.e. *time's*) instants exists, she seems to characterize time in such a way that it, too, has parts. But this seems to more closely resemble *eternalism*, where time is understood as having parts.[*] In light of these problems which face Brogaard's view, I find the presentist version of the worm theory to be less attractive than the eternalist version. But, I argue that there is still hope for presentist perdurantism, because the 'worm theory' is not the only perdurantist view of persistence.

2. STAGE VIEW AND PERSISTENCE

Ted Sider has offered another version of perdurantism, which he calls the "Stage View."[**] According to the stage view (which, one should note, entails an eternalist ontology on Sider's view), *we are stages.*[9] Persons, cats, rocks, cups, and all other everyday objects are not spacetime worms. Rather, each stage (or instantaneous temporal part) of a spacetime worm is the thing itself. Also, the predicate 'person' typically applies to the person stage and not the spacetime worm on this view.[10] Note, however, that Sider does *not* reject the existence of spacetime worms, he merely rejects that *we are* spacetime worms. How, then, if we are stages, is this a view of persistence?

Sider argues quite eloquently that persistence on the stage view happens in virtue of a counterpart theory of *de re* temporal predication. According to Sider's counterpart theory, objects persist by having distinct counterpart stages at every time at which they exist. So the utterance 'Ted was once a boy' is true just in case there exists some person stage x (prior to the utterance) that bears the temporal counterpart relation to the Ted-stage that currently exists.[11] Sider argues that even though these two stages (Ted the adult and Ted the boy) are

[8] Markosian, Ned. "A Defense of Presentism." *Oxford Studies in Metaphysics*, Volume 1 (Oxford: Oxford University Press, 2004): 47-82. 51.

[*] At least in the sense that time is spread out, and is analogous to space in that it has parts.

[**] Sider (2001) refers to perdurantism as 'four-dimensionalism' and Haslanger refers to persistence on the stage view as 'exdurantism.'

[9] Sider, Theodore. "All the World's a Stage." *Australian Journal of Philosophy* 74 (1996): 433-453. 433.

[10] Sider, Theodore. *Four-Dimensionalism: An Ontology of Persistence and Time*, (Oxford: Clarendon Press, 2001). 190.

[11] Ibid., 193.

distinct, it is still true that Ted *was* once a boy. The fact that Ted is no longer that exact same stage is not problematic for persistence on his view.*

Sider argues that the stage view solves many of the puzzles of persistence over time better than either the worm theory or three-dimensionalism.[12] For example, the stage view better explains how we count (namely, by *identity*). Sider states that "[c]ounting must be by identity when we count objects not in time (numbers, for example), and surely we count persons in the same sense in which we count numbers."[13] The idea here is that, if we count by identity, we must count person stages rather than worms. Whether we count stages or worms *can be* indeterminate, but typically it depends on the 'speaker's interests.' However, the stage view faces some difficulty when sentences include what Sider calls 'timeless counting.' To quote Sider: "sentences involving 'timeless counting' are ill-handled by temporal counterpart theory."[14] This difficulty arises because whenever statements involve *temporally extended predication* (for example: 'how many people were in your office *over the last hour*?') our reference in counting shifts from stages to worms. Thus, a concession is required for friends of the stage view, and Sider obliges with such a concession.[15]

Another type of puzzle that Sider argues the stage view solves better than the worm view or the three-dimensional view is that of the statue/lump. Because we typically quantify over stages rather than worms, the fact that there seem to exist both a statue and a lump in the same place at the same time is not so problematic.** Since the case entails two distinct spacetime worms, which share one or more stages in common, then, if we count stages instead of worms, the lump and statue never coincide. However, as Sider admits in the concession above, we *sometimes* quantify over spacetime worms. Thus, if one takes a 'timeless perspective' and adds temporally extended predication to a question or a sentence, such as 'how many statues and lumps were there *in the last minute?*,' Sider would be forced into a shift of reference, on his view, wherein he would count *worms* rather than stages. The result is problematic, in that it seems spacetime worms (which are objects on Sider's view) *sometimes* – albeit not typically – coincide. Indeed, Sider recognizes and explicitly discusses this very problem in an earlier essay, where he states:

---

* Sider's view (obviously) has more depth than how I have presented it here. However, since the focus of my paper is in critiquing and revising Brogaard's argument, I will not go into great depth regarding Sider's view. Although I will offer comparisons from my view to Sider's, for the purposes of this paper the description I've provided will suffice.

[12] Sider 1996, 433.

[13] Sider 2001, 189.

[14] Ibid., 197.

[15] Ibid.

** I use the phrase "not so problematic" rather than "not a problem" because, as I will argue, I think that *some* problem remains in Sider's version of the stage view.

> I do grant the existence of aggregates of stages, and such aggregates do sometimes coincide; but I deny that these aggregates are people, statues, coins, quarks, etc. Moreover, I deny that these objects are (typically) in the range of our quantifiers.[16]

So while Sider avoids the problem that *statues* and *lumps* coincide, his view still admits that *objects* sometimes coincide. Indeed, this is not a knockdown argument by any means; however, the problem of coinciding *objects* apparently remains on Sider's view.

Sider does recognize that the stage view faces some difficulties, as do most views, but he argues (and correctly so, I think) that its benefits vastly outweigh its costs. But if another view could handle these problems better, I think it would be worth considering.

3. STAGE VIEW PRESENTISM

While elucidating his view, Sider anticipates the emergence of a presentist version of the stage view. He argues that "a certain kind of presentist" could advocate the temporal counterpart theory of the stage view. But ultimately, this view would be much less attractive, considering that by endorsing such a view, the presentist would have to relinquish *de re* temporal predication[17] – which is one of the primary reasons the view is so appealing. Such a concession, Sider argues, is "obstacle enough" for the presentist. However, I argue that one can make a case for a presentist version of the stage view, *without* the use of a counterpart theoretic view of persistence.[*] But before expounding such a view, I must first lay further groundwork for the presentist stage view, so that its persistence conditions can be more salient once presented.

First, I argue for a presentist ontology which seems to differ from Brogaard's understanding of presentism. I argue that, in order to make proper sense of presentism, we must view all existing objects as being *wholly* present. Brogaard argues that existing objects are only *partially* present.[**] However, my ontology of existing things is one wherein 'exists' is predicated only of things which are wholly present at that time — otherwise, such a thing cannot be said to truly exist. When contrasting Brogaard's claim about partially existing

---

[16] Sider 1996, 15.

[17] Sider 2001, 208.

[*] For the detailed discussion of persistence on my view, see Section 4.3 of this paper.

[**] Note that I am not certain she would say that *all* objects are partially present. However, any object that is thought to persist cannot be wholly present on her view, given her claim that "no stage is wholly present at more than one time" (Brogaard 2000, 4).

objects with my claim about wholly existing objects, it becomes clear that my understanding of the presentist ontology differs deeply from the presentist ontology which her worm theory entails. For instance, on my view, no object can contain non-existent parts, whereas on Brogaard's view, this is simply how objects exist.

The fact that existing things must be wholly present on my view happens in virtue of the fact that any wholly present thing *is a stage*. Brogaard makes this same claim, but for different reasons. On my view, *objects are stages* and *vice versa* – and therefore, objects and stages are identical. Where Brogaard argues that objects are aggregates of stages, which are partially present at any given time, I argue that the object *is* the stage and that it is wholly present at any time at which it exists. However, it should be noted that my view of stages is *not* tantamount to Sider's, because the eternalist stage view posits that some objects only partially exist at a time — namely, spacetime worms. Therefore, on my view, aggregates of stages are *not* themselves objects.* This better accords with the presentist ontology, because, as the quote from Markosian in section one illustrates, presentism entails that there be no time except the present time. Thus, an aggregate of stages at different times would not necessarily entail the constitution of an object – which, remember, was Benovsky's primary charge against Brogaard's view.

Thus, both Sider and I agree with Brogaard when she states: "No stage is wholly present at more than one time; every stage is wholly present at exactly one time. There is a new stage for every moment at which a given thing exists."[18] But the three of us agree for fundamentally different reasons. Also, my reasons for affirming such a claim turn out to be opposite from Brogaard's. Brogaard makes this claim in an attempt to argue that objects are *never* wholly present, whereas I affirm the same claim to argue that objects are *always* wholly present. I also make the stronger claim that all objects are stages and all stages are objects — thus, aggregates of objects are not objects themselves. What then, on my view, do aggregates of stages constitute? *Events*. Where events remain ambiguous on Brogaard, Sider and David Lewis's views, I offer a precise description of the nature of events. (However, I will postpone such discussion until Section 4.2.)

To sum up: The Presentist Stage View is such that (i) there is no time except the present time, (ii) that only wholly present stages exist, and that (iii) everyday objects, such as persons, cats, cups, etc. *are* stages. If some thing is not wholly present, it cannot be said of that thing that it exists, nor that it is an object.

---

* For a more thorough treatment of why this is the case, see Sections 4.1 and 4.2 in this paper.

[18] Brogaard, 4.

Thus, on my view, 'exists' can only be predicated of *objects*, but not of events.*

So far, I have not yet provided a detailed description for objects or events. Indeed, the most important part of my view, *persistence*, has also been conspicuously absent. Now that I have laid much of the groundwork for the view, I will turn to the heart of it.

4. EXISTING, OCCURRING, AND PERSISTING

In this section, I will explicitly lay out the details my view. I will thoroughly examine the nature of *existing* objects and of *occurring* events. I will also elucidate how my view accounts for persistence *without* using temporal counterpart relations. If I am successful in this, the presentist stage view will avoid Sider's anticipated critiques and it will avoid the problems that Brogaard's view faces.

4.1 Existing

On the presentist stage view, 'existing' is predicated only of objects, which are *wholly* present at any given (present) time. Such objects, remember, *are* stages, and stages never exist as parts of other objects. Thus, objects and stages are synonymous on this view. If this were not the case, then *existing* things would necessarily have *non-existing* parts, which the presentist stage view rejects. This ontology of existing things (or objects) is not particularly controversial, so I feel this description will suffice. Formally put:

> *Existing things* = (i) actual, (ii) three-dimensional, (iii) spatially extended, (iv) objects, that are (v) wholly present (vi) at a time.

Such existing things are called 'stages' or 'objects.'

4.2. Occurring

'Occurring' on this view is something to be understood quite differently than 'existing.' 'Occurring' is predicated only of events, which are *partially* present at any given time, in virtue of having an existing thing as its part at any present time; an event must also be (partially) present at more than one

* This can be thought of similar to how Brentano and others argue for "different modes of being." See Brentano, Franz. *Philosophical Investigations on Space, Time, and the Continuum*. Trans. B. Smith. (London: Croom Helm, 1987) and McDaniel, Kris. "Ways of Being." *Metaphysics*. Ed. David Chalmers, David Manley, and Ryan Wasserman. (Oxford University Press, 2008).

successive time. Occurring, therefore, has a different 'mode of being' than existing, in that the ontology of occurring differs from the ontology of existing. Such entities can be understood as *successive aggregates of stages*, which have an ending at some time later than the time at which they began.* Formally put:

> *Occurring entities* = (i) non-actual, (ii) four-dimensional,** (iii) temporally extended (iv) events, that are (v) partially present (vi) at more than one successive time.

Such occurring entities are called 'worms' or 'events.'†

Prima facie, occurring entities seem much more controversial than existing things on the presentist stage view. Controversy arises primarily because of the first attribute ascribed to events: namely, that occurring entities are *non-actual*. But I argue that, because of the following two examples, this should not seem so counterintuitive. First, on the presentist view, events *cannot* be actual (if by *actual* we mean wholly present or existing or as containing no non-existent parts). Second, an event must not be viewed as something which spans *over* time, but rather as something that occurs *through* times.‡ An event occurs in virtue of having successive stages at many present times, which make up a part of that numerically same event. Thus, once the event no longer has an existing *part*, the event no longer occurs. Occurrence, then, depends upon existence for its being.

Although an event on this view is entirely made up of distinct stages at different times, it is not reducible to any one stage. When we quantify over events, we count *one* event, even though events are made up of many distinct parts. An event, then, has the property of ending at a later time than at which it began (or unfolding temporally) in virtue of each stage *at a time* existing as a part of that event, which, in a sense, can be understood as *pushing* time along.[19] So one event contains many parts – similar to how eternalists view spacetime worms, but it never actually exists, it only occurs. Thus, we have one stage existing *at a time* and one event occurring *through time* (or at many successive

* Again, the fact that only the present exists does not undermine the fact that there are events. This is because of the fact that events *occur* rather than *exist*. So if multiple stages exist successively and participate, say, in the same motion through space and time, then such a motion is understood as an occurrence, or an event, which begins at time t1 and ends at time t6. However, the motion *itself* cannot be said to 'exist' on this view.

** However, these are not four-dimensional in the standard sense. But rather, they are only four-dimensional in virtue of having begun at an earlier time, and where they are still occurring at a later (present) time, at which there exists a three-dimensional object as a part of the same event, which 'unfolds' or occurs *through* time. Also see the discussion later in this paper regarding Brogaard's 'temporally unfolding' fourth dimension to better understand how occurrences take up four dimensions rather than merely one dimension.

† Note that, like 'stage' and 'object,' 'worm' and 'event' are synonymous terms as well.

‡ Note the plurality of the word 'times.' As stated above, an event must be present at more than one *successive* time.

[19] Brentano, 101 and 215; Brogaard, 6.

times). I think Brogaard explains this quite well when formulating her view about how *objects* persist. Brogaard argues that only "stages of objects exist, but that objects have four dimensions in the sense that they have an unfolding temporal dimension in addition to the three spatial ones."[20] I argue, however, that if we apply this 'temporal unfolding' to events, rather than objects, it becomes much more plausible. Indeed, Brogaard also states that "events are commonly understood as having temporally extended parts even though these never exist as a whole but only through their successive stages." This explains events quite well. So I share Brogaard's intuition about how this fourth dimension 'unfolds,' but only in the sense where events, not objects, do the unfolding.*

Therefore, my view accounts for events quite nicely. Brogaard argues for the persistence of objects, but in a way that seems more fitting for the occurrence of events. Indeed, Brogaard never explicitly defines events on her view. And given that aggregates of stages constitute *objects*, events remain ambiguous. Sider's stage view and Lewis's worm view[21] seem to share this ambiguity about events as well. And another eternalist perdurantist, W. V. O. Quine, has stated:

> Physical objects, conceived thus four-dimensionally in space and time, are not to be distinguished from events, or, in the concrete sense of the term, processes. Each comprises simply the content, however heterogeneous, of some portion of space-time, however disconnected or gerrymandered.[22]

Thus, where eternalist views (and Brogaard's view) are lacking, my view provides a straightforward and distinct ontology of both objects *and* events, without leaving such things up to the imagination of the reader.

4.3. Persisting

On the presentist stage view, 'persisting' differs from 'existing' and 'occurring,' but only in that persistence is not reducible to either existing alone or occurring alone. Rather, persistence entails the occurrence of a continuous career-event, wherein an object exists at any present time. *Prima facie*, 'persisting' sounds like something that is synonymous with 'occurring.' I argue, however, that although objects do persist by occurring, that persistence

---

[20] Brogaard, 7.

* D. H. Mellor states: "All ... events take time, and none is wholly present at any one time. An instant of time indeed contains no part of such events; it merely separates temporal parts [or stages] of them, as an internal surface separates the spatial parts of a thing." "On Things and Causes in Spacetime." *The British Journal for Philosophy of Science* 31.3 (1980): 282-88. 283. Although he was arguing for Four Dimensionalism, this is relevant here.

[21] Lewis, Davis. *On the Plurality of Worlds*. (Oxford: Basil Blackwell, 1986).

[22] Quine, W.V.O. *Word and Object*. (Cambridge: MIT, 1960).

differs from occurrence in that persisting things are *wholly* present at any given time, whereas occurring things are only *partially* present at any given time. For instance, when we talk about what persists, we are talking about the cup or person itself, not just the event of which it is a part. I will elucidate this argument here.

Take 'Smith' for example (but note that the example could just as well be a cup). Smith is an object, and he is therefore wholly present at any given (present) time. But, he is also just a stage, so it seems as though he cannot persist through more than one instant. However, I argue that Smith not only exists, but he exists as part of one continuous life (or life-event).* Thus, the presently existing Smith has the property of *being an existing part of one continuous life-event*. The object that stood next to a computer just before sitting down to work on a paper, then, held the property of being an existing part of the same event to which the object that currently works on the paper belongs. They are both the existing part of the same life-event at any time they exist. Therefore, at any present time during Smith's continuous life-event, there is a wholly present object that holds the property of *being an existing part of one continuous life-event*. Both the object which stood by my desk and the object working on the paper have this same property. To further elucidate this point, I will examine this in two different ways: the first being what Sider calls a *typical* case, and the second being what Sider calls a *timeless* case.

The *typical* case pertains to Smith *at a time*. This is the typical way we view persons, according to Sider, in that we typically quantify over stages and not worms. So on the presentist view, this ostensibly present Smith is wholly present, and therefore exists. Thus, there is a stage which *is* Smith. The *timeless* case (or what I call a case which includes temporally extended predication) pertains to Smith 'over time' or at multiple successive times. Sider cites that in such cases, our reference shifts to quantify over worms rather than stages. Thus, on the presentist stage view, where worms (or aggregates of stages) are events, we then quantify over Smith's continuous life – which is still counted as *one*. So in the typical case we have a wholly present Smith, but what about in the timeless case? Once we regard Smith's continuous life on the presentist stage view, we recognize that such a life cannot be wholly present at a time for the same reason that a continuous war cannot be wholly present at one time. However, when regarding this continuous life-event (which ends at some time after the time at which it began), at any present time the event is instantiated, it is because of the fact that there is a wholly present Smith, which holds the property of *being an existing part of one continuous life-event*. Now, at a time

* One can substitute 'career (or career-event)' for an object, such as a cup, fork, etc.

five seconds later, there still exists an object Smith, which possesses that same property. Thus, to be straightforward in my claim:

> If *s* = a stage (Smith) at a time *t* and *w* = a worm (Smith's life) at multiple successive times; then *s* = a wholly present person and *w* = that person's continuous life; each *s* is the same type of *s* iff each *s* has the property of *being an existing part of the same numerical w*, and this is the case. So while each *s* remains distinct, they are *all* still a part of the same *w*, and since only one *s* ever *exists* at a time (and no two *s*'s ever exist at the same time), that *s* persists by existing at every time at which *w* occurs.

Thus, at any given time, Smith is wholly present, and his continuous life (or indeed any other continuous sub-event, such as reading for five seconds, running for ten minutes, etc.) occurs in virtue of having Smith exist at any present time it is occurring. Since there is no time except the present time, only one Smith ever exists: namely, this wholly present Smith. And since events occur at multiple successive times, only one life ever occurs. So, once again, when Brogaard states that

> According to presentist four-dimensionalism, then, to say that W. V. O. Quine exists (with tensed 'exists') must be characterized as loose talk. What we really mean is that an instantaneous stage of W. V. O. Quine exists.

We would agree (on the presentist stage view) with the second sentence, but the first sentence we would adamantly reject. For on the presentist stage view, when we say 'Quine exists' (in the tensed sense) we *literally* mean it. However, given that no such Quine presently exists, we would correctly say that no such Quine exists, nor does Quine's continuous life still occur, and hence, Quine no longer persists.

5. COUNTING, PUZZLES, AND AN OBJECTION

I will now briefly examine some of the implications of, and an objection to, this view.

5.1. Counting

Counting, on this view, happens similarly to how Sider presents counting on his view. That is, we typically quantify over stages (objects at a time), but sometimes our reference will shift and we will quantify over worms (which, on my view, are events), instead. But, I offer another reason for why this shift in reference happens, which stems from my view that objects and events have ontologically distinct modes of being, and which challenges Sider's view that referential shift typically depends on the 'speaker's interests.'

Sider argues that, when statements or questions include temporally extended predication (or 'timeless counting'), our reference shifts from counting stages to counting worms. And such referential shifts typically depend on the 'speaker's interests.'[23] What Sider means by 'typically' here entails that, if I ask somebody how many roads I must cross to get to the bakery – which happens to be across one road that winds back and forth six times between where I am and where it is – the appropriate response would be 'six roads,' even though there is only one continuous road. However, I argue that questions such as 'how many people were in Bob's office *over the past hour*?' should invoke different responses, since people certainly tend *not* to think of 'Bob' in terms of stages and aggregates of stages. If this were the case (that Bob is in fact a stage at a time or an aggregate of stages over the past hour), then people would remain entirely unaware of this fact and their 'interests' would not be a factor in the question. Therefore, to anchor such referential shifts in the *interests* of those whom remain unaware seems rather implausible.

I argue, on the other hand, that such a change in reference depends not upon the speaker's interests, but rather upon the fact that stage aggregates compose events instead of objects. If objects exist *at a time*, whereas events occur *through time* (or at multiple successive times), then the fact that we count stages when answering questions not in time, but count worms when answering questions which include temporally extended predication, makes much more sense. This does not depend on the speaker's interests, but rather, depends on the fact that *this is the way things are*, and our intuitions track such truths. Thus, whenever statements or questions contain temporally extended predication, we will typically conceptualize 'occurrences' (events or worms). But whenever questions or statements do include such predication, we count 'existents' (objects or stages). This is what causes the shift in reference between stages and worms – not the speaker's interests.

So, when someone asks me 'how many people are in Bob's office?' I will answer 'one' – because I will conceptualize an object, which contains mass and volume and which therefore fills up a spatial region. But if someone asks 'how many people were in Bob's office over the past hour?' I will be inclined to think

---

[23] Sider 2001, 189.

abstractly about Bob, which takes up many different possible spatial regions at different times, and I will answer 'one' in reference to Bob's worm (or event) which took place over the past hour. And since this does not depend upon my interests (which are hopefully conscious interests on Sider's view), I need not be aware of the fact that I am counting an event rather than an object. The fact that events are temporal, whereas objects are spatial, provides a good reason for why Sider claims that we count stages *at a time*, but count worms (which are events on my view) whenever we take such a 'timeless perspective.'

5.2. Puzzles

Here, I will give a very brief treatment of the puzzle of coincident entities – which, as noted above, caused a problematic consequence for Sider's view. But, I claim that my view can solve the statue/lump puzzle, while also avoiding the problem that Sider faces.

A sculptor gets a lump of clay on Thursday and molds it into a statue of David on Friday. Disappointed with his work, he destroys the statue of David on Monday, but the same lump of clay still exists thereafter. The statue and the lump hold different properties (the statue holds the property of coming into existence a day after the clay and going out of existence before the clay), and thus, they appear to be distinct objects. Do these two objects coincide on the stage view? Sider (mostly) avoids this problem, as we examined above, but is left with a view wherein we sometimes quantify over worms, and thus, worms (which are objects on his view) sometimes coincide – albeit never having persons, cats, etc. coincide. However, on my view since we *also* count stages rather than worms (at a time), the two never overlap. But, like Sider, once temporally extended predication is included in the statement or question, then our reference shifts and we count worms. But, on my view, *worms are events*. So, on the presentist stage view, when someone asks 'how many statues and lumps were there *in the last minute*?' our reference shifts to quantifying over *events*, rather than objects, but objects still *never* coincide — because worms are not objects.* The fact that events coincide is a fact on every ontology (as far as I know). I walk and think at the same time quite often, but this is not a problematic consequence of an ontology.

5.3. Objection

I have nearly exceeded the scope of this paper. But I will briefly address

* *Also note* that the statue and the lump differ only in temporal properties that pertain to events (on my view). This further shows that such 'worms' are merely events.

one objection.

One might object to my theory of persistence on the grounds that *it makes no sense that distinct objects can make up a continuous life-event, because this event would be composed of disjointed stages, which cannot bear relations to non-existent things.*

I will answer this in two ways. First, Sally Haslanger points out that she has the property of being the daughter of her parents, even though her parents don't presently exist. Thus she bears a relation to non-existent things.[24] They once existed, however, and she still exists, so this seems consistent to me. She was born to her parents and will always hold the property of being their daughter (for her entire life, anyway). Thus, there should be no problem for an existing stage to bear a relation to the same event, of which it is a part, and to which it came into existence in relation. Secondly, if one wishes to obstinately adhere to the claim that existing things can bear *no* relation whatsoever to non-existent things, then I would challenge such a person to explain why they view this presently existing stage to bear the property of *being distinct* from the previous stage(s). Distinctness, after all, is a relational property – one which, in this case, would bear relation to presently non-existent objects.

6. CONCLUSION

Ted Sider endorses an eternalist theory of the stage view and argues that it solves puzzles better than the worm view or three-dimensionalism – and he is right. Sider also admits the possibility of a presentist stage view, but with a caveat included. In this paper, I have advocated a presentist version of the stage view, which does not depend upon a temporal counterpart theory. The fact that I have presented such a view which both avoids Sider's admonitions and evades the most prominent problems that face Brogaard's view should attest to the appeal of stage view presentism. This view is also on par with Sider's view when it comes to solving puzzles, but it avoids the problems that his view faces. Thus, I conclude by claiming that stage view presentism is both plausible and appealing.

---

[24] Haslanger, Sally. "Persistence Through Time." *Oxford Handbook of Metaphysics*. Ed. Loux & Zimmerman. (Oxford: Oxford University Press, 2003): 315-54. 324.

BIBLIOGRAPHY

Benovsky, Jiri. "On Presentist Perdurantism." *Nordic Journal of Philosophy* 8:2 (2007): 79-88.

Brentano, Franz. *Philosophical Investigations on Space, Time, and the Continuum*. Trans. B. Smith. (London: Croom Helm, 1987).

Brogaard, Berit. "Presentist Four-Dimensionalism." *The Monist* 83:3 (2000): 341-356.

Haslanger, Sally. "Persistence Through Time." *Oxford Handbook of Metaphysics*. Ed. Loux & Zimmerman. (Oxford: Oxford University Press, 2003): 315-54. 324.

Lewis, Davis. *On the Plurality of Worlds*. (Oxford: Basil Blackwell, 1986).

Markosian, Ned. "A Defense of Presentism." *Oxford Studies in Metaphysics Vol. 1* (Oxford: Oxford University Press, 2004): 47-82.

McDaniel, Kris. "Ways of Being." *Metaphysics*. Ed. David Chalmers, David Manley, and Ryan Wasserman. (Oxford University Press, 2008).

Mellor, D.H. "On Things and Causes in Spacetime." *The British Journal for Philosophy of Science* 31:3 (1980): 282-88.

Quine, W.V.O. *Word and Object*. (Cambridge: MIT, 1960).

Sider, Theodore. "All the World's a Stage." *Australian Journal of Philosophy* 74 (1996): 433-453.

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

Consciousness as the Domain of a Computation

James Hodson
Princeton University

INTRODUCTION

Modern theories of mind and of cognition tend to revolve around the same basic question: how is it possible – given the assumption that we live in a world made of physical objects – for the brain to give rise to the rich conscious experiences that manifest themselves to us each and every waking or dreaming moment of each and every day? What is it about the physical brain that sets it apart from a piece of wood, or an oriental rug, or the computer that currently sits on my desk?

In conjunction with this question, we are led to ask the ever more pertinent-sounding follow-up: if the brain gives rise to conscious experience, then why not also all other physical objects – how do I know that my arm isn't currently having a direct conscious experience of its own? Any theory that addresses the question of consciousness, and its place in a physical universe, must offer a definitive answer to these questions and many more, in order to satisfy all aspects of this seemingly mind-boggling conundrum.

One theory that tries to address these issues is that of computationalism, whereby all that is required for any system to qualify as being conscious, is for that system to implement (run) an appropriate set of computations (the appropriate program). As long as we can map states of the system directly to states of the computation, then we can be justified in claiming that the system is in fact experiencing consciousness in exactly the same way as we would experience consciousness (if our brains were implementing the same computations). This theory has been discussed and developed at length by David Chalmers[1], and has received much attention from philosophers dissatisfied by the conclusions that may follow from it.[2] Because the computationalist is committed to viewing consciousness as essentially a by-product of a system, Maudlin argues, his theory is inadequate in explaining how a large mass of inactive components (as we assume must occur relatively often in the brain) — some of which may *never* become causally relevant – could possibly be necessary in order for consciousness to proceed from the system as a whole. If this were necessary, then the computationalist framework would be insufficient for explaining consciousness.

In this paper I will aim to show that Maudlin's attack on the computationalist theory need not cause any trouble to the view as a whole, and is in fact based upon a misunderstanding of the necessary and sufficient conditions for a system to be seen as implementing any given computation. However, Maudlin's

[1] Chalmers, David. "A Computational Foundation for the Study of Cognition." 1994. http://consc.net/papers/computation.html

[2] Maudlin, Tim. "Computation and Consciousness." *Journal of Philosophy* 86 (1989): 407-432; Searle, John. "Minds, Brains, and Programs." *Behavioural and Brain Sciences* 3 (1980): 417-424.

argument serves as a useful starting point for exploring possible weaknesses of the computationalist view, and therefore allows for the development of an even more robust theory.

This paper will be sub-divided into three further sections. The first of these sections will lay the foundations for the computationalist view, and expose each of its underlying assumptions to common criticisms. Firstly, I will clearly define the terms that are used when explicating the theory, and then show that if we did not take *the implementation of a particular computation* as being the necessary and sufficient condition for consciousness in a physical system, then it is difficult to see how we could proceed whilst maintaining our primary assumption that the universe is purely physical in nature.

The second section will be dedicated to a defense of the computationalist view from the argument that Maudlin gives in "Computation and Consciousness." The defense takes the form of a thought experiment designed to show that the computationalist is not bound to Maudlin's original implicit assumption, namely that consciousness is derived from an entire physical system that is implementing a computation in a sub-part of itself at any given time. These terms will be explained more fully in what follows.

Finally, the third section of this paper will conclude that although Maudlin's argument is misplaced, the computationalist theory is still not completely safe from potential attack. I will propose other areas that need restating, and possible starting points for such developments. Having said that, we will see that the computationalist approach remains the most hopeful for answering those questions originally posed, and that any modifications that may be required will not be so groundbreaking.

I

In order to understand the computationalist view it is helpful to clarify exactly what is meant by a *computation*. A computation is simply a high-level description of the steps required in order to transform a given range of inputs of a certain kind to an output of a certain kind. Usually, computations define States of a physical system, and the kinds of connections between those States (transitions) that are required in order for the system to achieve this goal. In short, a computation is an abstraction of a physical system that describes its workings in a formal manner. In order to implement a computation, then, it is necessary for each of the physical States and State-transitions of a system X to map correspondingly (on a 1:1 basis) to States and State-transitions of the computation Y that it is implementing. For a computationalist, if such a mapping exists, then we are able to say that system X implements computation Y. In this context, the abstraction of a Turing Machine is often seen as helpful in grasping the idea, so this is where we will begin.

A Turing Machine is a useful abstraction to have for this discussion, because it is very simple to understand, with few rules that govern its functioning, and yet provides the same computational power as much more complicated systems. Any finite computable function may be computed using a Turing Machine. It follows then, that for a computationalist, showing that States and State-transitions of a computation can be mapped to States and State-transitions of a given Turing Machine will provide a good basic model for a theory of implementation.

Essentially, a Turing Machine works as follows. A tape that is infinitely long in both directions, and sub-divided into an infinite number of squares of equal size, contains the input data. Any individual square along the tape can contain either a stroke (denoted by 1) or be empty (denoted by 0). Next, the Turing Machine has a head that can move sequentially back and forth along the tape, and perform one of three actions after reading a particular datum from a square on the tape: do nothing, write a 1 in place of what is currently in the square, or delete what is currently in the square in order to leave it blank (0). The action that the machine performs is determined by its final component, the State table. Each State tells the machine to perform a single action, and provides the next State that the machine should move to. Explicitly, the State may tell the machine to move one square to the left or right of the square that it is currently on, or read the square that it is currently on and branch (conditionally) to the next instruction based upon what has been read. For example, the compound statement, 'if (0) move right, else move left,' might translate to:

> State 1:
> Read the tape;
> If the Symbol is a 0, then move right and go to State 2;
> Otherwise, move left and go to State 3.

This simple instruction set is powerful enough to allow us to implement a wide variety of computations, and can be expanded in various ways to add explanatory power — although its computational power will not increase. That is to say, a Turing Machine can implement the same set of sequential algorithms as any modern computer theoretically could (ignoring time and space constraints).

With this in mind, let us turn now to the central tenet of computationalist thought, the thesis of *Computational Sufficiency*. Put simply, this states that having the appropriate computational structure, by itself, is enough to qualify as possessing a mind and a wide variety of mental properties. As long as a system is running the correct *program* (or, as we have been discussing, implementing an appropriate computation), then at any point within that *program* it will have the same conscious states as any other system that possesses a mind. This means that all of our conscious experience is a product of the computation that (presumably) our brain is instantiating – nothing else is required in order to create the wide variety of experiences we have as human beings.

In addition, a second thesis, that of *Computational Explanation*, flows directly out of that of *Computational Sufficiency*. In general, if a computation suffices to give rise to these phenomena (a mind), then *necessarily* that which we associate with consciousness – behaviour and cognitive processes – will be derived from the computation being implemented, and nothing more. This thesis is strongly linked to the assumption a computationalist must make, that it does not make a difference what *kind* of matter is involved in the physical implementation of a computation. Neurons, or silicon, as will be later explained, are fully equivalent and produce the same implementation of a computation both qualitatively and quantitatively.

In his paper, "A Computational Foundation for the Study of Cognition," Chalmers puts forward a computationalist theory based on a slightly enhanced version of a Turing Machine, and addresses several possible objections to the general computationalist position. I will explicate the theory that Chalmers puts forward, along with insights from another paper of his, "Absent Qualia, Fading Qualia, Dancing Qualia,"[3] which addresses the reasons for believing that the implementation of a computation is in fact the best way to frame our problem.

Chalmers (1994) attempts to explain why Computation rather than any other discipline is the most valid and fruitful approach to help us understand the mind and cognitive properties. He believes that although a Turing Machine is powerful enough to represent any computation, a more intuitive machine is required for the purpose of bridging the gap between our conception of consciousness and our conception of a computation. The machine that he introduces is the *Combinatorial State Automaton* (CSA), which essentially acts as a Turing Machine that instead of being in only one state at a time, participates in an over-all State that is defined by a State-vector. Therefore, rather than being in State S_1 for example, a given machine will be in State $\{S_1, S_2, S_3\}$. This, Chalmers believes, allows for a clearer mapping from States of a computation, to the States of a physical system.[4]

Chalmers describes implementation using the terms that we defined

[3] Chalmers, David. "Absent Qualia, Fading Qualia, Dancing Qualia." *Conscious Experience*. Ed. Thomas Metzinger. (Paderborn: Imprint Academic, 1995): 309-330.

[4] For a further discussion of the adequacy/inadequacy of the CSA model proposed by Chalmers, see: Brown, Curtis. "Implementation and Indeterminacy." *Canberra Conferences in Research and Practice in Information Technology* 37 (2004).

earlier, whereby a system can be seen as implementing a computation when there is a direct mapping from physical states of that system, to states of the computation. The next step in piecing together the computationalist theory is addressing how computation relates to cognition — that is, how we get from the implementation of a computation, to any appropriate physical system actually having conscious experience in a similar way to humans.

Firstly, Chalmers states that a behaviouristic approach to this explanation would be insufficient. A behaviouristic approach would be to see how a system behaves when given certain input (stimuli), and concluding that two systems are isomorphic if the behaviour displayed by them is the same in all respects. Chalmers believes such an explanation to be insufficient because we can easily envisage two distinct computations giving exactly the same output. To make this point even clearer, let us imagine two systems that compute the product of three numbers: a, b, and c. The first system multiplies a by b, and then multiplies that total by c; the second system adds a to itself b*c times. Although the two systems invariably produce the same output, they have very different internal processes, different States that they can enter, and a different order in which they are entered. Another problem with the behaviouristic approach is that depending on the input we give to two systems, we might end up classifying two systems as equal when in fact they only happened to overlap in behaviour for the given input. So, such a method is insufficient in that it only concentrates on the overall output of a system. What is needed is a stronger link to each of the sub-computations occurring — that is, the causal organisation. Chalmers asserts that the invariance of support for cognitive function (conscious experience) is maintained by any system that can be seen as the original system with any number of specific transformations applied – moving, distorting/stretching, sufficiently localised re-circuiting, and any other changes that do not alter *causal* interaction. Mental processes, that is, cognitive processes, are organisational invariants according to the computationalist view. They depend entirely upon the causal structure of the system – whereas many other properties do not (flying, for instance, as the causal organisation of something that flies is retained even when not flying, whilst the property of flying is not).

But why ought we to think that cognitive processes are organisational invariants rather than physical invariants (i.e. dependant entirely upon the material in which they are instantiated)? Chalmers offers us several reasons in his paper, "Absent Qualia, Fading Qualia, Dancing Qualia."

If the causal organisation of a system is all that is required for it to evince cognitive properties, then it is difficult to see why anything more could be required to produce conscious experience (such as the material from which its constituent parts are crafted). Chalmers attempts to show the empirical improbability of envisaging such dependence. The thought experiments referred to are those of ***absent qualia, fading qualia, inverted qualia,*** and ***dancing qualia***.

A case of absent qualia* occurs when any introduction of a non-neuronal pathway in the critical conscious sections of a system induce an entire absence of conscious experience in the system – although cognitive processes continue as per usual. That is to say, that we take a fully functioning brain, and replace one of the millions of neurons controlling sight perception with a silicon-based duplicate of that neuron (imagine that we have invented a procedure that *allows* us to do this without interrupting any of the brain's functions). It seems highly unlikely that replacing one neuronal link among billions with a silicon link would cause an end to conscious experience. Bearing in mind that the particular neuron being replaced is perhaps not even directly related to other parts of our conscious life, and that the silicon neuron performs in exactly the same way as the original, such a small change seems rather unlikely to cause a complete loss of our conscious experience. There would need to be something extremely special about our chemical-biological make-up in order for this to occur, and what's more, this still would not solve the question of what is *sufficient* for conscious experience to arise in a system (for if it did, then we would be forced to say that anything that is made from neurons would have conscious experience, regardless of the causal connections between those neurons).

One temporary way around this problem is to accept the thesis of *fading qualia*, which states that the more neurons are replaced, the less vivid our experience becomes. In this way, there is no magic moment at which consciousness ends, but rather, certain conscious experiences will fade out sooner than others, eventually leaving no conscious experience at all when the system has been completely replaced by silicon neurons. Of course, regardless of whether or not qualia are fading from the system, it will in no way affect the cognitive processes occurring in the brain – as these *are* a product of organisational invariance, not of conscious experience in and of itself. For example, even if the faded qualia patient does not experience vivid red when he sees a red car, if asked he will state that the experience he is having of red is very vivid, as vivid as ever. This is due to the fact that neurons send each other signals in order to effectuate actions, a purely causal relationship – regardless of whether or not there is *feeling* attached to them.

Finally Chalmers considers the case where the replacement of neurons with silicon causes localised qualia inversions — i.e. instead of experiencing the

* *Qualia* are the qualitative experiences associated with our conscious operation — such as what the colour *Red* looks like to us, or how *smoked salmon* tastes. These individual experiences are wholly subjective, and are only open to introspection.

same red when seeing red, we experience what is by all other accounts blue. Chalmers shows us this case because it draws together the idea of the causal layout of the system determining experience, and a refutation of the thought that qualia could change for any other reasons except the causal structure (the implementation of the appropriate computation). The experiment is simple: we replace a portion of the brain large enough to have a significant experiential difference on a specific variable (such as the colour red), with a silicon system that has the same causal structure.* We will keep the portion as small as possible to avoid too much conscious experience changing (we might infer that we had given rise to a different person if in fact the causal structure were not the only thing at play). We could imagine placing these two systems (neural and silicon) in parallel, with a switch between them. Each time we flick the switch the experience changes from red to blue, and back again — and yet, because of the organisational invariant, the person will not display any different behaviour to normal, or manifest any peculiar behaviour – their cognitive processing will continue as normal even though their experience of red continues to switch back and forth from red to blue (and they would not even express this difference if asked!). If this were possible, Chalmers argues, then psychology and phenomenology would be radically out of step – appearing to have very little influence on each other. Furthermore, this could lead to a situation in which qualia is continuously shifting regardless of whether or not neurons are replaced by silicon chips — simply from the constant molecular-level changes occurring in the brain — and we would never be aware of it!

These hypotheses seem to Chalmers to be empirically improbable — as the idea of *unconscious* qualia inversions is absurd, since qualia are *conscious* manifestations. The most plausible conclusion would be that no change in qualia would result from any of the experiments. Thus, it seems that the most rational explanation of conscious experience is as an organisational invariant.

Chalmers (1994) provides responses to possible objections to the computationalist view that we have outlined. One of the most famous of these objections is a thought-experiment proposed by John Searle. Searle proposes an experiment involving a man inside a room that has an input slot, and an output slot. The man is given a set of instructions regarding how to manipulate formal symbols that he receives through the in-slot, in order to provide a set of different symbols through the out-slot. The man inside the box has no idea what the symbols mean, he just follows the set of rules in order to produce output. Unbeknownst to him, people outside the room use the in-slot and out-slot in order to post questions in Chinese, and receive coherent Chinese

* It is important for the systems to have the same causal structure, and not just the same input/output – because the high-level implementation details ***are*** important, as noted in the ***product of three numbers*** experiment.

answers from the room. As time passes, the man inside the room becomes so good at the symbol manipulations, that to everybody outside, the answers are indistinguishable from those of a native speaker of Chinese. Yet, Searle says, the man cannot be said to *understand* Chinese – after all, he is simply using the given instructions in order to produce output that is based on the input; the symbols have no meaning for him. Furthermore, the system cannot be said to understand Chinese either, because we could imagine the man not being inside a room, and internalising all of the rules so that they come naturally to him – the output would be the same as a native speaker, but he would not know a word of Chinese!

Searle's argument is supposed to show that the room is exactly like a computer implementing a program that 'understands' Chinese in the behaviouristic sense, and yet it is inaccurate to say that the room understands Chinese in the same way that the man understands English. This is because, although the formal structure might be the same, the system lacks the semantics required in order for it to understand the content it is processing. Therefore, the implementation of a computation alone cannot suffice for the instantiation of a mind.

Chalmers' response is to say that although the man does not understand Chinese per se, the system as a whole inevitably does. If the computation being implemented by the man in the room is equivalent to the computation being implemented by a native speaker of Chinese, then we are quite justified in saying that the system understands Chinese in exactly the same way as a native speaker understands Chinese.

Searle has a response to this, which is to imagine that the man internalises all of the instructions to the point that the room and the instruction set are no longer required. Then we are left with a man who knows all there is to know about the implementation of a system, and yet does not understand Chinese – despite appearing to everyone else as if he does. This again invites the response of "if the computations are equivalent, then the understanding derived would also be equivalent."

Instead, let us try to look beyond this latter response, as to where the divergence lies between these two views, and see if it cannot be reconciled. Searle's objection is that the semantics provide an obstruction between simply following rules, and actually understanding Chinese. As Searle states that machines do not have semantics, but brains do, then this at first appears to be an insurmountable barrier. However, if we were to properly analyse the computation occurring in the brain of a native Chinese speaker, we would be able to see that the knowledge of each symbol is grounded in a representation of that symbol's meaning in the world. For example, the symbol for a table is linked to whatever idea is evoked in the native Chinese speaker's mind when he/she hears the word 'table' uttered. In this way, if the computation that is being implemented by the man in the room did not somehow incorporate the elements of understanding that link the word 'table' to the idea of a table in the mind of a native Chinese speaker, then of course the man in the room would not understand Chinese. However, if the computation were to incorporate these aspects, it seems harder to object that the man-in-the-room system would not understand Chinese.

So, to recapitulate, the reason why Searle's thought-experiment seems to show that the system would not understand Chinese, is because it purposefully does not take into account the fact that the native Chinese speaker's computation is tied to the non-linguistic representation of the symbols (as an interwoven part of the computation). Instead, the man in the room has no instructions relating the symbols to their representations, and is therefore not implementing the appropriate computation for the understanding of language.

Now that we have laid down the foundation for considering issues relating to the computationalist's framework, we can move forward to consider an objection that was not dealt with at all in Chalmers' papers — that of Maudlin.

II

Maudlin argues that the mere implementation of a computation cannot suffice for consciousness. This intuition is based on the fact that we can imagine a complex system to which a computationalist would assign mentality (consciousness), then consider only a small task carried out by a tiny subsection of this system (the active part). If we isolate this subsystem then the computationalist is supposed to assert that this new subsystem would have no conscious experience (as the conscious experience is a product of the entire system). But then, Maudlin argues, how could the existence of consciousness depend upon the presence or absence of a huge mass of inert matter being present and connected to the subsystem? Surely, if there is no conscious experience in the subsystem, having all of this excess paraphernalia will not help one bit to create it.

The intuition here could be described in a different way. A computationalist would perhaps argue that the system as a whole is conscious, rather than say that the sub-part of the system that handles vision is conscious. Then, if we were to isolate the system that handles vision, a computationalist would certainly deny the idea that such a subsystem could be conscious. The reason for this is that a computationalist sees consciousness related not to localised events in the system, but rather to the way those local subsystems interact with the parts of the system that serve consciousness. That is to say, if all of the system that is required for consciousness is present in the subsystem at any time, then that subsystem will be conscious, otherwise it will not be. It is also important to note that the part of the system responsible for consciousness would be (one would think) always active.

In order to illustrate these points more clearly I propose a thought experiment as follows: Let us imagine a human brain, made of billions of neurons. Scientists have recently developed an extremely hi-tech procedure which allows neurons to be connected and disconnected without affecting their individual health — therefore, when a neuron is not firing, or being fired at, it can be removed without losing its potential. We could imagine that such a technique could be used for maintenance purposes in order to extend the lives of future generations. The way this procedure works is that it removes the neuron when it is playing no causal role in the system, fixes it up, and then puts it back in just before it is needed again.* Under Chalmers' view of computation, the causal structure of the system has not been affected when we remove and replace these neurons. Now, further imagine that this procedure is highly efficient, and is capable of removing many neurons from the brain at once — only those that are playing no causal role (which according to Maudlin's example is the majority of the brain). Causally, the brain works exactly as it does at any other time, and certainly no computationalist with Chalmers' view of implementation need admit that at any moment this system has conscious experience that differs from the full system when not being maintained. Finally, we can imagine this same machine being used to perform Maudlin's experiment, removing so many neurons at a single time that only those related to a tiny subsection remain in place. As we have just shown, this does not affect the causal structure of the system, and therefore need not be regarded as a threat to the computationalist view of consciousness as being directly linked to the implementation of the correct computation.

One of the side effects of Maudlin's claim is that, if true, conscious systems could pop into and out of existence all of the time, simply by momentarily having the right state transitions. The computationalist response to this claim is that, yes, if an appropriately structured system were to momentarily have the correct state transitions, then it would momentarily be conscious. However, it is important to note here that it is highly unlikely that such a large number of inactive neurons would ever exist – as any system implementing a conscious being would most

* This is a rather clever procedure, which is able to keep track of the current firings of all neurons in the brain, and of the shortest time period possible before relevant neurons could fire, and there could be a need for the neurons out for maintenance. Counterfactuals of the form: "if the stimulus were suddenly changed...", cannot be applied here, as 1) this would place the behaviour of the procedure outside of its causal structure, 2) our clever procedure would keep track of such a change and modify its actions accordingly.

probably be continuously active across large portions asymmetrically. So, the chances of such a complex system fleeting into and out of existence in nature is rather low — though certainly possible.

III

Now that we have shown Maudlin's argument to not be such a dangerous threat to the computationalist view after all, we can move on to consider whether our formulation is strong enough to withstand other possible attacks. These attacks are aimed not at whether or not implementation is sufficient for consciousness, but rather, at whether the theory is capable of expressing the relationship between implementation and consciousness adequately. If our computationalist theory were found to be lacking in explanatory power, then it would need to be restated, which may be just as detrimental as an attack to the central tenets of the theory. I will not offer a restatement of the theory here, but I will provide insights into possible solutions to the problems about to be posed.

The first of these explanatory gaps in the computationalist theory is a product of the computational model used pervasively — that of the Turing Machine. As described at the beginning of Section I of this paper, the Turing Machine is capable of computing any computable function, and as such would be certainly capable of carrying out any computation that modelled the conscious mind. However, being powerful enough to compute a given function does not suffice in and of itself. Implementation requires not only the power to compute, but also, a direct mapping of States of the computation, to states of the physical system involved. Such a direct mapping appears at first to be a simple matter of spatial and causal correspondence, but in fact, a third correspondence is most certainly also relevant — temporal. In order for a system to count as implementing a computation, we ought to require of it that State transitions throughout the system occur in parallel with State transitions in the computation. A Turing Machine can be in only one state at a time, whilst a mind is most probably in more than one state at any given time. Chalmers offers us the Combinatorial State Automaton as a replacement for the Turing Machine — it is capable of being in several States at once, by virtue of a State vector representing the current overall State of the machine. The problem, nonetheless, is not that of an inability to represent enough States, but rather, that of only being able to represent the *synchronous* execution of a multi-state computation — whereas a mind does not have *synchronous* limitations. Conscious experience is the conjunction of all of the sub-computations occurring within a conscious system. If these sub-computations do not overlap appropriately causally and temporally, then the system would have very disjointed experiences at best.

Therefore, computationalism needs a better model of computation if it is to offer us a way of examining its claims fully.

A further, and final note is meant to frame future objections to computationalism in a manner that does not ignore the practical requirements of a system that implements consciousness. Namely, the fact that such a system would be in constant flux over the majority of its physical domain. Actions do not happen spontaneously — the build-up is a constant process, and consciousness/understanding cannot be understood independently of the whole. For example, Searle's Chinese room argument offers convincing anecdotal evidence that the system involved does not understand Chinese, but not in any interesting sense. If the room were equipped with the necessary apparatus for understanding Chinese, then it *would* understand Chinese. When objecting to computationalism, it is not possible to break the problem of consciousness down into smaller chunks — it must be attacked as a concerted whole. Seen in this way, the theory is perhaps stronger than first perceived.

In conclusion, although the Turing Machine by itself may not be a powerful enough model of computation, there are several ways of envisaging a system that runs asynchronously and is capable of supporting the localised and variable development of sub-computations. This may even be achieved through an array of slightly modified Turing Machines, but I will not go into the details of such an arrangement. When taken in the way described in this paper, the computationalist stance offers firm responses to all criticisms laid at its door to date. The task now is to expand our understanding of the conditions of a physical system capable of implementing a conscious being, and ensure that the framework developed permits these.

BIBLIOGRAPHY

Brown, Curtis. "Implementation and Indeterminacy." *Canberra Conferences in Research and Practice in Information Technology 37* (2004).

Chalmers, David. "A Computational Foundation for the Study of Cognition." 1994. http://consc.net/papers/computation.html.

Chalmers, David. "Absent Qualia, Fading Qualia, Dancing Qualia." *Conscious Experience*. Ed. Thomas Metzinger. (Paderborn: Imprint Academic, 1995): 309-330.

Maudlin, Tim. "Computation and Consciousness." *Journal of Philosophy* 86 (1989): 407-432.

Searle, John. "Minds, Brains, and Programs." *Behavioural and Brain Sciences* 3 (1980): 417-424.

ACKNOWLEDGEMENT:

Thanks to Professor John Burgess (Princeton University) for help, advice, comments and editing throughout the drafting and writing of this paper; and to Anastassia Fedyk (Princeton University) for thoroughly proofreading my originally ungrammatical text.