The Yale PHILOSOPHY REVIEW

An Undergraduate Publication

ISSUE 2 | 2006

Issue II, Spring 2006

Email the editors at questions@yalephilosophy.fastmail.fm
For our online edition and submission guidelines, visit <www.yale.edu/ypr>.

EDITORS' NOTE

"This is patently absurd; but whoever wishes to become a philosopher must learn not to be frightened by absurdities."

Bertrand Russell

The editors of the *Yale Philosophy Review* take Russell's advice to heart—from debates about three versus four-dimensionality to discussions of what makes an event an event, we are never afraid of the strange, the esoteric, or the absurd. For, as fellow philosophers know well, the path to truth is marked by paradoxical road signs, infinite detours, and a constantly shifting landscape. The true philosopher, as Russell reminds us, is not one who shies away from these hurdles but one who embraces them. The journey, as the cliché goes, is the most important part.

With this edition of the *YPR*, we hope that you, the reader, will have another opportunity to explore the absurd and yet urgently important pursuit that is philosophy. This volume of the *YPR* is our second, and we are very proud of it, and of the growing institution that helped create it. Since last year, our ranks have filled, our submissions multiplied, and our love of absurdity deepened. The result is a journal that reflects some of the best work by undergraduate philosophers worldwide.

Inside this volume, you will find an array of philosophic investigations, from metaphysics to ethics to the history of ideas. Our chosen five essays, however, all share one trait: a willingness to explore, to suggest the absurd, and to play with our common conceptions. And finally, our interview with Richard Rorty celebrates the central role of intellectual curiosity and playfulness in the highest of philosophic pursuits.

We do hope you enjoy this second volume of the *Yale Philosophy Review*!

With Fondness and Absurdity,

James Martin & Amia P. Srinivasan
Editors-in-Chief

THE Yale Philosophy Review
Issue II, Spring 2006

Objects, Worms, and Slices in 3 and 4D

Daniel Koffler
Yale University

According to one theory of persistence, objects persist through change over time in virtue of being wholly present at all points in time at which they exist. According to another theory, objects are spacetime worms that persist in virtue of having temporal parts at all points at which they exist. According to a third theory, objects, strictly speaking, are momentary time slices, and ordinary medium-sized dry goods are sums of time slices related by a temporal counterpart relation. Whole object theory, parthood theory, and counterpart theory are theories of extension (in time, in this case, but they can also be applied to ordinary space and modal space). These theories are distinct from three-dimensionalism and four-dimensionalism, which do not explain how objects are extended in a domain, but rather the number and kind of dimensions an object's extension occupies. Each theory of extension is consistent with exactly one of 3Dism and 4Dism, and 4Dism is vastly more plausible than 3Dism independent of any considerations of persistence and extension. Therefore it remains to weigh those theories of extension consistent with 4Dism against one another. Slice theory comes with an ontological price that worm theory is free of: Slice theory rises or falls with an orthogonal metaphysical principle, i.e. unrestricted mereological composition, whereas worm theory is on equally good footing given restricted or unrestricted composition. On the other hand, worm theory bears a semantic price that slice theory avoids: Worm theory is in a difficult position vis-a-vis the semantics of counting and the problem of coincidence. Koffler argues that worm theory can survive its semantic cost but slice theory cannot survive its ontological costs.

The philosophical literature on persistence and extension has produced three conceptions of how entities in space, time and modal space can be extended: whole-presence theory (entension), part-theory (pertension), and

counterpart theory (obtension).[1] Theodore Sider suggests that the problem of coincidence, i.e., the problem of objects with overlapping parts, might motivate one to adopt temporal obtensionism just in case one is committed to four-dimensionalism, because (a) the problem—if it really is a problem (as I will argue it is not subsequently)—rules out temporal pertension and (b) temporal entension entails three-dimensionalism. Hence, the committed four-dimensionalist who deems the problem of coincidence a fatal objection to temporal pertension may not adopt temporal entension (without giving up four-dimensionalism), and thus must develop an alternative view of temporal extension that is consistent with four-dimensionalism.

Something is missing, and that is a definition of three- and four-dimensionalism. There are, of course, the traditional accounts of both terms: three-dimensionalism is endurantism, and four-dimensionalism is perdurantism. The addition of temporal obtension throws a wrench into this framework, however. Claiming that temporal obtension and temporal pertension are alternative forms of perdurantism merely redefines perdurance as the disjunction of part-theoretic and counterpart-theoretic temporal extension, and hence begs the question of what property temporal pertension and temporal obtension have in common that temporal entension does not have. What the introduction of temporal counterpart theory does to the traditional accounts of endurance and perdurance, in other words, is to restructure them as disjunctions of two different sets, one of them a set with one member, temporal entension, the other a set with two members, temporal pertension and temporal obtension.

These sets cry out for an explanatorily rich definition, a principle behind the dichotomy of temporal entension on one side and temporal pertension and obtension on the other. Let perdurance continue to have its tra-

[1] The terms "entension" and "pertension" are introduced by Josh Parsons (2003) to refer to the spatial analogues of endurance and perdurance. I borrow Parsons's terms, but in my usage they are neutral to the domain in which an entity is located. Entension and pertension can thus be spatial, temporal, or modal. Additionally, the term "obtension," though an original coinage, refers to the form of extension described by counterpart theory, whether modal (as in Lewis), temporal (as in Sider), or spatial (a position that has yet to be advocated, but has a prima facie resemblance to compositional nihilism).

ditional meaning, namely, nothing besides part-theoretic persistence. Call counterpart-theoretic persistence "obdurance" if it needs a name. In what way are perdurance and obdurance alike, and endurance unlike either? Perdurance and obdurance are both consistent with four-dimensionalism (4Dism) and inconsistent with three-dimensionalism (3Dism), whereas endurance is consistent with three-dimensionalism and inconsistent with four-dimension-alism; alternatively, endurance and 3Dism entail one another, and perdurance and obdurance entail 4Dism while 4Dism entails one of the two.

So we are back at square one: what are the definitions of 3Dism and 4Dism? But what the preceding discussion has made clear (I hope) is that if these definitions are to account for the common membership of temporal pertension and obtension in one set, they have to be free of reference to endurance and perdurance. That such definitions can be given becomes clear once we recognize that the traditional conceptions of 3Dism and 4Dism are conceptions of the predication of objecthood in spacetime, whereas perdurance and endurance are conceptions of form-domain pairs (i.e., temporal pertension and temporal entension, respectively). The question of how to decide what in an ontology is to be ascribed the predicate "is an object" is conceptually distinct from the question of how an object's extension in a domain is formed. It is, in other words, an a posteriori truth that 3Dism entails temporal entension, and 4Dism entails either temporal pertension or obtension, if it is a truth at all. Only when 3Dism and 4Dism are considered independently will the reason for these entailments emerge; 3Dist temporal pertension is not a contradiction, nor is 4Dist temporal entension.

Fortunately for the purposes of non-circularly defining 3Dism and 4Dism, the literature on these concepts provides insight into what the terms mean purely as theories of the predication of objecthood and not as conjoint theories of objecthood and theories of persistence. Hidden within an inessential aside in a paper on composition, Peter van Inwagen, an unambiguous 3Dist, provides the rudiments of 3Dism as a theory of objects:

> I will remark that in the remainder of this paper I am going to be making a controversial assumption about material ob-

> jects: that material objects are three-dimensional and strictly persist through time. I assume, for example, that a cat is a three-dimensional object (and not a "space-time worm," whatever that is). And I assume that the three-dimensional cat I took to the vet last September is numerically identical with the three-dimensional cat I stroked last week. (I shall be told that the cat I took to the vet and the cat I stroked had different properties and hence cannot have been numerically identical. I reply that *it* had different properties at different times.) I oppose this view to the view that *two* three-dimensional cats—or perhaps I am supposed to call them "cat-slices"—figure in my history, once occupying a point in time last September and the other occupying a point in time last week, these numerically distinct three-dimensional objects having no more intimate connection than that established by their both being slices of one four-dimensional object. (van Inwagen 1987: 25)

Notice: one and the same three-dimensional cat is present last September and last week. There is a fourth dimension, time, but it is merely a dimension in which the three-dimensional object is located; if time is necessarily pointed in one direction, or more strongly, if the correct account of time is A-theoretic,[2] three-dimensional objects move in time, from the past towards the future. Three dimensions, in other words, define the boundaries of the object. The remaining dimension is related to it in the same way that a river is related to a ship sailing on it. So strict 3Dism, independent of any views about the nature of the domains and forms of extension, can be formalized this way:

[2]Briefly, A-theory states that every point in time is either past, present, or future, presentness being a property continuously shifting from being had by one time to the being had by the next in the unidirectional series of points in the temporal dimension, the movement of presentness being the cause of the experienced phenomenon of the progression of time. B-theory states that there is no absolute past, present, or future, only "earlier than" and "later than" relations, of which each time bears exactly one to every other time and neither to itself. The dichotomy originates in J.M.E. McTaggart's "The Unreality of Time." Mind 17 (1908): 457-474.

> 3Dism: "x is an object" is true iff. the boundaries of x are completely contained within three dimensions, no matter how many other dimensions there are besides those three.

According to strict 3Dism, an object could have boundaries in two of the dimensions of space as well as the one dimension of time, and float along the river of the remaining spatial dimension. No 3Dist, of course, believes that there are any objects that fit such a description, and because of other views they hold, such objects are impossible. But prior to adopting any views apart from 3Dism, the fact that all the objects are bounded in space and floating in time is an epistemically contingent fact.

The strict meaning of the opposing view, 4Dism, should now be easy enough to predict. And indeed, there are ample 4Dist accounts of objecthood that make it a relatively uncomplicated task to prize the concept of 4D objects out of any coincident views on extension and perdurance. In his response to Judith Jarvis Thomson's famous "crazy metaphysic" objection to temporal parts theory,[3] Mark Heller provides exactly the sort of account we are looking for:

> I propose that a physical object is not an enduring spatial hunk of matter, but is, rather, a spatiotemporal hunk of matter as filling up regions of spacetime. A physical object is the material content of a region of spacetime. Just as such an object that is created at noon and destroyed at one. If we think of the object as three dimensional and enduring through time, it would

[3]In "Parthood and Identity across Time," Thomson writes: "It [temporal parts theory] seems to me a crazy metaphysic—obviously false... It seems to me that its full craziness comes out only when we take the spatial analogy seriously. The metaphysic yields that if I had exactly one bit of chalk in my hand for the last hour, then there is something in my hand which is white, roughly cylindrical in shape... which was not in my hand three minutes ago, and indeed, such that there was no part of it in my hand three minutes ago. As I hold the bit of chalk in my hand, new stuff, new chalk keeps constantly coming into existence *ex nihilo*. That strikes me as obviously false" (Thomson: 210, 213). Crazy or not, one of my central reasons for upholding temporal parts theory is the seriousness with which I take the spatial analogy.

> be appropriate to say that the object exists at different times; the same object exists at noon and at one. Such an object has boundaries along only three dimensions. The whole object is that hunk of matter which entirely fills up those boundaries. The whole object, therefore, exists at noon and still exists at one. A four dimensional object, on the other hand, has boundaries along an additional dimension. The whole object must fill up all its boundaries and, therefore, does not exist at a single moment. If we accept that physical objects are four dimensional, the appropriate thing to say about the object under construction is that it takes up more than an instantaneous region of time. It does not exist *at* noon *and* one; rather it exists *from* noon *until* one. Instead of thinking of an object as existing at various times, we should think of it as existing within regions of time. (Heller: 322)

This is about as complete a description of four-dimensional objects as we could ever hope for. The formalization of 4Dism follows straightforwardly:

> 4Dism: "x is an object" is true iff. the boundaries of x are completely contained within four dimensions.

One immediate observation to make about 4Dism is that although the formal definition does not say so, if it turned out tomorrow that there were actually two dimensions of time, there would likely not be one 4Dist who did not instantly become a 5Dist. This brings to light an important contrast between the formal definitions of 4Dism and 3Dism. The 4Dist conception of objecthood is governed by the scientific account of dimensionality: science (apart from what goes on in the most advanced levels of theoretical physics) accounts for four dimensions, three of space and one of time. 4Dism might be better described as *N*dism, the *n* filled in by physics. 3Dism, on the other hand, is a conception of objecthood that is constant no matter what science says about the dimensions and their number. The discovery of a second dimension of time would not transform 3Dists into 4Dists; the 3Dists would merely modify their description so that three-dimensional objects are

wholly located at coordinate pairs in time. One and the same van Inwagen family cat was present at ((t_{x1}, t_{y1}) and (t_{x2}, t_{y2}).

Thus, the 4Dist can readily identify the principle that his theory of objecthood relies on to state the number of dimensions containing an object's boundaries: that number is whatever science says is the number of dimensions of space-time. The 4Dist has no personal stake in objects being four-dimensional, and he would readily shift to five-, six-, seven-dimensional object predication if science told him to, perhaps 11Dism if the conclusions of string theory warrant such a view of space-time. By contrast 3Dism has two alternatives: it can either declare objects to be brutely three-dimensional, which does not sound terribly promising; or it can ground its view of objecthood on a disanalogy between space and time.

Such a disanalogy, if it were compelling, would be a compelling argument in favor of 3Dism, and what is just as important, is the only compelling argument for 3Dism over 4Dism. No one who does not affirm a strong version of a space/time disanalogy has any reason to be a 3Dist, and no plausible 3Dism fails to be grounded in one. (Such a disanalogy is strongly implied, I think it is clear, by the offhand dismissals of 4Dism on van Inwagen's part.) If, somehow, it turned out that there were two dimensions of space, the 3Dists would be 2Dists (and the 4Dists would be 3Dists); they would not suddenly adopt what is now the 4Dist view of objects filling up all the dimensions of spacetime just because the total number of dimensions turned out to match the number of dimensions that their theory had previously claimed completely contained the boundaries of objects. Strict 3Dism does not entail the following point, but no 3Dist fails to uphold it: the three dimensions in which an object's boundaries are entirely located are the spatial dimensions. Whatever the disanalogy between space and time is, it must be such that a spatial dimension is one in which an object's boundaries can be located, whereas a temporal dimension is not. So we are in a position to give two further definitions of 3Dism and 4Dism, which are not yet conceptual entailments of any theory of temporal extension, but do explain how each theory arrives at the number in its title:

Principled 4Dism: if x is an object, the number of dimensions

> in which the boundaries of x are located is given by and equal to the total number of dimensions of space and time; on the assumption that there are four dimensions of space and time, an object's boundaries are entirely located within four dimensions.
>
> Principled 3Dism: if x is an object, the numbers of dimensions in which the boundaries of x are located is given by and equal to the total number of dimensions of space; on the assumption that there are three dimensions of space, an object's boundaries are entirely located within those three dimensions and no others.

Notice that the principled reformulations of 3Dism and 4Dism bring the terms "space" and "time"—the terms of the debate over persistence and extension—back into the picture. The strict definitions of 3Dism and 4Dism do not incline towards any view of the form of temporal extension; they are all equally compatible with temporal entension, pertension, and obtension; in fact, strict 3Dism and 4Dism are rather silly doctrines, arbitrary precisifications of the *n*-dimensionality of objecthood that are not a priori better ways to carve up nature at its joints, as it were, than 1Dism or 6Dism or 156Dism. The principled articulations of the theories, however, do have bearing on a conception of temporal extension.

Now as we have seen, the principle of principled 4Dism is the scientific numeration of the dimensions of space-time, and the principle of principled 3Dism is the disanalogy of space and time. This latter principle I cannot regard as even remotely plausible, no matter how charitably I endeavor to interpret it. The implausibility of the disanalogy, of course, stands in stark contrast to the fact that, at least until the 20th century, virtually everyone, including virtually every philosopher, was a de facto 3Dist in deference to it. Today, almost everyone is still a de facto 3Dist in deference to the disanalogy; what has changed is that many and perhaps most philosophers have abandoned it, as has the scientific community, de facto if not in terms of overt affirmation of the philosophical doctrine of 4Dism. The reason the disanalogy is not remotely plausible is that modern physics, specifically the

special and general theories of relativity, establish beyond much reasonable doubt that the disanalogy is a cognitive illusion. Any extension in space is an extension in time as well; any communication between two objects in space is a communication in time; any movement in space is a movement in time. The Hubble telescope produces images of events that took place billions of light-years from our planet, and when we look at those images, we are looking just as much at faraway times as at faraway places. The only difference, however, between visual perception of the images in the Hubble telescope and visual perception in our everyday lives is the degree to which we are separated from the percepts in each case. We look back in time whenever we gaze at our own navels, exactly as far back as the time it took light from our navels to reach our eyes.

Perhaps this is too strong a line to take against 3Dism. The 3Dist will, of course, have a story to tell by which relativity and 3Dism turn out compatible. But the crucial point is that any such story is just that, a story. Ptolemaic astronomers had their own geocentric story to tell that was in fact a better predictor of data than Galilean astronomy, at least until it was corrected by Kepler. Continuing to affirm a disanalogy of space and time strong enough to provide sufficient motivation for 3Dism in the face of the Einsteinian revolution in physics is less a metaphysical doctrine than simple metaphysical chauvinism.

Now, of the alternate theories of temporal extension, only one, temporal entension, is explicitly and obviously consistent with principled 3Dism.[4] Temporal entension, that is, endurantism, is the doctrine that all non-momentary objects are wholly present at the multiple points in time at which they exist, and whole presence, according to the intuitive Parsons definition,[5] reduces to the presence of the object along with any proper parts it has. Temporal entension agrees with 3Dism, because 3Dism claims that an ob-

[4]Hereafter, "3Dism" and "4Dism" are abbreviations of "principled 3Dism" and "principled 4Dism."

[5]"Definition of 'entirely located': x is entirely located at r iff x is located at r and there is no region of space-time disjoint (i.e., not sharing a subregion) from r at which x is located. Definition of 'wholly located': x is wholly located at r iff x is located at r and there is no proper part of x (i.e. a part of x not identical to x) not located at r." (Parsons, 2003: 4).

ject fills up, as it were, boundaries in the three spatial dimensions, and is wholly, multiply located—i.e., all its parts, without exception, are multiply located—at points along the temporal dimension, in just the same way that the boundaries of a boat do not include the river on which it is floating.

A non-momentary temporally entended object would clearly be impossible under 4Dism: for an object to be temporally entended, it must be wholly present at every point in time at which it exists; for any object to be wholly present at any point, no part of it can be missing from any point; but since, under 4Dism, a non-momentary object fills up a region of time, the only way none of its parts could be missing from every point in time at which it exists is if it is entirely located at multiple points in time; but entire location at multiple points is a contradiction; hence, under 4Dist entension, there are only momentary objects. Thus, if we assume both 4Dism and the existence of non-momentary objects as premises, temporal entension is simply not a metaphysically viable thesis.

Temporal pertension and temporal obtension remain to be analyzed in the 3Dism/4Dism framework. Let us consider temporal pertension first. Any non-momentary object that is temporally pertended has parts at multiple points in time; it exists, as Heller puts it, *from* t_1 *until* t_2; t_1 and t_2 are its temporal boundaries. Hence it fills up the region of time bounded by t_1 and t_2; as long as its spatial extent is also non-zero, it is bounded in space; and this is as much as to say that a non-momentary, spatially non-zero extended object is a four-dimensional object. The case with temporal entension, therefore, is reversed precisely: the only way that temporal pertension could be false given 4Dism is if there are only momentary objects, but even if that were the case, it would not be enough to establish that temporal pertension is false—one would have to find independent evidence that the momentary objects are temporally extended in a way other than pertension. Similarly, the temporal pertension of a non-momentary object is impossible under 3Dism, because 3Dism claims that an object's boundaries are entirely located in the three dimensions of space, but a non-momentary temporally pertended object has boundaries in time as well as in space. Hence, temporal pertension could only obtain for momentary objects given 3Dism, but

we need not be concerned with this outcome, because 3Dism is false.

The case of temporal obtension is slightly more complicated. There is a superficial manner in which it is equally compatible with both 3Dism and 4Dism, because objecthood under temporal obtension is only a predicate of momentary extensions, and since 3Dism and 4Dism are only in disagreement about objects having boundaries in time in addition to space, temporal obtension is simply indifferent to the dispute between the two. We could just leave it here, noting that what is important for there to be any reason to believe in a theory of temporal extension is for that theory to be consistent with 4Dism. Since temporal obtension clearly is consistent with 4Dism, the temporal obtensionist could stop right there without considering the status of temporal obtension under the false doctrine of 3Dism. But, as it turns out, temporal obtension is like temporal pertension in being incompatible with 3Dism. At least, temporal obtension, in the sense that its proponents generally use it, is inconsistent with 3Dism. This is because the proponents of temporal obtension claim no ontological dispute with temporal pertensionism, only a semantic one. The same things exist, they say, only the term "object" is predicated in a different way. Sider, declaring an ontological virtue of the temporal obtension move, "grant[s] the existence of the worm theorist's worm. My ontology is the same as the worm theorist's: four-dimensionalism." (Sider: 191). In other words, temporal obtension and temporal pertension are ways of describing the same arrangement of the furniture of ontology; under Siderian temporal obtension, there might not be any "objects" filling up boundaries in time, but something surely is, and Siderian temporal obtension is inconsistent with 3Dism to the same degree and for the same reason.

We thus rule out 3Dism and spatiotemporal entension as fatally implausible metaphysical doctrines, no matter what brilliant arguments can be mustered on their behalf, on the assumption that the disanalogy of space and time is false, or at least so weak as to preclude the notion that objects can only fill up regions of space but not of time-or, conceivably, regions of time and not space, but no one construes the disanalogy that way, because everyone agrees with the manifest truth that objects do fill up regions of

space. This conclusion warrants a further conclusion: if space and time are sufficiently similar that they do not differ with regard to the physical possibility that objects fill up regions in either, there is no reason to posit variation in the form of extension between time and space. Hence, spatial extension and temporal extension are formed in the same way, because ultimately space and time are the same (or awfully close to the same); there is just space-time, just spatiotemporal extension, and no ontological discrepancy between the forms of extension in time and space.

The obvious next step is to decide between spatiotemporal pertension and obtension. Making such a decision, however, is a hopeless task if Sider is correct that the two views are semantic alternatives for the same ontology. It could, in other words, be preferable to *describe* spatiotemporal extension in one form rather than the other, but that preference, clearly, has nothing to do with the ontology of extension. If, on the other hand, there is an ontological discrepancy between them, the matter of deciding between the remaining forms of spatiotemporal extension becomes a matter of deciding which is a more plausible ontological theory—in other words, a direct evaluation of their merits as metaphysical doctrines.

Sider places a crucial condition on his claim that he and the "worm theorist" have the same ontology: they have the same ontology "given unrestricted composition" (Sider: 191). But unrestricted mereological composition is a rather large theoretical item to give away without any fight. Unrestricted mereological composition is the view that two x's compose a y anytime that the two x's exist; there are no additional conditions or restraints placed on composition. Or as van Inwagen characterizes it:

> It is impossible for one to bring it about that something is such that the xs compose it, because, necessarily (if the xs are disjoint), something is such that the xs compose it. (van Inwagen 1987: 74)

The consequences of this view are quite extraordinary. Given unrestricted composition, there is an object whose proper parts are the desk I am sitting at, the instantaneous temporal part of Hubert Humphrey's left shoe on No-

vember 4, 1967 at 7:43:33 P.M., and a molecule of ammonia on the outermost ring of Saturn. If, in addition to unrestricted composition, modal realism is true, there is an object composed of the temporal part of Humphrey's head on election night in 1968, the temporal part of Humphrey's (or Humphrey's counterpart's) torso in the nearest world in which Humphrey won, and the legs and feet of Humphrey or Humphrey's counterpart in the nearest world in which Humphrey or his counterpart is a golden retriever.

Such a view cries out, as much as anything else does, for the crazy metaphysic objection. But that objection cannot be sufficient to disproving unrestricted composition; for one thing, the conclusions in favor of spatiotemporal pertension and obtension at the expense of entension are subject to the crazy metaphysic objection (paradigmatically, we might say, as the objection was designed for them). But we might want to say, minimally, that we should be reluctant to embrace a view that prompts the crazy metaphysic objection without having a very good reason for doing so—a reason we had in the case of spatiotemporal pertension and obtension, that reason being the scientific conception of space-time. I think a reasonable rule of thumb might say that intuitions do count as evidence, hence the crazy metaphysic objection is really an objection anytime a theory produces egregiously counterintuitive results, but no amount of intuitive evidence can ever be conclusive, or else we would be warranted in regarding the supremely counterintuitive results of theoretical physics as false, whereas they are true, or at least have significant confirmation.

So Sider needs to make a case for unrestricted composition. What is at stake for him in unrestricted composition is the same as what is at stake for us, namely, ontological discrepancy between temporal obtension and temporal pertension. If composition is unrestricted, then there is a something that has as parts any given set of temporal slices, and of course, therefore, a something for all the things temporal pertension takes to be the continuants in ontology; the obtensionist disagrees with the pertensionist about what to call an object, but not about the elements of ontology, what things exist and what things do not. Furthermore, the disanalogy of space and time, whose falsity we regarded as fatal for 3Dism, can be more precisely stated as a

metaphysical disanalogy. Hence, as long as a theory entailed no ontological discrepancy between space and time, we would not say that the theory should be rejected as 3Dism should be rejected if it proposes merely alternative semantics for describing extension in space and time. And Sider's respective positions on temporal and spatial extension follow just such a pattern; i.e., assuming Sider is not motivated to take up spatial counterpart theory, he upholds spatial pertension along with temporal obtension. But he may do so without claiming a disanalogy between space and time, as long as the conjunction of spatial pertension and temporal obtension does not entail an ontological discrepancy between space and time.

If there are restrictions on composition, however, the ontological picture changes considerably. Given restricted composition, the temporal pertensionist claims that a person is a space-time worm, and the obtensionist maintains that what we typically call a person is an arrangement of temporal counterpart slices, and the two really disagree, because for the obtensionist, there is nothing whose parts are all the temporal slices which stand in the counterpart relation to one another, not just not an "object," but nothing at all. What's more, the temporal obtensionist is now in a difficult position with regard to his position on spatial extension, as he may not affirm both temporal obtension and spatial pertension without claiming an ontologically significant disanalogy between space and time, a claim we rule out presumptively, and so he is committed to spatial obtension. Perhaps the theoretical benefits of temporal counterpart theory are strong enough, even given the ontological consequences of restricted composition, to justify upholding spatiotemporal obtension. But given both the wild counterintuitiveness of spatial obtension, and the fact that the temporal obtensionist, as Sider claims, is not likely to want to deny that the pertensionist's spacetime worms do indeed exist, the prospects for temporal obtension really may be said to rise or fall with unrestricted composition—at least, the temporal obtensionist would have to demonstrate a metaphysical advantage of his theory over temporal pertension; semantic advantages would be insufficient to make the case.

Sider's argument for unrestricted composition, which comes out of Lew-

is's argument for it, rests on the claim that restricted composition would entail vagueness about precisely when composition occurs, such vagueness being presumptively fatal to the proposed restriction, because it would not be a matter of semantic indecision—not a matter of language, to borrow a classic case, having no precise definition of "heap" that determines the sharp cut-off point such that any fewer number of grains is not a heap and any number of grains greater than or equal to it is a heap—but a matter of truly ontological indeterminacy, of composites that straddle ontology somewhere between existing and not existing. As Lewis puts it, "It is a vague matter whether a given class satisfies our intuitive *desiderata* for composition…What is this thing such that it sort of is so, and sort of isn't, that there is any such thing? No restriction on composition can be vague. But unless it is vague, it cannot fit the intuitive *desiderata*. So no restriction on composition can serve the intuitions that motivate it. So restriction would be gratuitous. Composition is unrestricted" (Lewis: 212-213).

Sider further develops this view by expanding it to an alternative theory of vagueness on which a proponent of restricted composition might ground a claim that the matter of whether any particular objects satisfy the intuitive criteria of composition is not vague. That theory is epistemicism, characterized by Sider as the claim that:

> [V]agueness never results from indeterminacy of truth value… According to the epistemicist there will be a single hair whose removal results in the man becoming bald. Even though no one could ever know where it lies, this sharp cut-off for the predicate 'bald' exists. Since epistemicists are already accustomed to accepting sharp cut-offs for predicates like 'heap' and 'bald,' one might think they would also be happy with a sharp cut-off in a continuous series of cases of composition. (Sider: 130-131)

It is far beyond the scope of this paper to explore the merits of any of the theories of vagueness, but I will say that my answer to the Lewis/Sider argument from vagueness against restricted composition is something like

epistemicism. Sider derives his argument that even the epistemicist should accept unrestricted composition from two alternative epistemicist characterizations of how it is that there is "a sharp cut-off in a continuous series of cases of composition":

> (1) [O]ne candidate [for a cut-off point] is more intrinsically eligible, carves nature at the joints better than the rest, thus granting it *metaphysical* privilege, or
> (2) one candidate fits use better than the rest granting it *semantic* privilege (Ibid.).

Sider claims the epistemicist "surely" should adopt the second view, and he believes this of the epistemicist in all cases, not just determining cut-off points in cases of composition.

He is sure this is so because of the nature of the traditional cases of indeterminacy in vagueness theory—how many hairs a man loses to become bald, how many grains make up a heap, at what point the outback begins and ends, etc. I think, *contra* Sider, that there is nothing the least bit amiss with an epistemicist adopting (1), and indeed, that in the specific case of indeterminacy over when composition occurs, (1) is a far better explanation of why there would be a sharp cut-off point than (2), and that by not simply assuming the epistemicist position to be consistently (1) or consistently (2) in all cases of indeterminacy, we can see why the composition cases are better candidates for (1) and traditional vagueness problems for (2). Cases of heaps and baldness, if they have sharp-cut off points, have them undoubtedly in virtue of the fact that there is, unbeknownst to us, a privileged precisification of "heap" and "baldness," i.e., a rule for the determinate predication of these terms that for whatever reason better expresses the contents of the concepts. The reason that epistemicism should take the semantic approach for baldness is that baldness itself is not, in a fundamental sense, a metaphysical phenomenon; the predicate "bald" is only ever ascribed because of accidents in the development of language and culture that included the development of a word for what happens to men when they lost their hair. Baldness is a phenomenon characterized by a kind of seman-

tic indeterminacy because social, cultural, and linguistic processes created a concept of imprecise semantic content.

By contrast, composition is a metaphysical fundamental if it ever occurs. It might be true that no x's would be ascribed composition were there no language in which to couch the ascription, and that therefore no concept "composition" would exist, but metaphysically, x's would go on composing y, or not, regardless of anything else in the world. The non-existence of the *concept* of composition does not entail or have any bearing on the metaphysical reality of composition; the non-existence of the *concept* of baldness, conversely, entails that nothing is bald. What the (1) characterization of epistemicism in the case of composition provides is a sharp metaphysical cut-off point; every instance of the statement "the x's compose a y" has a determinate truth value. That we cannot know that truth value should not be problematic: the concern over restricted composition was that it would yield an ontology of composites existing to varying degrees rather than existing or not existing simpliciter; as long as every y is either composed simpliciter, or not composed simpliciter, from any x's, it does not matter whether the cut-off point between cases of composition and non-composition is known or knowable. Indeed, as long as any x's determinately do or do not compose a y, metaphysical soundness is preserved regardless of whether or not any conscious beings exist to ask the special and general composition questions.

So goes, in any case, my response to the vagueness challenge to restricted composition. Further development of the theory might account for the (2) characterization of epistemicism in cases like baldness, or even linguistic indeterminacy, in which case the theory would be one that rejected a unitary phenomenon of vagueness, on the grounds that any instance of apparent indeterminacy may seem similar in a variety of ways, but that nevertheless the root causes of particular cases of indeterminacy need not be the same. I do not claim that I have conclusively disproved unrestricted composition, or anything of the sort, but I do claim to have shown one segmented way that unrestricted composition could be false: (1) it has wildly counterintuitive consequences, and hence merits the crazy metaphysic objection; (2) an epistemic account of indeterminacy in composition allows

for a non-self-defeating claim of restricted composition; (3) that epistemic account takes a metaphysical rather than a semantic position on the determinacy of composition, which is plausible in the case of composition because it is a pre-linguistic ontological fundament—composition would exist if language did not in a way that is not (or not obviously) true of baldness. From this point forward, I will proceed from the assumption of restricted composition through some kind of epistemic account of vagueness. If that assumption is wrong, then entension is still ruled out as the form of spatial and temporal extension, but otherwise, there is no metaphysical advantage that obtension or pertension could be said to enjoy over one another, though one may turn out to be better at describing spatial or temporal extension than the other. This, indeed, is Sider's view: part-theoretic extension in space and counterpart-theoretic extension in time, which come out to the same ontology given unrestricted composition.

Given restricted composition, however, the differences between spatiotemporal obtension and pertension take on ontological significance and come into sharp relief. The immediate consequence of restricted composition is that we return to where we were in the beginning of the section, united spatiotemporal extension; there is no more wiggle room for a hybrid view, as any hybrid view is now one for which a fatal consequence, i.e., an ontologically significant disanalogy between space and time, is unavoidable. This being the case, we might have an immediate, prima facie argument for rejecting spatiotemporal obtension, namely, spatial obtension is both intuitively bizarre and has to its credit no apparent explanatory advantages over spatial pertension, and so it simply does not warrant rational belief. But if a hybrid view is metaphysically untenable given restricted composition, the failure of spatial obtension entails the failure of temporal obtension: it does not matter how strong the argument for temporal counterpart theory is, spatial obtension is an anchor on the fate of temporal obtension, and pending a strong case for spatial obtension over spatial pertension, temporal obtension sinks. Such a conclusion is tempting but perhaps premature. Spatial obtension has a suggestive resemblance to compositional nihilism (see fn. 1), and if spatial counterpart theory could serve as the se-

mantic theory behind nihilism's paraphrases of compositional appearance, it could be that there is a reason, at least for nihilists, to uphold spatial obtension. Let us not pre-emptively assume, therefore, that spatial obtension is false. Even so, I hope to demonstrate, temporal obtension should be judged on its own merits and rejected.

The argument for temporal obtension was based on the difficulties that the problem of coincidence poses for temporal pertension along with the need to preserve 4Dism. But is the problem of coincidence a problem at all? For Sider, it is a linguistic problem, a problem for the semantics of counting, but not a metaphysical problem: "Coincidence between spacetime worms is *metaphysically* unobjectionable" (Sider: 191). On the other hand, as we saw in the original discussion of the problem of coincidence, there is a view on which coincidence is indeed a metaphysical problem. This is the view expressed by Michael Della Rocca as the "[apparent] conceptual truth to the effect that there cannot be distinct indiscernible things that occupy precisely the same location at all the same times and have all the same parts" (Della Rocca: 10). The response, I think, is that temporal pertensionism results in two possible alternative construals of the problem of coincidence: (1) affirmation of the existence of "distinct indiscernible things that occupy precisely the same location at all the same times and have all the same parts"; (2) denial that partial overlap produces metaphysically problematic complete overlap.

That is, unless a temporal pertensionist wishes to deny the possibility of Parfit's fission cases—a denial that could only be an ad hoc avoidance of coincidence—she must accept that there can be partial overlap of spacetime worms. Hence, there are possible worlds in which two or more worms share some of their parts. But if there are possible worlds in which two or more worms share some of their parts, there are also worlds—given an intuitive recombination principle like the Paull/Sider Isolation Principle[6]—in which only the shared parts exist. In those worlds, it turns out, distinct

[6]"For any object x in any world w, there is a world w' containing a duplicate of x in isolation," (Paull and Sider:838.) "X exists in isolation" means that x exists in a world containing only x, x's parts, and anything whose existence is entailed by the existence of x.

spacetime worms—it could be two or twenty or two hundred of them—occupy precisely the same location throughout spacetime and have all the same parts. This is response (1) from above: the problem of coincidence is not a problem; coincidence is unproblematic whether partial or complete. If Della Rocca's conceptual truth is to be upheld, on the other hand, either worm theory is false, or the recombination principle involved is false. So the temporal pertensionist who maintains that complete overlap is problematic denies the recombination that would generate complete overlap from partial overlap. This is response (2): coincidence is unproblematic for temporal pertensionism because unproblematic partial overlap does not entail problematic complete overlap.

Thus, the temporal pertensionist has two possible answers to the problem of coincidence. Depending on the views of the worm theorist in question on recombination, the problem is either obviated by not being a problem at all (1) or by not being a problem for worm theory (2). Sider, who presumably believes in his own recombination principle, and who also believes that coinciding worms exist—just that they are not the bearers of the predicate "is an object"—should surely accept that the worlds of completely overlapping worms exist. This is my view as well. Now, aside from merely reciting a catechism of "I'll bite that bullet" with regard to "distinct indiscernible things that occupy precisely the same location at all the same times and have all the same parts," what can someone who shares my and Sider's view say to a proponent of Della Rocca's conceptual truth that this is impossible? Are the two camps in an unresolvable stalemate? I think that they need not be, and that the proponent of the possibility of complete overlap can give an account that distinguishes the metaphysical possibility, the completely overlapping worms, from the counterintuitive epistemic result that worlds of two, twenty, and two hundred qualitative indiscernibles exist. Thus, we would say that provided there is a fact of the matter in every case, a determinate number, about how many worms are present, the epistemic impossibility of knowing that number has no bearing on the metaphysical possibility of overlap. Notice the similarity between this account of the problem of coincidence and my earlier account of restricted composition as

a matter of epistemic indeterminacy; if this account succeeds in not only affirming but explaining the metaphysical possibility of complete overlap, the parallel account of composition may just succeed with regard to restricted composition.

Provided that at least one of the two alternative temporal pertensionist responses to the problem of coincidence is sufficient to establish temporal pertensionism's metaphysical soundness, the basic unsoundness of temporal obtension becomes apparent. Since we are assuming restricted composition and therefore an ontological discrepancy between the two theories, and have just seen that pertensionism is sound, I do not see any basis for a claim that temporal obtensionism is a more plausible metaphysical doctrine. None of the ontological virtues of pertensionism are available to it any longer, and it is left with the strange result that nothing—not just no bearers of object predicates, but nothing at all—exists for more than a moment. This ontologically strong form of temporal obtensionism, furthermore, falls prey to the same objection Lewis raised against temporal entension: it does away with monadic properties and substitutes either relations or relational properties in their place. The resulting account of properties, furthermore, violates some very basic intuitions about properties in general.

The property that I'm fairly certain I have, for example, of having been born on November 27, 1983—whether it is an intrinsic property of me (perhaps birthdates are essential properties) or relational property to the time—turns out not to be a property of me at all, but a property of someone else who is (tenselessly) born on November 27, 1983. The counterpart relation, it is true, gives a semantics according to which "I was born on November 27, 1983" comes out true, but not in virtue of anything to do with me; if it is true that I was born on such and such a date, that is because someone else who is not me was born on that date. I do not see how this account of properties could possibly be true. And with its falsity, the prospects for temporal obtension begin to look significantly poor. It enjoys no theoretical ontological advantage over temporal pertension, either in the account it gives of what objects exist or how those objects have properties, and indeed, the opposite is true, temporal pertension is theoretically stronger on

both counts. The final remaining argument for temporal obtension is that it gives the correct count of persons in Parfit's fission and fusion cases, and temporal pertension does not. But it seems to me a borderline conceptual truth that a metaphysically sound theory is automatically preferable to a metaphysically unsound theory prior to any semantic considerations.

If temporal pertensionism gives by far the best account of the ontology of extension in time, as it seems to, and a consequence of this fact is that it poses problems for the semantics of counting, so much the worse for the semantics of counting. Temporal obtension fails, temporal pertension succeeds, and because the disanalogy between time and space is false, the correct account of spatiotemporal extension is unified: spatiotemporal pertension, spacetime worms.

Works Cited

Della Rocca, Michael. "Two Spheres, Twenty Spheres, and the Identity of Indiscernibles." *Pacific Philosophical Quarterly* 86 (2005): 480-492.

Heller, Mark. "Temporal Parts of Four Dimensional Objects." *Philosophical Studies* 46 (1984): 323-334.

Kripke, Saul. *Naming and Necessity*. Cambridge: Harvard University Press, 1972.

Lewis, David. *On the Plurality of Worlds*. Oxford: Blackwell, 1986.

Parfit, Derek. "Personal Identity." *Philosophical Review* 80.1 (1971): 3-27.

Parsons, Josh. "Entension, or How it Could Happen That an Object Is Wholly Located in Each of Many Places." March 4, 2003. <http://weka.ucdavis.edu/ jp30/papers/>.

Paull, R. Cranston, and Theodore R. Sider. "In Defense of Global Supervenience." *Philosophy and Phenomenological Research* 52.4 (1992): 833-854.

Sider, Theodore. *Four Dimensionalism*. New York: Oxford University Press, 2001.

Thomson, Judith Jarvis. "Parthood and Identity Across Time." *Journal of Philosophy* 80.4 (1983): 201-220.

van Inwagen, Peter. "When Are Objects Parts?" *Philosophical Perspectives* 1 (1987): 21-47.

———. *Metaphysics*. Boulder: Westview Press, 2002.

How Act-Utilitarianism is Directly Collectively Self-Defeating

Nick Day
Nottingham University

In this paper Day argues that there are particular actions that will *always* have an imperceptible effect when performed individually, but that when these actions are performed by a large number of people the collective effects can be perceptibly large. Thus, the act-utilitarian may find herself in a situation relevantly similar to a prisoner's dilemma; as such, act-utilitarianism is directly collectively self-defeating. The paper then discusses the practical implications of this problem, such as in the case of the dilemma of greenhouse gas emissions and global warming, where imperceptible individual actions sum into perceptible effects. Day concludes by arguing that if no solution can be found to this problem, then unanimous support for act-utilitarianism will be theoretically disastrous.

In *Reasons and Persons*, Derek Parfit (1984) argues that commonsense morality, by making agent-relative claims, falls foul of a prisoner's dilemma, and thus can be seen to be directly collectively self-defeating. Parfit (103) suggests that this is a reason to reject commonsense morality. I will argue, against Parfit, that act-utilitarianism similarly succumbs to this problem, as it too is directly collectively self-defeating.

To begin, if we call some theory T, and call the aims that this theory gives us our T-given aims, then T is *directly collectively self-defeating* when "it is certain that, if we *all* successfully follow T, we will thereby cause the T-given aims *of each* to be worse achieved than they would have been if *none* of us had successfully followed T" (Parfit 1984: 55) ("All" and "none" giving us the simplest cases). A many-person dilemma (a prisoner's

dilemma involving more than two people) is an example of a situation in which *self-interest* is directly collectively self-defeating. A many-person dilemma occurs only when, in a group of many people, each has a choice that affects others (in terms of benefits and harms) and each person's choice does not affect the choices of the other people. Suppose, in such a situation, I can choose either to give a small benefit to each person in a group (including myself) or give only a large benefit (say, twice that of the small) to myself. Consider this example:

> We are the team of ten people that win a quiz show. In order to determine our winnings we are given a final task. Each person has the choice to either give $200 in prize money to only him or herself, or $100 of prize money to each person in the group, including him or herself; no one will know what the others have chosen.

The greatest *total* amount of winnings that can be taken home is $10,000. This sum is obtained by each member of the team choosing to give $100 to every other person, thus leaving each of the ten people with $1,000 to take home. However, it is also true that the greatest amount of winnings that I can get *for myself*, whatever the others choose, is secured by giving myself $200; whatever the others choose this will *always* give me $100 more than I would have otherwise gotten. If all the people reason in this way, choosing to give themselves the $200, then each person ends up with only that $200—only a fifth of what they could each have had, if *all* had chosen differently. In this kind of situation self-interest is directly collectively self-defeating. By all correctly following self-interest they have each done worse, *in self-interested terms*, than they would have done if none of them had successfully followed self-interest.

This type of situation is not a dilemma for the act-utilitarian. An act-utilitarian must impartially choose the course of action that maximizes the sum of *all* benefits and harms. In this case the utilitarian choice is clearly to give $100 to each person, in order to achieve the greatest total winnings of $10,000; this also happens to be the best possible outcome for each person

individually. Thus a group of act-utilitarians does better in self-interested terms than a group of self-interested people. There are many everyday situations in which this kind of dilemma occurs. Parfit gives the example of a group of commuters that have the choice between driving and taking a bus: "each goes faster if he drives, but if all drive each goes slower than if *all* take buses" (1984: 61).

Now let us look at the type of dilemma that occurs when imperceptible effects are involved. It has a structure very similar to that of the many-person dilemma described above. Suppose I have the choice of either giving myself a large benefit or affecting everyone (including me) *imperceptibly* (this replaces the small benefit of the many-person dilemma). An example of this sort of situation would be the following:

> We are 10,000 intensely thirsty people in a desert. We are carrying with us a water cart capable of holding many thousands of liters of water, but unfortunately it is now empty. We come across an oasis that is rapidly drying up. We have very little time before it disappears completely. We each have only enough time *either* to drink one liter of water *or* to add two liters of water to the water cart. There will be no way of telling who has chosen to contribute to the water cart so it is decided that this water will later be shared out equally amongst all 10,000 people when matters are less urgent.

At first it may seem that we must reach the same conclusion in this case as we did in the standard many-person dilemma. Each person has a choice of either drinking an extra one liter or giving 0.2ml (2 liters/10,000) to each of the 10,000 people. The self-interested person chooses to just drink the one liter of water because he knows that, whatever the other people choose, he will receive most water this way (i.e. always an extra one liter on top of whatever the other people's contributions amount to). However, if all choose this option, each person ends up with significantly less water than if everyone had chosen to contribute two liters to the water cart: if all choose to drink one liter then this is all they ever receive, but if all choose to con-

tribute two liters, then each receives two liters back. Self-interest again appears to be directly collectively self-defeating—if everyone were to act on self-interest, each person would be worse off, in self-interested terms, than if none of them had been self-interested. However, can act-utilitarianism be seen as a solution in this case, as it was in our earlier many-person dilemma?

Here, I make two assumptions: firstly, that drinking an extra 0.2ml of water is *always* imperceptible, and secondly, that the utilitarian judges the correctness of an action by how it affects perceptual experiences. The act-utilitarian must consider the two choices: *either* give an extra liter to himself or give an extra 0.2ml to everyone including himself. Being an act-utilitarian, he must choose the course of action that maximizes the sum of all benefits minus harms. From our second assumption we view those benefits and harms as involving changes in perceptual experience. From our first assumption we see that choosing to give an extra 0.2ml to everyone will benefit *no one, no matter what others do*—the suggestion is that this extra 0.2ml will not make any difference to the recipient's perceptual experiences. On the other hand, it is clear that the act-utilitarian, being desperately thirsty, will be greatly benefited in terms of effects on perceptual experience by having an extra liter of water. Therefore in order to maximize the sum of benefits minus harms the act-utilitarian *must* choose to drink the one liter of water rather than contributing two liters, as his consumption of one liter produces some benefit and, under our first assumption, the contribution of two liters produce none. This means that if all were to correctly follow act-utilitarianism in this situation, they would each choose to drink the one liter. However, it is clear that *all* drinking one liter is worse, in utilitarian terms, than all contributing two liters. This would mean that the situation I have described is a dilemma both for groups of self-interested people and groups of act-utilitarians. If everyone were to correctly follow act-utilitarianism in this situation, they would all be much worse off than if no one had correctly followed act-utilitarianism (i.e. if they had instead contributed two liters each). Therefore, according to Parfit's definition, but contrary to Parfit's view, act-utilitarianism is *directly collectively self-defeating*.

This conclusion rests on the two initial assumptions outlined above. Before attempting to justify the assumptions, I will suggest a way of understanding perpetual experience that makes sense of the first assumption, that drinking an extra 0.2ml is *always* imperceptible. More generally (since the number of people in the desert example could be increased) this is the claim that there are such things as actions that are always imperceptible, even though collectively the effects of those actions are perceptible.

How is this kind of effect possible? I suggest that it involves vagueness. A concept is considered vague when it can be analyzed in "terms of an underlying continuum along which an imperceptible or unimportant change occurs" (Pelletier 1999: 946), i.e., when there are no strict boundaries along this continuum. This contrasts with some non-vague concepts that can be analyzed in terms of an underlying continuum, but one that has particular points at which perceptible or important change occurs, i.e. strict boundaries. For instance, the predicate "is 21 years old" (when applied to people) can be analyzed in terms of an underlying continuum (age) with *strict* boundaries that govern its application, namely at the precise point when a person is 21 years old and another precise point one year later when he no longer is. A vague term like "heap," however, cannot be described in this way. There is an underlying continuum of grains of sand that is relevant to its application, but along this continuum there is no point at which an important change occurs; each individual grain of sand is not important to the application of the term "heap." This is a case of *linguistic vagueness*: it is the boundaries of the concept that are vague. This kind of vagueness provides the basis for one set of Sorites paradoxes, for example Cicero's heap paradox: a single grain of sand cannot have the term "heap" applied to it, and adding one grain of sand to something that is not a heap does not allow you to apply "heap," therefore you can never apply the term "heap." Zeno has a similar paradox: dropping one millet seed makes no sound, and therefore unloading a basketful of millet seeds makes no sound. The vagueness involved here is not linguistic: it does not involve blurred boundaries to the use of a concept. It is, rather, *perceptual vagueness*. The suggestion here is that perceptual experience is vague in an analogous way to the vagueness of

concepts like "heap": it can be analyzed in terms of an underlying continuum (color, sound, pain) along which imperceptible or unimportant change occurs. Though dropping one millet seed is inaudible, and the addition of each *individual* millet seed to a group adds nothing to the loudness of the sound that is heard when they are dropped, *collectively* they *do* make an audible sound. Perceptual vagueness is key to showing that act-utilitarianism is self-defeating; it is the basis of the claim that an individual act can have an imperceptible effect, but collectively be perceptible.

I have argued that there are situations in which act-utilitarianism is self-defeating on the basis of two assumptions: (1) perceptual vagueness exists and (2) right and wrong are defined in terms of the effects on perceptual experience of an act. However, there are theories that suggest that either (1) or (2) is false, and thus deny the possibility of such self-defeating situations. I will discuss these theories and argue against them.

Parfit (1984) accepts the possibility of (1), but suggests that consequentialists must reject (2) in order to avoid a (morally) absurd conclusion. Consider this example:

> 1000 torturers have a single victim attached to a torture device. Each torturer presses a button that causes an imperceptible increase in the power to the torture device. Though the victim cannot notice the difference made by each single press of the button, after 1000 presses he is in unbearable pain.

If we accept (1) and (2) we must also accept that none of the torturers do anything morally wrong. This is an important example: it represents the form of the problem that occurs when *only* imperceptible effects are involved. A way of showing that the torturers are wrong in terms of act-utilitarianism would most likely be a solution to situations in which act-utilitarianism appears to be self-defeating. Parfit suggests that it is (morally) absurd to suggest that the torturers are not wrong, and proposes two theories that allow us to avoid this conclusion. First he shows that we must reject one of the following two claims in order to avoid a Sorites Paradox:

> a) Someone's pain cannot become *imperceptibly* better or worse.

> Someone's pain cannot become either less bad, or worse, if this person could not possibly notice any difference.
> b) 'At least as bad as' and 'not worse than' are, when applied to pains, transitive relations. Thus, if someone's pain in Outcome (2) is at least as bad as it was in outcome (1), and his pain in outcome (3) is at least as bad as it was in Outcome (2), his pain in Outcome (3) must be at least as bad as it was in Outcome (1)." (Parfit 1984: 78)

If we accept both (a) and (b), then we must accept a (logically) absurd conclusion. From (a) we must accept that since each press of the button is imperceptible, and therefore not harmful, each press of the button creates a situation that is 'not worse than' the one before. With (b) we must then also accept that since the outcome created by 1000 button presses is not worse than the outcome of 999 button presses, and 999 this is not worse than 998 button presses, etc., the outcome of 1000 button presses is not worse than 1 button press. This is absurd.

We must therefore reject (a) or (b), and Parfit (1984: 79) suggests that by rejecting either we must also reject (2). If we reject (a) then we accept that benefits and harms can be imperceptible. While still accepting the utilitarian principle that right and wrong are defined in terms of benefits and harms, we reject (2) by suggesting that those benefits can be defined in terms other than effects on perceptual experience. Thus, in the example given above, each torturer is morally wrong in virtue of doing something that is slightly harmful, though imperceptible, to the victim. By accepting the theory of imperceptible benefits we have a solution to situations in which act-utilitarianism is self-defeating: though an individual act might always have imperceptible effects, each of these acts might be slightly harmful or beneficial. For instance, in the example of the thirsty people in the desert we might find that we should each contribute two liters because this has 10,000 very small (imperceptible) benefits that outweigh the benefit to me of drinking the one liter. Is this an acceptable solution for act-utilitarianism?

It seems not: utilitarian theories do not have the capability to consider imperceptible effects as benefits in the required way. Consider forms of

utilitarianism that prescribe in terms of mental states. These forms can vary widely in how they evaluate consequences. For instance, hedonistic utilitarianism values pleasure or happiness, preference utilitarianism values the satisfaction of desires, and idealistic utilitarianism values particular experiences independent of their being valued or creating happiness. However, all these (as mental state theories) value some mode of sensation; what matters is "how people's experiences feel 'from the inside"' (Nozick 1974: 42). This focus makes Nozick's experience machine an all-encompassing argument against mental state utilitarianism. The idea of imperceptible benefits goes directly against this claim; by suggesting that there is value in something that by definition has no effect on our experiences, we reject the claim that is common to all these utilitarian theories. If it even makes sense to talk about imperceptible pain, this is certainly not a kind of pain that a mental state utilitarian will care about. The acceptance of imperceptible effects as beneficial is therefore not possible for the mental state utilitarian.

Now consider a non-mental state preference utilitarian theory like that of Peter Singer (1987: 171). On this view, imperceptible effects can be counted as beneficial; it is precisely that unknown consequences can still be viewed as beneficial to a person that distinguishes this form of preference utilitarianism from the mental state form. However, the preference utilitarian's shift of focus is insufficient to provide a solution to self-defeating situations. For these scenarios involve supposed perceptual vagueness; in these cases it is *a mental state* for which a preference has been formed-for example, the relieving of thirst. Though we may accept that there are some preferences whose satisfaction has value even if that satisfaction is unknown to the owner of the preference, such as the desire for true friends or for a successful life, it is not the case that all satisfaction of preferences can have value like this. In particular, preferences for particular mental states seem necessarily to be preferences for which the satisfaction must be known in order for it to have value, and, in fact, preferences that *cannot* be satisfied unbeknownst to the owner of the preference. Since it is exactly this kind of preference that is involved in self-defeating situations for act-utilitarianism, it is of no use to the non-mental state preference utilitarian that he can accept

some imperceptible benefits. Rather, he must accept that the imperceptible effects in these cases are not beneficial, and thus must accept (a), *in at least these cases*. There is, thus, no form of act-utilitarianism that can consider imperceptible effects as benefits in a way that allows a solution to the self-defeating situations.

Given such failure, we must reject (b), that "not worse than" is transitive, instead of rejecting (a), that benefits cannot be imperceptible. However, Parfit (1984: 79) suggests that accepting (b) merely means that we have to find another reason to reject (2), the claim that right and wrong are defined in terms of an action's perceptible effects. If we do not, then we must accept a morally absurd conclusion: that the torturers do not do anything wrong in causing their victim unbearable pain. Without being able to appeal to imperceptible benefits we must find another way of showing that the torturers are immoral. Parfit suggests that we must appeal to what they do *together*. This seems plausible as it fits with perceptual vagueness in an intuitive way. The argument might be made: "The torturers have done something wrong by causing perceptible pain. Individual acts, though each on its own imperceptible, were collectively perceptible. It is therefore *collectively* that these torturers are wrong." However, the problem here is that it is impossible to determine whose acts it was that collectively caused harm, since each individual act caused none. Parfit offers this principle for judging the individual acts:

> (C7) Even if an act harms no one, this act may be wrong because it is one of a set of acts that *together* harm other people. (1986: 847)

How are we to demarcate this set of acts that together cause harm? It is not related to any harmful consequences of the individual acts because it has already been noted in (C7) and above that *individually* no harm is caused. Nor can we appeal to Parfit's (1984: 77) claim that the torturers "would cause these people to suffer *most* if they *all* acted in this way": the victims suffer just as much when the button has been pressed by 999 torturers as when it has been pressed by 1000. Parfit suggests that "often, however, we

can tell" (1986: 848) what this set is; for instance "it may be clear who are the people who together cause pollution" (Id.). However, though the set of acts that cause pollution may be clear, this does not show that the set of acts that cause *harm* through pollution is clear, given that an act might cause an amount of pollution that has only imperceptible, and therefore not harmful, effects. To define the "set of acts that together harm other people" in this case as simply all pollution-causing acts, without reference to the actual harmful consequences of each individual act, is to claim that "the rightness and wrongness of an action is a function of the consequences of others doing the same *kind* of action" (Gruzalski 1986: 781). As Gruzalski points out, this is to abandon act-utilitarianism "for a multi-act version of utilitarianism" (Id.). While accepting that benefits and harms are defined in terms of effects on perceptual experience, it denies (2) by suggesting that the rightness and wrongness of an act is *not* to be understood solely in terms of the beneficial or harmful consequences of that act. It is therefore not compatible with act-utilitarianism, which states that the right action is that which itself maximizes the sum of benefits minus harms.

To summarize, Parfit suggests two solutions that reject (2). The first solution gives moral weight to the imperceptible effects of the *individual* action. This approach fails because such weighting is incompatible with utilitarianism. The second solution gives weight to the effects of the actions *taken as a group*. This fails because this grouping of effects cannot be achieved within an *act*-utilitarian theory. Neither of these can therefore be a solution to the dilemmas for act-utilitarianism. This leaves us with the difficulty of accepting that the torturers have not acted wrongly.

Other philosophers (e.g. Lyons (1965), Gruzalski (1986), Matheny (2002)) have suggested that the solution is to reject (1), perceptual vagueness. They instead propose a theory of perceptual thresholds. This is the suggestion that perceptual experience can be analyzed in terms of an underlying continuum with particular points at which important or perceptible change occurs, i.e. there are strict boundaries on perceptible change. I will label this the *Threshold Theory*. I will first describe how this theory is suggested as a solution to utilitarianism's problems with imperceptible effects.

I will argue that the initial plausibility of this theory's solution relies on two mistaken assumptions, and that what the Threshold Theory actually shows is unclear; its supposed solution is lost in the problems of overdetermination. Secondly, I will argue that the Threshold Theory is less plausible than perceptual vagueness as a view of perceptual experience.

Matheny (2002) suggests that the Threshold Theory can show how the torturers were wrong. Suppose the pain caused by the torturers can be divided into 100 threshold units (i.e. a threshold for perceptible change every 10 button presses). The claim that no individual torturer causes harm is not true: every 10th button press causes harm and therefore *some* of the torturers are wrong. Matheny suggests that the act-utilitarian must use *expected benefits* (i.e. probability x benefit of a consequence) as a decision procedure and, because of this, *all* the torturers are wrong. Each torturer should consider his button pressing to have an expected harm calculated by the probability that his act will be the one that completes a threshold, which is 1/10 (the number of button presses he causes / the size of a threshold unit measured in numbers of button presses), multiplied by the harm caused by 1 threshold unit. Each torturer is responsible for the harm from 1/10 of a threshold unit; all the torturers together (1/10 * 1000) are responsible for the harm from 100 threshold units—the total harm caused. It is possible to see how, if this were true, it would allow a solution to the utilitarian dilemma by showing how each person would take responsibility for a portion of the total harm, and the portions added together would equal the total harm. It is an intuitive conclusion but has two major flaws.

Firstly, Matheny's approach implies the fallacy that only the torturer who completes a threshold unit has caused harm. This is a mistake about who causes *actual* harm. In fact, if a threshold unit is completed by the last act, then it is true of every torturer that his act caused the harm; it is true of each that if only they had chosen not to press the button then the last threshold unit would not have been completed and this much less harm would have been caused. Secondly, it mistakenly suggests that a threshold unit will always be completed by the last act. This is true in the above calculations only because we assumed a threshold unit that divided the tor-

turers' actions exactly (see Matheny's examples (2002: 294)). But it is not plausible to suggest this would always be the case. If the last act does not complete a threshold unit then we have a case of *overdetermination*: it is not true of any torturer that by acting differently any less harm would have been caused—the same number of threshold units would still have been completed. Only by making both these mistakes would we suggest that the subjective probability of an individual's act causing harm is 1/10 in the torturers' case. This leads to a mistake in calculating *expected* harm. The probability that I must instead consider when calculating expected harms and benefits in these cases is the probability that the sum of effects from all acts except mine will be on a threshold. If it will not, my act can be of no consequence.

Parfit (1984: 74) points this out in his discussion of elections. The only situation in which my vote can cause harm or benefit is when the election would be tied without it. To judge the expected benefit of my voting I must estimate the probability of this occurring. Calculating the expected benefit in threshold situations therefore involves establishing the probability that others will act in a certain way, i.e. such that the sum of effects from all acts except mine will be on a threshold. It does not involve, as Matheny (2002: 296) suggests, a mathematically deducible probability that is inversely proportional to the threshold unit size.

This conclusion leaves the Threshold Theory's implications for utilitarian dilemmas unclear. Consider the desert example. Act-utilitarianism would be directly collectively self-defeating if by *everyone* following it *correctly*, everyone does worse in utilitarian terms. But it is not clear in threshold cases what everyone following act-utilitarianism *correctly* would entail. The *actual* harms and benefits of each individual's actions depend upon the action of each other individual. We seem stranded in a circle of justification, analogous to Rawls's criticism of perfect altruism: "suppose that in deciding what to do all vote to do what everyone else wants to do. Obviously nothing gets settled" (Rawls 1999: 65). It seems impossible to determine the objective fact about what each act-utilitarian would be correct to do; there is no way of deciding how each individual would actually

maximize benefits without first deciding how another individual would do so. Nor is it clear that there is an answer in terms of *expected* benefits. For this we need to calculate a probability that others will act such that the sum of the effects of their actions will be on a threshold. If this probability were high enough then the greatest expected benefit would lie in contributing two liters and thus act-utilitarianism would not be self-defeating. But, if this probability were too low, e.g. if we had reason to expect that the effects of others' actions would be in between thresholds (as we might in an election), then act-utilitarianism would be self-defeating. But without information on the position of these thresholds or the reasons others might have for acting, the calculation of this probability is entirely arbitrary. Neither of these pieces of information is obtainable—the latter because, even if we knew that all would attempt to follow act-utilitarianism, it is not clear how they would act. For these reasons we cannot say whether or not, if the Threshold Theory were true, act-utilitarianism would be self-defeating.

Having shown the problems that the Threshold Theory would face as a solution even if it were true, I will now look at reasons for rejecting it on its own grounds. Consider again the absurd conclusion in the torturers example of accepting both (a) that benefits cannot be imperceptible and (b) that 'not worse than' is transitive. Gruzalski (1986: 780) suggests that a Threshold Theorist would reject (b), claiming that 'not worse than' is intransitive. This is incorrect. The Threshold Theorist avoids the absurd conclusion by rejecting a claim that was assumed in the example: that "each torturer presses a button that causes an imperceptible increase in the power." Instead it is suggested that, for instance, for every 10 button presses there is a perceptible increase because a threshold unit has been completed, while only 9 button presses has no effect. This is to say that 2 presses is as bad as 1, 3 as bad as 2, and 9 as bad as 8, *but* that 10 *is* worse than 9, and this is why 10 is worse than 1; transitivity is preserved. Due to affirming this transitivity the Threshold Theorist must admit that *however small* the difference between two states of affairs, even the addition of a single molecule of water to someone's drink, that difference could be perceptible. This occurs when the difference lies on a perceptual threshold. This view of percep-

tion seems implausible. Suppose a computer screen displays a wavelength of visible light to an accuracy of 0.01nm. It decreases this wavelength by 0.01nm every 20 seconds. If you watch this display you will eventually notice that it has changed color, but this may take many hours. The Threshold Theorist, by affirming the above transitivity, must be saying that when this occurs it means that you can tell the difference between the current color and the color before the last 0.01nm change. This is to say, for instance, that you could tell the difference between the shade of red at 720.35nm and the shade at 720.34nm. This splits the color red into around 16,000 different shades. It just does not seem true that we can ever make such fine distinctions. Similarly, it does not seem true that we can ever notice the difference in the reduction of thirst between drinking an amount of water one molecule more than another amount. It is much more plausible to suggest that our perceptual experience is vague, as discussed earlier.

Thus, the sharp lines drawn by the Threshold Theory just do not seem plausible. Consider this example: look at a clock to see when the hour hand moves. Does this movement appear as a succession of discrete units of movement (formed each time an individual imperceptible movement completes a threshold unit) or is it that "even though that hand never seems to move, we can soon see that it must have moved" (Parfit 1986: 849)? The Threshold Theory suggests that the movement of the hour hand will appear disjointed, and this appears to be incorrect; it is instead Parfit's description that fits with our actual experience, and for this reason we should reject the Threshold Theory in favor of perceptual vagueness.

If (1) perceptual vagueness exists and (2) right and wrong are defined in terms of effects on perceptual experience, then there are situations in which act-utilitarianism is directly collectively self-defeating. We cannot reject (2) in a way that provides a solution for the act-utilitarian. And though the denial of (1) is implausible, even if it were accepted it is unclear how it could provide a solution for act-utilitarianism. For these reasons I suggest that there are situations in which act-utilitarianism is directly collectively self-defeating. If everyone were to correctly follow act-utilitarianism this could have disastrous results, in utilitarian terms.

Works Cited

Gruzalski, Bart. “Parfit and Utilitarianism.” *Ethics* 96 (1986): 777-782.

Lyons, David. *Forms and Limits of Utilitarianism*. Oxford: Oxford University Press, 1965.

Matheney, Gaverick. “Expected Utility, Contributory Causation, and Vegetarianism.” *Journal of Applied Philosophy* 19.3 (2002): 293-297.

Nozick, Robert. *Anarchy, State and Utopia*. Oxford: Blackwell Publishing, 1974.

Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.

———. “Comments.” *Ethics* 96.4 (1986): 832-872

Pelletier, F. J. “Vagueness” in The Cambridge Dictionary of Philosophy. Ed. R. Audi. Cambridge: Cambridge University Press, 1999.

Rawls, John. *A Theory of Justice*. Oxford: Oxford University Press, 1999.

Singer, Peter. “Life’s Uncertain Journey” in *Metaphysics and Morality*. Ed. Philip Pettit et al. Oxford: Basil Blackwell, 1987.

Physical Modeling and Event Individuation

Martin Glazier
Yale University

If we include events in our ontology, then it is incumbent upon us to define a criterion of identity for those events. Glazier surveys the identity criteria of Davidson, Quine, and Kim, and argues that their criteria are either circular or fail to respect our intuitions in certain problematic cases. He proposes a new criterion according to which two events *x* and *y* are identical if and only if they cannot be physically analyzed as separate components of a process. Glazier argues that his criterion successfully deals with the cases that are problematic for the other three criteria.

Davidson (1969) argues that events are genuine entities that should be countenanced by our ontology, and I find this claim plausible in light of his arguments. But the serious hurdle remains of defining a criterion of identity for events. It is crucial to the ontological claim that we be able to specify a method of individuation for events, lest we risk violating Quine's dictum of "no entity without identity." In this essay I examine three criteria proposed in the literature and conclude that they are unsatisfactory. I then propose a new criterion that I hope charts a course between the three I reject.

Davidson takes a first stab at a criterion, postulating that events are identical if and only if they have exactly the same causes and effects. However, it is questionable whether this criterion supports what even Davidson himself wants to hold about the identity of specific events. He gives the example of Jones, who apologizes by saying, "I apologize." While introducing this example, Davidson suggests, rightly I think, that Jones's apology and Jones's utterance "I apologize" are the same event. However, by his criterion, these events are probably distinct. Some set S of events may have caused Jones to realize that he needed to apologize, but why did he decide to apologize

by saying, "I apologize," rather than by saying, "I'm sorry," or by leaving a note? It seems likely that the specific form of his apology was partially caused by at least one event not in S.

More importantly, as Quine (1985) points out, and as Davidson (1985) later concedes, the criterion is circular. The circularity becomes obvious when we express the criterion formally: events x and y are identical if and only if ((z) (z caused x $\iff$ z caused y) and (z) (x caused z $\iff$ y caused z)), where z ranges over all events. In Davidson's criterion, event identity is defined by quantifying over all events. But in order to carry out this quantification, we must already be able to individuate events. Hence, Davidson's criterion fails.

Quine manages to simultaneously give both a clearer proposal for just what an event is and a better criterion for identity. He suggests that an event is just a time-slice of a physical object. Actually, for Quine, the "material content of any portion of space-time," no matter how disconnected, counts as an object. But agreement with this conception of objecthood is not crucial to understanding the basic idea behind Quine's definition of an event, phrased as I have put it above. The criterion of identity follows naturally from Quine's definition of event; events x and y are identical if and only if they are the same time-slice of the same object.

Davidson poses a dilemma for Quine's criterion of identity. He gives the example of a metal ball that simultaneously heats up and rotates, with the duration of each of these transformations completely filling the same stretch of time. Quine's criterion requires him to say one of two things. First option: there are two distinct objects that fill the region of space-time occupied by the ball during its heating and rotation; one of the objects heats up but does not rotate, while the other rotates but does not heat up. This option violates so many of our intuitions about physical objects that it is surely hopeless. Second option: the heating of the ball and the rotation of the ball are one and the same event. Quine bites the second bullet quite willingly, claiming that he is "not put off by the oddity of such identifications" (167).

But the oddity becomes more disturbing when we remember that events are related through causal dependence. For example, I might say, "I was

punched in the face yesterday, so that caused me to have a black eye today." But I could also say, grammatically if not eloquently, "My receipt of a punch in the face yesterday caused my possession of a black eye today," thus making the events explicit. If our criterion of identity for events is to be valid, it must certainly be compatible with our notion of causal dependence.

Now suppose we place a thermometer near the ball. We then heat and rotate the ball, and the thermometer rises. Someone may now ask us what caused the thermometer to rise. Clearly, one proper response is

(1) The heating of the ball caused the rise of the thermometer.

But for Quine, "the heating of the ball" and "the rotation of the ball" refer to the same event. Hence (1) will remain valid if we substitute the second of these expressions for the first. But then we are forced to agree that

(2) The rotation of the ball caused the rise of the thermometer.

Quine glibly concedes this point, but it seems deeply counterintuitive. The ball's rotation, it seems, has nothing at all to do with the rise of the thermometer.

Kim (1973) offers an alternative criterion of event identity that avoids this implausible result. His criterion arises from a characterization of events distinct from Davidson's and Quine's. Kim defines an event to be a "complex" consisting of an object, a property, and a time, such that the object exemplifies the given property at the given time. The dependence of events on their constituent properties gives Kim a way of accommodating our intuition about the distinctness of the heating and the rotation of the ball. Clearly, "heating up" and "rotating" are distinct properties; for one thing, they have different extensions. Hence Kim finds, as we hoped, that the heating of the ball and its rotation are two distinct events.

But Kim's criterion goes too far. Consider another example of Davidson's: Brutus killed Caesar by stabbing him. Hence we have the two events:

(3) Brutus's killing of Caesar

(4) Brutus's stabbing of Caesar

The properties "killing Caesar" and "stabbing Caesar" are presumably distinct, although they may actually have the same extensions. If we consider their extensions in other possible worlds, it is clear that they cannot be the same property. So Kim is obligated to declare (3) and (4) to be two distinct events. But this is quite hard to admit, since (4) seems to be merely a more specific description of (3). Indeed, suppose we told someone that Brutus killed Caesar. They might say, "Tell me more about the killing," to which we could surely respond, "The killing was a stabbing." This simple, natural exchange reveals the identity of (3) and (4).

So what went wrong with Kim's criterion? It seemed like a good way to eliminate the false causal claims we derived using Quine's criterion. We managed to separate the heating up and rotating of the ball into two distinct events, and we did this by pointing out that *heating up* and *rotating* are distinct properties. But this expedient gave us distinct events where we didn't want any. Indeed, Kim's property criterion implies the existence of a probably infinite number of distinct, yet incredibly similar events at every object at every point in time. Consider:

(5) I had the property of eating lunch at 12:05 today.

(6) I had the property of eating lunch and of being American at 12:05 today.

(7) I had the property of eating lunch and of being American and of believing that the Beatles are better than the Beach Boys at 12:05 today.

(5)-(7) are all true, and the three properties attributed to me by these sentences are surely distinct. Hence according to Kim's criterion, these three

distinct events all occurred simultaneously at 12:05 today. It seems to strain our concept of event to the breaking point, however, to believe that (5)-(7) really assert the existence of distinct events. We can continue constructing sentences of the form (5)-(7), by adding more properties true of me, in order to create an infinite number of distinct, yet simultaneous events. And what is so unsettling about this infinitude of events is that they all pertain to the same chunk of space-time, and indeed, to exactly the same action. How then can they be distinct?

Can we find a middle ground between the dearth of events under Quine's criterion of identity and the surplus of events under Kim's criterion? In order to sketch my proposal for the criterion of identity for events, which was suggested to me by Bennett (1996), I want to begin by examining the everyday activity of reporting an occurrence. There seems to be a difference between the way we would tell the story of the metal ball and the story of Caesar's assassination, and the difference is not only one of enthusiasm. In recounting the saga of the ball, we would probably say something like "The ball heated up and rotated at the same time," noting the heating and the rotation conjunctively. But in telling of Caesar's death, we would most likely say "Brutus killed Caesar by stabbing him," linking the two verbs together more closely through the use of the word "by." And these methods of recounting cannot be interchanged. Hence, we would not say "The ball heated up by rotating" or "Brutus killed and stabbed Caesar at the same time."

I am not suggesting that such a grammatical difference can always be used to tell when we have two events and when we have only one. Indeed, Bennett effectively demolishes the idea that "locution can be used as a criterion for individuation." And even if we could find a constant grammatical feature of sentences describing, say, two distinct events, that seems to leave undone the deeper philosophical task of explaining what metaphysical feature of events gives rise to this grammatical pattern.

However, I think that we can formalize the difference between the ball story and the Caesar story if we realize that physically modeling an occurrence is a formalization of recounting it. We simply replace words with

equations. Now, when modeling physical processes, we often analyze components of the process separately. Bennett gives the example of a cannonball arcing its way over the city wall while simultaneously rotating on an internal axis. We can completely describe the motion of the ball by writing down an equation that describes its parabolic path relative to the earth and another equation that describes its rotation around an axis intersecting the ball and traveling with it. Then if we want to find the motion of any particular point inside the ball, we need only combine these equations in the proper way.

I claim that this procedure of analyzing a physical process by components is the formal equivalent of recounting the occurrence of two simultaneous, distinct events. Here I use the word "components" in a technical sense, as in the components of a vector. The description "the cannonball flew on a parabolic path, and it rotated about its own axis" becomes the two equations describing the components of the ball's motion. Similarly, in our example of the metal ball, we may describe the rotation of the ball by one equation, and its rise in temperature by another. But notice that no such analysis by components is possible in the case of Jones's apology or Caesar's assassination.

Hence I propose that two events *x* and *y* are identical if and only if they cannot be physically analyzed as separate components of a process. Consider once again the assassination of Caesar. This physical process cannot be analyzed as two components, one the stabbing of Caesar and one the killing of Caesar. Why is this? I cannot answer, "Because they are the same motion," since the parabolic path and rotation of the cannonball comprise one motion, and yet that motion can be analyzed as components as described above. I confess I can answer nothing other than, "Because the stabbing of Caesar *is* the killing of Caesar." That this is my only answer increases my confidence that the proposed criterion captures something important about the nature of event identity.

Although the proposed criterion deals correctly with all of the examples given in this essay and still others that I invite the reader to dream up, some worries may remain. The criterion may seem to still imply the existence of

too many distinct events. In the cannonball example, it seems correct to say that the cannonball's traveling of the parabolic path is a distinct event from its rotation. But the criterion also implies that its horizontal movement is a distinct event from its vertical movement. Some may have the intuition that these are not distinct events. However, I would argue that there is an at least equally strong intuition that they are distinct events, and this intuition is brought out by an argument from causal dependence like the one we used to dispense with Quine's criterion. If we are asked what caused the cannonball's entrance into the city, it seems correct to say that its horizontal movement alone caused its entrance into the city. (Actually, vertical movement would also be required to get the ball over the city wall, so suppose now that the city wall has been knocked down by the invading army.) But it would be wrong to say that the ball's vertical movement caused its entrance into the city. Just as with the example of the metal ball, we can avoid this second, false causal claim if we hold that the horizontal and vertical movements of the ball are distinct events.

We should not be surprised that the proposed criterion deals properly with our intuitions about causal dependence. After all, one of the consequences of being able to analyze a physical process by components is that each component generates separate predictions. Each component gives rise to a set of events that are causally dependent on that component. For example, just by calculating the cannonball's parabolic motion, we know that it got over the city wall. We don't need to know anything about its rotation to predict this. But this implies that in saying what caused the cannonball's arcing over the city wall, we need only make reference to its parabolic motion, not its rotation.

Finally, I want to return to an issue I raised earlier. I argued that a mere grammatical pattern does not provide the kind of metaphysical understanding a theory of events should provide. But then I went on to base my proposed criterion upon a supposed formalization of grammar. Does this mean that this criterion also fails to illuminate the nature of events in the desired way? I do not think that it does. In moving the criterion from the realm of grammar to the realm of science, we have achieved a degree of objectivity

not available in language. Analysis of a process by components is an elementary yet extraordinarily useful technique in physics, a theory which has had phenomenal success in predicting and explaining events in the physical world. It seems plausible that in order to have achieved this success, physics must rest on a nearly correct, if implicit, metaphysics. Hence the fact that our intuitions about events accord with a natural categorization in the language of physics suggests that not only does the proposed criterion give necessary and sufficient conditions for the identity of events, but that it also captures the underlying metaphysics of those events.

Works Cited

Bennett, J. "What Events Are." *Events*. Ed. Casati, R., and A. C. Varzi. Aldershot: Dartmouth, 1996.

Ekman, Paul. "Moods, Emotions, Traits." *The Nature of Emotion*. Ed. Paul Ekman and Richard Davidson. New York: Oxford University Press, 1994.

Davidson, D. "The Individuation of Events." *Essays in Honor of Carl G. Hempel*. Ed. N. Rescher. Dordrecht: Reidel, 1969.

———. "Reply to Quine on Events." *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Eds. Lepore, E. and B. Mc Laughlin. Oxford: Blackwell, 1985.

Kim, J. "Causation, Nomic Subsumption, and the Concept of Event." *Journal of Philosophy* 70 (1973): 217-236.

Quine, W.V. "Events and Reification." *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Eds. Lepore, E. and B. Mc Laughlin. Oxford: Blackwell, 1985.

Ockham's Modal Moves: Crossing the Threshold of Modernity

José Luis Fernández
Purchase College, State University of New York

The Blumenberg-Löwith debate over the secularization hypothesis has attracted the attention of scholars interested in the history of ideas. In this paper, Fernández draws from Hans Blumenberg's *The Legitimacy of the Modern Age* in order to suggest how appeal to modality, in the form of William of Ockham's notion that God's absolute power entails the radical contingency of the universe, helped set the stage for the crossing of the epochal threshold of modernity. In the first section, Fernández gives a preliminary background to the famed dispute over the secularization hypothesis. In the second section, he argues that Blumenberg's thesis of man's self-assertion of reason trades on Ockham's notion of radical contingency. And in the third section, Fernández suggests how notions of God's absolute power helped to replace ideas of "providence" with "progress," thus ushering in the Modern Age.

The Paternity Suit: Epochal Authorship

Karl Löwith's *Meaning in History* argues for what is known as the 'secularization hypothesis': namely, that the notion of progress in modernity is a secularization of Christian eschatology (see *infra*, fn 4). In *The Legitimacy of the Modern Age*, Hans Blumenberg offers a critical challenge to Löwith's argument, specifically disagreeing with Löwith's contention that the idea of progress within history may be credited to expectations of future fulfillment.

On Löwith's account, the development of systematic inquiry or *Wissenschaft* so characteristic of the modern age was driven by a Christian teleology, aiming for worldly deliverance:

> The ideal of *modern science* of mastering the forces of nature and *the idea of progress* emerged neither in the classical world nor in the East, but in the West. But what enabled us to remake the world in the image of man? It is perhaps. . . *the hope in the future kingdom of God*. . . (203, emphasis mine)[1]

In this respect, Löwith's account appears quasi-Hegelian, as his teleological view of history shares various similarities with Hegel's realization of *Geist*.[2] Indeed, like Hegel, Löwith allows for the influence of chance and contingency in the formation of historical destinies.[3] Moreover, Löwith's account takes chance happenings in history as following from inevitability:

[1]Pippin states, "[Löwith] is not claiming that the modern notion of progress is Christian eschatology" (270). However, *pace* Pippin's analysis of the issue, his view seems mistaken, at least by the standards of textual evidence. Eschatology, which derives from the Greek, literally means "discourse about last things." Moreover, the Christian usage of eschatology deals exclusively with the coming of the kingdom of God, which will transform history by transcending and ending it. Therefore when Löwith, as in the passage above, suggests that the secular idea of progress is perhaps caused by the hope for God's kingdom, he is *ipso facto* stating what Pippin has him disclaiming.

[2]See Löwith 1964: 32: "Hegel completes the history of the spirit in the sense of its ultimate fulfillment, in which everything which has taken place hitherto or has been conceived is comprehended in a unity; but he completes it also in the sense of an eschatological end, in which the history of the spirit is finally realized."

[3]See Löwith 1949: 199-200. I realize that this claim is controversial, as there is some dispute among Hegel scholars over whether Hegel is a rigid teleological determinist or whether he is in fact an historical contingentist. My own reading of Hegel reveals that he is not a determinist, though we can plainly find textual evidence for the view that world historical events represent the necessary unfolding of Spirit through time. However, I also find that Hegel is not a contingentist, though, again, we can find examples where he states that chance occurrences do in fact take place in historical events and affect its course. Therefore, I believe that Hegel allows for compatibility between these extremes. Hence, my reading is amenable to the views espoused by Dieter Henrich, "Hegel's Theory about the Coincidence." *Kant-Studien*, 50 (1958/59), and Stephen Houlgate, "Necessity and Contingency in Hegel's Science of Logic," *The Owl of Minerva*, 27.1 (1995): 171-186.

> In the reality of that agitated sea which we call "history," it makes little difference whether man feels himself in the hands of God's inscrutable will or in the hands of chance and fate. *Ducunt volentum fata, nolentem trahunt*, could easily be translated into terms of a theology which believes that God works not only through those who obey his will but also through those who perforce serve him against their will (199).[4]

Ideas of preordination and fate are pivotal to Löwith's secularization thesis, as it turns on the teleological orientation of historical progress. We may attempt to provide a catchphrase for this view in rather un-Kantian terms: purposiveness *with* a purpose. The following distillation from Löwith's "Introduction" provides us with a glimpse of his idea of purposiveness:

> [The] philosophy of history is... entirely dependent on theology of history, in particular on the theological concept of history as a history of fulfillment and salvation... The same is true in regard to the formal structure of history. History, too, is meaningful only by indicating some transcendent purpose beyond the actual facts. But, since history is a movement in time, the purpose is a goal... If we reflect on the whole course of history, imagining its beginning and anticipating its end, we think of its meaning in terms of an ultimate purpose (1, 5).

Löwith regards the modern age as nothing more than an epoch forgetful of its historical goal: salvation.[5]

The Blumenberg-Löwith debate, following the articulation of Löwith's secularization hypothesis, is a battle over what I call *epochal authorship*.

[4]The Latin in this passage is a textual example of the Chinese box within a box motif, as Löwith is quoting Seneca, who in turn is quoting Cicero: "Fate leads the willing, and drags the unwilling." See Seneca, "Letter CVII, 11," *Letters from a Stoic*, (London and New York: Penguin, 1969), p. 200.

[5]Ibid., p. 18. "Not only does the *eschaton* delimit the process of history by an end, it also articulates and fulfils it by a definite goal." See also p. 114: Man's "expect[ing] the eschaton not only in history but eventually from it" is an expression of man's hope for salvation during lifetime rather than after death.

On one side, Löwith discounts the modern idea of progress due to its secularization of the Christian eschatological pattern. On Löwith's view, the desires of modernity are merely a masquerading version of what came before: namely, the Christian yearning for worldly deliverance.[6] On the other, Blumenberg defends the legitimacy of the modern age with his thesis of man's "self-assertion of reason."[7]

Blumenberg disagrees with Löwith's genetic account thusly:

> The eschatological future had…lost its connection with the blessings of salvation that had already been conveyed to redeemed mankind. Consequently, the basic eschatological attitude of the Christian epoch could no longer be one of hope for the final events but was rather one of fear of judgment and the destruction of the world…Early Christianity found itself in what was…the difficult position of having to demonstrate the trustworthiness of its God to an unbelieving surrounding world not by the fulfillment of His promises but by the postponement of this fulfillment…For the Middle Ages there was both a *cosmic* and an *individual* eschatology. This split made it inevitable that man's interest would be absorbed by the question of his own "last things." (44-46)[8]

Blumenberg points out that while it may well be the case that medieval eschatological expectations might have been built by a long progression of soteriological studies and hopes, in due course, this prospect turned into a

[6]Martin Jay notes, "Borrowing the old trick of early Christian polemicists, who accused the ancient Greeks of having secretly stolen their best ideas from the Bible, the secularization theorists located the unacknowledged paternity of the modern in originally religious ideas. Illegitimacy comes therefore not from lacking a proper parent, but from denying his generative power." (Jay, 184)

[7]Blumenberg, p. 138: "Thus 'self-assertion'…means an existential program, according to which man posits his existence in a historical situation and indicates to himself how he is going to deal with the reality surrounding him and what use he will make of the possibilities that are open to him."

[8]See also Arendt 1958: 73-74 for an explanation of the tie between "the obvious nonfulfillment of…eschatological hopes" and the *vita activa*.

crisis of grace.[9] Indeed, his refutation of Löwith's thesis turns on a radical repositioning of the medieval mindframe due to a loss of confidence in the grace of salvation: "Since He has fixed the eternal judgment after the end of the world, He does not carry out the separation presupposed by that judgment before the end of the world"[10] (44).

Commenting on this shift, Robert M. Wallace notes, "Both salvation and the creation had thus been deprived of all accessible meaning and reliability" (Wallace: 63). No longer able to rely on the prevailing world-view, man turns his faith and focus not towards the conceptually opaque and perhaps spiritually unreliable, but rather to the immediate and dependable: namely, himself and his world. Frank B. Farrell provides an excellent analysis of this turn:

> Man is no longer at home in a meaningful universe in which he has a proper place, in which it is something about him and what he is that gives him a role in God's dealing with creatures. The radical contingency of the world is matched by a radical insecurity of the self... Instead of trying for a theoretical understanding of the metaphysical character of the universe (an understanding that has become impossible), [man] shall instead turn to what [he] can take reality to be through [his] own labor and construction. (Farrell : 153)

Thus, it is Blumenberg's contention that these frustrations led man to tailor his actions not towards taking part in a chance, merely possible, and uncertain dwelling in God's kingdom, but rather towards the project of "self-assertion."

While making his case for "self-assertion," Blumenberg often mentions the medieval appropriation and transformation of ancient theories and postulations, e.g., from Platonic beliefs, and Aristotle's idea of contemplation

[9] "Eschatology may have been, for a shorter or longer moment in history, an aggregate of hopes; but when the time has come for the emergence of the idea of progress, it was more nearly an aggregate of terror and dread." (31)

[10] Blumenberg here quotes the second and third century ecclesiastical writer Tertullian.

as the highest human fulfillment.[11] On Blumenberg's account, theoretical inconstancies are needed in order to ensure that the wheels of progress keep spinning. New problems that need solving must arise or be resurrected. As new arrays of ideas and beliefs come to overhaul the ones of the *ancien regime*, the transformation, or what Blumenberg calls "the re-occupation," of conceptual schema ushers in a period of epochal change. One such paradigm shift is found in the epochal threshold between the medieval and the modern world, during which man ascribed to the world the infinite creative power previously reserved for God. This re-allocation of power does violence to the old relation between God, man, and world. As Elizabeth Brient points out:

> The cosmos of the Middle Ages is a finite, well-ordered whole, a closed hierarchy, whose order and value…is granted by an infinite and benevolent God. In the transition to the modern age, the world comes to "acquire" the divine attribute of infinite being, but only at the price of destruction of this ancient order. (Brient: 98)

The destruction of the ancient order has the effect of bringing on the existential mode of *unheimlichkeit*, what both Heidegger[12] and Freud[13] call, respectively, "unhomey" and "uncanny": a feeling of radical unfamiliarity or being not-at-home. Reality presents itself as indifferent and arbitrary to the human individual. The traditional metaphysical framework that shaped man's perception of reality, ontologically grounded in God, is now replaced

[11] See Aristotle (2001):1104: "If happiness is activity in accordance with virtue, it is reasonable that it should be in accordance with the highest virtue…this activity is contemplative."

[12] Heidegger uses the term *unheimlich* or "unhomey" to express *nichtzuhausen-sein* or not-being-at-home. Used in this way, *unheimlich* is a state of mind denoting feelings of anxiety over alienation and unfamiliarity. For our purposes, man's unfamiliarity with his new metaphysical picture of the world. See Heidegger: 233.

[13] Following Schelling's definition of *unheimlich* or "uncanny," Freud states that *unheimlich* is the name for everything that "ought to have remained hidden and secret, and yet comes to light" (Freud: 224). For our purposes, the coming to light of the extreme intensification of God's transcendence. 1953-74

by one in which the term "reality" simply comes to mean an actual state of affairs.

Blumenberg tracks down this "disappearance of order" (139) to, in general, late medieval nominalism, and, in particular, to the thoughts of William of Ockham. In what follows, I will focus on the relation between God, man and world on the stage set by Ockham's modal thought, which, if not set forth, may have delayed the epochal threshold to modernity for later centuries.

The Ockhamite Origins of Modernity

Blumenberg suggests how late medieval nominalists were joined in defending the idea of God's absolute power from any "immanent laws" (160-161); that is, from apparent nomological laws of necessity. God, the *sui generis* necessary being, cannot be tied to the contingent laws governing nature. Subsequently, the created world is to be understood not principally as an expression of God's omnibenevolence or of His omniscience, but rather of His omnipotence.

As we have seen, man's loss of confidence in God's redemptive plan already led to his viewing reality in terms of its utter *statu*. Producing a similar, though arguably more terrifying, effect, Ockham elevated God's omnipotence to radical predominance, implying the radically *contingent* nature of existence. Blumenberg's use of the Ockhamite influence on modernity is complex and subtle, but not very elaborate. By way of adding a supplement to Blumenberg's reading, I will note some of the central ideas held by Ockham on modality and God's absolute power.

Prior to Ockham's theory of modality, the prevailing view of modality was dominated by Aristotle's affirmation of actuality over potency or possibility.[14] We see this idea at work in the notion that no genuine possibility can remain forever unactualized, what Arthur Lovejoy coined as "the

[14] "[I]t is clear that actuality is prior to potency." See Aristotle, *Metaphysics*, IX, 8, 1049b5-1050a18, pp. 828-31, in Mckeon, *The Basic Works of Aristotle*, ed. Richard McKeon, trans. W.D. Ross, (New York: The Modern Library, 2001).

principle of plenitude" (Lovejoy: 52).[15] In contrast, Ockham's theory of modality took the notion of possibility not as leading to a necessary actual state of affairs, but rather as involving a consideration of alternative states of affairs. According to Ockham's theory, possibility just names a state of affairs whose formulation is not a logical contradiction. As such, there is nothing necessitating that every possibility be actualized at some time or other.[16]

Ockham maintains that God's will is bound only by the principle of non-contradiction, i.e., for Him anything is possible as long as it does not violate this principle.[17] However, this idea needs some unpacking, as it, *prima facie*, suggests that God's power is limited. God, if He exists, is either omnipotent or not omnipotent, but not both: God's omnipotence entails, it seems, that he is *not* not-omnipotent. As such, it seems reasonable to think that the law of non-contradiction *does* apply to God.

If one is going to say that the laws of non-contradiction do *not* have application to God, then one will have to admit that (i) his statement that God is omnipotent *could* be true, while (ii) it is *equally* true that God is not omnipotent; indeed one will have to admit (iii) that it is equally true that God is without any power at all. Therefore, if the laws of logic do not apply to God, then it follows that even if it is true that God is omnipotent, God could still be limited in power or completely powerless.

We see, then, that on the one hand, by understanding Ockham's assertion that God is bound only by the law of non-contradiction, if we are to say anything *meaningful* about God, then we have to assert something that, if true, rules out such ideas as God's power being circumscribed or deficient. On the other hand, if the laws of logic do not apply to God, then nothing we say about God, even if true, will rule out anything. Hence, if we accept the

[15] See also Sorabji: Sorabji argues, contra W.D. Ross, Jaakko Hintikka and some others who hold Aristotle to be a strict determinist, that Aristotle did not give wholesale acceptance to the principle of plenitude. Rather, "he accepted the principle of plenitude only for a very restricted range of cases…" (128).

[16] See Ockham 1988: 314-36.

[17] See Ockham 1990: 25: "Anything is to be attributed to the divine power, when it does not contain a manifest contradiction."

notion that God is *not* bound to the law of non-contradiction, we will not have succeeded in making a meaningful claim about what God's properties *might* be like. Therefore, if we want to make any meaningful assertions about God at all, it seems that we must at least commit ourselves to the claim that the laws of logic have application to God.

Ockham further states that God can do anything a secondary cause can do; subsequently, no event or act, at any time, determines the next by any intrinsic, internal or logical necessity. Recall that God's absolute power has no necessary tie with physical or nomological necessity. God can produce Y with or without need of X. The lawful, causal structure of the created world does not interfere with God's omnipotent agency. For He can choose at any time to override or bypass the natural order. God is, at any time, capable of interrupting the natural causal sequence of events and directly producing any given effect.[18] Here we see Ockham countering the Thomistic view that since God is the creator of the world He transfers some of His own necessity to the natural world. For Thomas, the patterns we apprehend in nature, what he calls "intelligible species," represent the designs of the Creator, and it is through His will that we gain knowledge through "active intellect."[19] Ockham's answer to Thomas is that "intelligible species" are merely mental concepts or linguistic terms abstracted by human beings. God is indeed the final cause of everything, but unnecessary metaphysical complications do not help us to understand the natural world. Hence, on Ockham's view, theology (founded on God, necessity, etc.) and the natural sciences (founded on natural laws, contingency, etc.) must go their separate ways.

Potentia Absoluta and *Potentia Ordinata*: Progress, not Providence

The world picture that emerges from Ockham's conception of modality is one that is radically *contingent* on a Divine will. Furthermore, God's will is not bound by any nomological necessities that would somehow have influ-

[18] Ibid., p. xx.

[19] See MacDonald: 161.

ence on His actions. Subsequently, the natural, physical order of the world is neither an absolute nor *necessary* order (the way things must only be), but merely an *ordained* order (the way things are out of numerous alternative ways of being). The distinction is subtle, as there is a fine difference between the absolute and ordained Divine will.

Hence, we ought to note that God's *potentia ordinata*, i.e., the way things *actually are*, is only one expression of His *potentia absoluta*, i.e. the way things *could be* at any given time.[20] Subsequently, the created world comes to be seen as not tied to God's creative power through any *necessary* relation. Indeed, in terms of ontology, as much as this world is, it could just as easily *not have been*:

> The contingency of creation is the true meaning of Ockham's account of *potentia absoluta* and *potentia ordinata*. Everything other than God exists contingently. Since the world is not logically necessary, God equally could have chosen not to create it.[21]

Consequently, an unlimited number of possible worlds[22] correspond to God's *potentia absoluta*, with God's *potentia ordinata* representing the divine choice of one actual world. Moreover, it is futile trying to rationally account for God's creative choices.[23] For God's *potentia absoluta* not only releases Him from his own ordinances in the nomological order, but it also frees Him from the order of grace and human intelligibility. For instance, in the natural order, God's absolute power means that His actions need not be subordinated to His own laws, e.g., during the performing of miracles.

[20] Respectively, "ordered or ordained power" and "complete, absolute power," Blumenberg: 153.

[21] See Osler: 30.

[22] Blumenberg: 160-61.: "To the *potentia absoluta*... corresponded an infinity of possible worlds, but no infinity of actual worlds was allowed to correspond to it."

[23] Ibid.: 154: "Ockham's distinction between the *potentia absoluta* and the *potentia ordinata* does not alleviate the situation for rationality because although it does imply that once chosen, the *ordo* [order] will be observed, it does not provide any access to the contents of the chosen order."

Furthermore, God could easily choose to destroy this world and create a radically different one, with a set of natural laws that we could not imagine. Moreover, while we have no choice but to abide by the rules for human salvation that God ordains, He is not obligated to keep to the promise of these rules. Conceivably, He can save people who do not follow His rules. The fact that we do follow these rules gives us no purchase on salvation. As I touched upon earlier, with an awareness of these radical notions, man begins to feel not-at-home, his world being only one state of affairs, just "the 'facticity' of reality" (138).

Blumenberg's thesis argues that it was precisely this *unheimlichkeit* or alienating quality of the world that opened up new conceptual and existential possibilities: possibilities not having to do with God, but rather having to do with man's self-assertion of reason. Such self-assertion of reason became the characteristic of the modern age not as Löwith would have it, i.e., as a product of secularization, but rather, as noted in Farrell's observation (*supra*, p. 4), by the fact that it became all but impossible to choose the alternative of salvation. As soon as salvation became tied to the arbitrary will of God, its function broke down inside the old paradigm between God, world and man. For these reasons, Blumenberg contends that man turned his gaze not towards heaven and the redemptive promise, but rather peripherally towards the new mechanistic science. Notions of progress usurped the role once held by Providence. Blumenberg maintains that this break in the old eschatological pattern welcomed in the naturalistic epoch in a historical sense, and the modern epoch in a conceptual sense. In both of these accounts, Blumenberg's use of Ockham's modal theology provides a strong challenge to Löwith's secularization thesis. When notions of God's *potentia absoluta* began to entail the radical contingency of the universe, the cord tying notions of a transcendent necessary determinism with the unfolding of contingent historical occurrences was cut, thus ushering in the Modern Age.

Works Cited

Arendt, Hannah. *The Human Condition*. Chicago: The University of Chicago Press, 1958.

Aristotle. *Metaphysics*. *The Basic Works of Aristotle*. Ed. Richard McKeon. Trans. W.D. Ross. New York: The Modern Library, 2001.

———. *Nichomachean Ethics*. *The Basic Works of Aristotle*. Ed. Richard McKeon. Trans. W.D. Ross. New York: The Modern Library, 2001.

Augustine. *Confessions*. Trans. R.S. Pine-Coffin. London and New York: Penguin, 1961.

Blumenberg, Hans. *The Legitimacy of the Modern Age*. Trans. Robert M. Wallace. Cambridge: MIT Press, 1983.

Brient, Elizabeth. *The Immanence of the Infinite: Hans Blumenberg and the Threshold of Modernity*. Washington: Catholic University of America Press, 2002.

Colish, Marcia C. *Medieval Foundations of the Western Intellectual Tradition*. New Haven: Yale University Press, 1997.

Farrell, Frank B. *Subjectivity, Realism, and Postmodernism: The Recovery of the World in Recent Philosophy*. Cambridge: Cambridge University Press, 1994.

Freud, Sigmund. *Das Unheimlich*. *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Ed. and Trans. James Strachey. London: Hogarth Press, 1953-74.

Heidegger, Martin. *Being and Time*. Trans. John Macquarrie and Edward

Robinson. New York: Harper & Row, 1962.

Houlgate, Stephen. "Necessity and Contingency in Hegel's Science of Logic." *The Owl of Minerva*. 27,1, Fall 1995.

Jay, Martin. "Review Essay." *History and Theory*. 1985, Vol. 24.

Kant, Immanuel. *Critique of the Power of Judgment*. Ed. and Trans. Paul Guyer. Cambridge: Cambridge University Press, 2000.

Lovejoy, Arthur O. *The Great Chain of Being*. Cambridge, Mass: Harvard University Press, 1936.

Löwith, Karl. *Meaning in History: The Theological Presuppositions of the Philosophy of History*. Chicago: University of Chicago Press, 1949.

———. *From Hegel to Nietzsche: The Revolution in Nineteenth-Century Thought*. New York: Columbia, 1964.

MacDonald, Scott. "Theory of Knowledge." *The Cambridge Companion to Aquinas*. Cambridge: Cambridge University Press, 1993.

Osler, Margaret J. *Divine Will and Mechanical Philosophy: Gassendi and Descartes on Contingency and Necessity in the Created World*. Cambridge: Cambridge University Press, 1994.

Pippin, Robert. "Blumenberg and the Modernity Problem." *Idealism As Modernism: Hegelian Variations*. Cambridge: Cambridge University Press, 1997. Originally appeared in Review of Metaphysics, 40.3 (1987): 535-57.

Sorabji, Richard. *Necessity, Cause, and Blame: Perspectives on*

Aristotle's Theory. Ithaca: Cornell University Press, 1980.

Wallace, Robert M. "Progress, Secularization, and Modernity: The Blumenberg-Löwith Debate." *New German Critique* 22 (1981): 63-79.

William of Ockham. "Modal Consequences." *The Cambridge Translation of Medieval Texts: Volume One, Logic and the Philosophy of Language*. Eds. Norman Kretzmann and Eleonore Stump. Cambridge: Cambridge University Press, 1988.

———. "Prologue to the Ordinatio, q. i, N ssq." *Philosophical Writings*. Ed. and Trans. Philotheus Boehner. Indianapolis: Hackett, 1990.

Phenomenal Externalism
Cross-Modal Matching and the Threat of Epiphenomenalism

Wesley H. Holliday
Stanford University

Phenomenal externalism, or the externalist version of representationism, is the view that the phenomenal character of experience—how the world *seems* to you—is exhausted by the intentional content of experience—how your experience *represents* the world. In the 1970's, Hilary Putnam argued for *externalism about meaning*: what a person means by a word is in part dependent on her external environment, such that molecular duplicates whose nervous systems are in the exact same states can nonetheless mean *different things* by the same word, just by virtue of a difference in their environmental conditions. The lesson drawn by externalists is that *meaning ain't in the head*, since it can change without any change in the brain. The phenomenal externalist wishes to say something similar about experience—it is in part dependent on the environment, such that molecular duplicates in the same states can nonetheless have *different experiences*, just by virtue of a difference in their environmental conditions. *Phenomenolgy ain't in the head* either, phenomenal externalists maintain, and can also change without any changes in the brain. Some philosophers find this view attractive since it obviates various traditional problems with locating phenomenology *in* the head. However, there have also been serious objections to the view. Holliday reviews Ned Block's famous "Inverted Earth" objection to phenomenal externalism, considers several recent replies to the objection by Michael Tye and Fred Dretske and then makes the case that these replies fail when we consider an empirical, psychological phenomenon known as *cross-modal matching*. Indeed, the replies of Tye and Dretske, which try to negotiate both Inverted Earth and another infamous thought-experiment involving "Swampman," threaten to render the phenomenal character of experience *epiphenomenal—i.e.*, causally impotent—which seems to undermine the point of providing an externalist representationist theory of it.

Phenomenal Character and Phenomenal Externalism

Introspection plainly reveals the difference between a sensation of heat and a sensation of cold or between a visual experience of red and a visual experience of green. From the first-person perspective, these experiences differ in their *qualitative* characteristics (*qualia*, for short) or, to use some other philosophical terms of art, in their *phenomenological* or *phenomenal* characteristics. To say basically the same thing in still different language, there is a difference between *what it is like* to feel heat and to feel cold or to see red and to see green. Of course, the same goes for olfactory experiences, gustatory experiences, auditory experiences, and so on: these experiences can all fall into difference types by virtue of their distinctive qualitative or phenomenal properties. Whenever there is a difference in how two things *feel*, *appear*, or *seem* to a subject, there is a difference in the phenomenal properties of that subject's experiences.

Many philosophers have taken the phenomenal character of experience, the fact that there are these *raw feels*, as it were, to be a defining mark of consciousness. The question, which has long plagued philosophers, is how to find a place for such phenomenal consciousness in the natural world. Phenomenal properties seem to be very real properties (consider the immediate phenomenological quality of pain!). At the same time, they seem to be rather strange properties, insofar as it is not clear exactly how they fit in among the *physical* properties that we take to be most secure in our catalogue of the world's real properties. Initially, there may not seem to be any problem: to find the phenomenal properties, we just look inside the head, at the brain, of course. We may then *identify* the phenomenal properties of an experience with some physical properties of the brain, such as properties of neurons or other structures. However, as soon as we try to do so, it may seem that we come up empty-handed: if one hallucinates a rotating, red triangle, one's neurosurgeon cannot find a corresponding piece of one's brain that is rotating, red, and triangular, which present at least a *prima facie* complication for identifying the phenomenal characteristics of that experience with some physical properties in the brain. Faced with such puzzles, a few philosophers have endeavored to push the phenomenal properties of ex-

perience *outside of the head*, thereby avoiding problems with finding them there. As Michael Tye suggests:

> . . . *phenomenology ain't in the head*. Just as you cannot read semantics out of syntax, so you cannot read phenomenology out of physiology. This is why you cannot find any technicolor qualia, any raw feels, by peering around inside the brain (with or without a flashlight). They simply are not in there. To discover what it's like, you need to look outside the head to what the brain states represent. Phenomenology is, in this way, externally based. (Tye 1995: 151)

The view that Tye defends is what we may call *phenomenal externalism*, the view that the phenomenal character of experience does not just depend on the state of one's brain, but also depends on the *environment* outside of the brain (in a way to be spelled out below). Some philosophers have found this sort of view attractive, since it promises to sidestep various traditional difficulties with locating phenomenology inside the head. However, because it makes phenomenology externally based in some sense, phenomenal externalism has the consequence that people whose brain states are the same can nonetheless have *different types* of experiences, which is to say, experiences with different phenomenal characters, just in virtue of a difference in their environments. This stands in opposition to a thesis that many philosophers have regarded as fundamental, namely, the *natural supervenience* of phenomenal character on the brain: in our world, there can be no difference in the phenomenal character of experience without a difference in the brain. However, phenomenal externalism denies natural supervenience. Pushing phenomenology outside the brain may promise to sidestep difficulties with finding it there, but it also has a price, insofar as it leads to the denial of natural supervenience. It is this denial that is at the root of some of the serious objections that have been raised against phenomenal externalism.

In this paper, I review Ned Block's famous "Inverted Earth" (Block 1990) objection to phenomenal externalism, consider several recent replies to the objection by Michael Tye (1998) and Fred Dretske (1995a), and then

make the case that these replies fail when we consider an empirical, psychological phenomenon known as *cross-modal matching*. Indeed, the replies of Tye and Dretske, which try to negotiate both Inverted Earth and another infamous thought-experiment involving "Swampman," threaten to render the phenomenal character of experience causally irrelevant, or *epiphenomenal*, which seems to undermine the point of providing an externalist theory of phenomenology.

Varieties of Externalism

In the 1970's, Hilary Putnam argued for *externalism about meaning* or "semantic externalism," the thesis that what a person *means* by a particular word is in part dependent on the external environment in which she has used that word; this environmental dependence is supposed to have the consequence that people whose brain states are the same can nonetheless refer to *different things* by the same word, just by virtue of a difference in their environments. The idea, popularized by Putnam's famous "Twin Earth" examples, is that two molecular-duplicate twins can mean different things by their use of the word 'water,' if one lives on Earth, a planet covered by H_20, and the other lives on Twin Earth, a twin planet covered by XYZ, a chemical indistinguishable from H_20. Imagine an Earthling traveling to Twin Earth, pointing at the liquid falling from the sky and saying, "It's water!" Putnam suggests that the Earthling has accidentally misapplied his term 'water' in this case, due to the superficial similarity of H_20 and XYZ. The Earthling has always applied the term 'water' to the stuff that falls from the sky on Earth, so by 'water' *he means* the stuff, whatever it is, that falls from the sky on Earth; as it turns out, that stuff to which he has always referred is H_20 (though we may suppose that for whatever reason, such as his living in 1750, our Earthling does not know this). Now, of course, by "water," his *twin* on Twin Earth refers to XYZ, since *that* is the stuff to which he has always applied the term 'water.' The upshot is that the twins' statements about 'water' are supposed to refer to different things—are supposed to have different semantic content—even though the twins are molecular du-

plicates whose brain states are the same. Putnam's conclusion is that there is a kind of environmentally-dependent semantic content, "wide" semantic content, that does not naturally supervene on the brain. We may have thought that in our world there can be no difference in semantic content without a difference in the brain, but semantic externalism denies this.

Soon after Putnam advanced his arguments for semantic externalism, some philosophers extended externalism beyond the semantic content of words and statements and on to cover mental content such as belief and thought (see McGinn). Beliefs and thoughts have *intentional content*, meaning that they can be *about* the world and that they can *represent* the world to be a certain way (i.e., my belief that it is raining in Paris is about Paris and represents Paris to be a certain way); externalism about belief and thought, then, is the view that the intentional content of belief or thought may, like semantic content, be *wide* in some cases: intentional content may be dependent on the environment such that people whose brain states are the same may nonetheless have beliefs or thoughts with *different* intentional contents—they may represent the world differently—just by virtue of a difference in the thinkers' environments. Thus, when an Earthling drinking H_20 thinks *water is refreshing* and when a Twin Earthling drinking XYZ thinks *water is refreshing*, the intentional contents of their beliefs/thoughts are supposed to be different, since the one represents something about H_20 and the other represents something about XYZ.

Whether or not one finds these arguments for externalism convincing, the point is that recently the tide of externalism has swept even further, past semantics, past intentional mental states like belief and thought, and now even on to mental states that are paradigmatically phenomenal, like experiences. Thus, we come to the view of phenomenal externalism. Phenomenal externalism combines two theses. First, there is *strong representationism*: as with mental states like belief and thought, experiential states have representational/intentional content, in that they represent the world to be a certain way (this is representationism); moreover, the content of experiential states, including their phenomenal character, is *purely* representational/intentional (this is *strong* representationism). The second component

of phenomenal externalism is externalism about the mental content in question: the representational/intentional content of experience is *wide* content, meaning that people whose brain states are the same can nonetheless have experiences with different representational/intentional contents and thus, by strong representationism, different phenomenal characters, just by virtue of a difference in their environments. It follows from the conjunction of strong representationism about experience and externalism about experiential content that the phenomenal character of experience does not naturally supervene on the brain.

Representationism Plus Externalism

Let's spend a moment on the issue of *representation*. As we saw above, Tye suggests that to discover the place of phenomenal characteristics in the natural world, you cannot just look around in the brain with a flashlight; rather, "…you need to look outside the head to what the brain states represent. Phenomenology is, in this way, externally based (Tye 1995: 151). But what does the idea that phenomenology is "externally based" amount to? It seems the idea is that we should not expect to locate blueness *in the brain* with some phenomenal property that is to be identified with a physical property there (which we may have been tempted to do in order to explain, for example, how a yellow object may *appear* blue to a perceiver who's retinal connections have been suitably re-wired; *that* sort of blueness must be "in the head," one might think). The idea is instead to locate *blueness* as a property of an object in the external world, which the brain *represents* the object as having. How does the brain represent an object as having the property of blueness? A basic idea would be that if, under normal conditions, a brain state, S, is tokened in me *if and only if* there is a blue object in front of me and *because* there is a blue object in front of me, then that brain state, S, *is representing* that there is a blue object in front of me (Ibid.: 101).[1] If I have that brain state when and only when there is something blue in front

[1]Here and in what follows, I use color terms in such a way that saying that an object *is blue* is not saying that it *necessarily appears blue* to any perceiver; the perceiver may be color-blind or may have just had her retinal connections re-wired, such that 'blue' objects

of me that causes me to have that state, then we may say that that brain state *tracks* blueness; the intentional content of that brain state, what that brain state represents, is therefore blueness. And, as we noted above, strong representationism suggests that the phenomenal character of experience—how the object *appears* to you—is exhausted by the intentional content of the experience—the blueness of the object, in this case—which seems to work out.

According to the externalist or wide version of strong representationism, which I have been calling "phenomenal externalism" (following Dretske 1995b), the phenomenal character of experience is wholly identified with *wide* intentional content. The externalist move here is to say that just as Twin Earth examples showed that *meaning* is not entirely "in the head," *neither is experience*: both depend on the external environment. This dependence of experience on the environment amounts to the claim that the phenomenal character of experience "is constituted by the properties in the world that things are represented as having" (Dretske 1995a: 1). Since it is the properties *out there* in the world that things are represented as having that constitute the phenomenal character of experience, if there is a change in the environment, there is a change in experience, even if there is no physical change in the brain. This is the denial of natural supervenience that I have made much of already and that will be at stake in some of the arguments that follow.

Inverted Earth

Ned Block (Block 1990) has devised a thought experiment that is supposed to show how phenomenal character and wide intentional content (hereafter "intentional content") can come apart, thereby refuting phenomenal externalism. Imagine a planet called "Inverted Earth," which is exactly like Earth in all respects but one: everything on Inverted Earth is colored the complementary color of its counterpart on Earth. For instance, on Inverted Earth,

appear colorless or even a different color. Likewise, it is open whether an object that instantiates the property of 'blueness' appears blue or some other color to a particular perceiver.

roses are *green* and the sky is *yellow*. Now imagine Ned, who lives on Earth, and his molecular-duplicate twin, Twin Ned, who lives on Inverted Earth. What is interesting about Twin Ned (besides his being a molecular duplicate of Ned) is that he unknowingly wears inverting lenses in his eyes, inverting lenses that neutralize the neural effects of Inverted Earth's inverted color scheme: in fact, when Ned looks at the sky on Earth and when Twin Ned looks at the sky on Inverted Earth, we are to imagine that the same retinal stimulation occurs and that their brain states are the same (Block 1990: 65).

Consider Ned's experience of the Earth sky. The thought Ned might express by "the blue of the sky is very saturated" is about, mundanely enough, blue. The intentional content of Ned's belief about the sky includes the property of blueness. Furthermore, the perceptual state Ned is in when he looks at the sky normally tracks blue things (i.e. it is tokened in him when and only when he looks at something blue and because he looks at something blue), so the intentional content of his experience of the sky includes the property of blueness that the sky is represented as having.

Now consider Twin Ned on Inverted Earth (hereafter "IE"). The first thing to note is that although everything on IE is colored the complementary color of its counterpart on Earth, the Twin Earthlings still use the word "blue" to describe their yellow sky, only they mean something else by that word than Ned and the rest of us do. The same is true of Twin Ned: although Twin Ned is unlike the Twin Earthlings in that he unknowingly wears inverting lenses, he has grown up on Twin Earth talking just like the Twin Earthlings and so his semantics are supposed to be inverted relative to Ned's and ours as well, such that 'blue' for him means yellow (he uses that word to refer to objects that are yellow, like the other Twin Earthling), similar to the way that for Putnam's imagined Twin, 'water' means XYZ, not H_20. Likewise, the intentional content of Twin Ned's belief about the sky is different than Ned's: the thought Twin Ned might express by, "The blue of the sky is very saturated," is supposed to be about yellow (Ibid.: 66), not blue, similar to the way that the thought Putnam's Twin might express by, "Water is refreshing," is about XYZ, not H_20. Finally, the perceptual state Twin Ned is in when he looks at the IE sky normally tracks yellow things (i.e. it

is tokened in him when and only when he looks at something yellow and because he looks at something yellow), not blue things, so the intentional content of his experience of the sky includes the property of yellowness, not blueness. The intentional content of his experience of the sky is thus inverted with respect to Ned's.

Now, Block presses, is it also true that the *phenomenal character* of Twin Ned's experience is inverted with respect to Ned's, along with the relative inversion of their intentional content? That is, when Twin Ned looks at a 'blue' [yellow] sky on IE through his inverting lenses, is the phenomenal character of his experience really inverted with respect to the phenomenal character of Ned's experience when he looks at a 'blue' [blue] sky on Earth through no lenses? If phenomenal externalism is true, the phenomenal character must be inverted, since the intentional content is inverted; on this view, the IE sky *appears differently* to Twin Ned than the Earth sky appears to Ned, even though their brain states are the same when they look at the sky! But intuitively, Block says, it is clear that the phenomenal character does not invert in this way (Ibid.: 63): a *yellow* IE sky *seen through inverting lenses* appears the same to Twin Ned as a blue Earth sky seen through no lenses appears to Ned. Intuitively, there is no phenomenal inversion, despite the inversion in intentional content. Thus, phenomenal character cannot be identified with (wide) intentional content and phenomenal externalism is false.

Biting the Bullet on Inverted Earth

But is intuition enough to undermine phenomenal externalism? One externalist strategy here is to take what I will call the Bullet Biting approach to Inverted Earth: agree with Block that intentional content is relatively inverted, but argue against Block that the phenomenal character is also relatively inverted, despite powerful intuitions to the contrary. One externalist reply along these lines questions one's *self-knowledge* of phenomenal character: Fred Dretske argues that the phenomenal character of experience might well be different between molecular-duplicate twins, because know-

ing about the phenomenal character of one's experience collapses into having beliefs about one's experience and *beliefs* (as we were supposed to have learned above in the Twin Earth examples) can be environmentally dependent such that they are different even for molecular-duplicate twins; if the twins' beliefs about their own phenomenal characters can thus be different, why assume that their phenomenal characters must be the same? Another externalist reply questions one's *phenomenal memory* (this reply refers to a slightly modified version of the IE story in which, rather than comparing twin Neds, we are to compare Ned on Earth and Ned, later, on IE, after having been kidnapped, outfitted with inverting lenses, etc.): Michael Tye suggests that if Ned thinks that the sky looks phenomenally the same on IE as it did on Earth, perhaps he is just *misremembering*; after all, "external factors are relevant to the individuation of memory-contents" (Tye 1998: 465) and external factors are different on IE and Earth, so maybe the trip somehow interfered with his memory. Both of these replies rely on the following thought: even if the phenomenal character were the same on IE and Earth, *you could not know it*, so why hold onto the (brute?) intuition that it must be?

Dretske's Bullet Biting approach to Inverted Earth relies on a distinction between two ways that something can *look* to a perceiver. Imagine a red rose. I am looking at its red color and my cat is looking at its red color. To the cat, the rose looks red in the sense that this rose looks to it the way red roses normally look to it (this rose looks phenomenally the same as other red roses) and in the sense that this rose looks different to it than non-red roses (the cat can discriminate this rose from non-red roses by how they look phenomenally). Take this sense of 'look' to be the phenomenal sense of 'look,' look$_p$ (Dretske 1995a: 68). So the rose looks$_p$ red to the cat. Does the rose look$_p$ red to me too? It does, since it looks the way other red roses look to me and since I can tell it apart from non-red roses. However, the rose does not just look$_p$ red to me. I also describe (even if just to myself) the rose *as* looking red. I have the concept of RED and I take this rose to be red. In this case, the rose looks red to me in a *doxastic* (=belief) sense of 'look,' look$_d$ (67), a sense in which the rose does not look red to the cat,

which does not possess the requisite concept of RED:

> I do not need the concept RED to see red. But I do need this concept to become aware of the *quale* red, to become aware that I am having an experience of this sort. This being so, qualia (understood as qualities that distinguish experiences from one another) remain "hidden," inaccessible, until one acquires the conceptual resources for becoming aware of them. . . For to become aware of the quale red is to become aware *that* one is having an experience of a reddish sort, and this is something one cannot be made aware of without understanding what it means to be red. (Ibid.: 139-140)

With the distinction between look$_p$ and look$_d$ available, consider Inverted Earth. Block claims that when Ned introspects the phenomenal character of his experience of the Earth sky, he will find the same thing that Twin Ned finds when he introspects the phenomenal character of his experience of the Inverted Earth sky, despite their relative inversion in intentional content. But Dretske asks us: how do Ned and Twin Ned know about the phenomenal characters of their experiences? Dretske's answer is simply that each asks himself how the object of his experience—the sky, say—looks$_d$ to him. Joseph Levine summarizes Dretske's point nicely: "What else could there be to each's knowledge of what his experience is like than his knowing how what he's looking at looks to him? But to know what something looks like, in the sense that yields knowledge, is to know how it looks$_d$" (Levine: 114). The key point, now, is that how something looks$_d$ is determined by the environment outside the head, in the same way that the content of beliefs caused by experiences is determined by the environment outside the head, as we were supposed to have learned from the earlier Twin Earth examples. Since Ned and Twin Ned's environments are different, it follows that the sky looks$_d$ one way to Ned and looks$_d$ a different way to Twin Ned (they have different look-beliefs, following their inverted semantics), and *that is all they can know* by introspecting their phenomenal experiences. Now, perhaps the phenomenal look of the sky is really the same for Ned and Twin

Ned after all, despite the fact that they cannot know it; but then again, "If it is not something they can be made aware of, why suppose it must be the same?" (Dretske 1995a: 141). In fact, given that the sky looks$_d$ differently to them, "what better explanation is there for why the experience of Fred and Twin Fred seems so different to them than that it is?" (141).

Dretske is not saying that the phenomenal character must be different. He is only challenging the intuition that the phenomenal character must be the same. In fact, his general appraisal of the intuition that phenomenal character must be the same between molecular-duplicate twins, what he calls the "Internalist Intuition," is that "the Internalist Intuition is a brute intuition, one that is not justified by any defensible claim about the nature of thought or experience" (Ibid.: 150).

Tye (1998) has advanced a related argument for the possibility of phenomenal inversion between molecular duplicates. It is related because, just as Dretske tries to raise doubts about how Ned could know that his experience was phenomenally the same as Twin Ned's (in the interpersonal case), Tye tries to raise doubts about how Ned could know that his experience on IE was phenomenally the same as his earlier experience on Earth (in the intrapersonal case). Tye asks us to think back to Putnam's Twin Earth example. Suppose that I have been transported to Twin Earth and that I have spent sufficient time there for my semantics to change, such that when I say, "Water is wet," I am now referring to *twater* (XYZ), no longer water (H_20):

> Suppose now I say, 'I take my gin with water just as I did in my undergraduate years.' My word 'water' now means twater; so, the belief I express here is false (assuming I switched to Twin Earth after getting my B.A.). As an undergraduate, I drank water, not twater. The 'memory' on which my belief is based is really a mismemory, induced by the deep shift in my external relations: I am no longer referring to the same liquid by the word 'water' as I did in my youth. (Tye 1998: 464)

Now, Tye wants to draw a parallel between this case of misremembering and the case in which Ned, having traveled to IE from Earth, thinks to himself

that the sky looks the same as it always did (which Tye will suggest is also a case of misremembering):

> . . . the strong representationist can say that my report of no change in phenomenal character is like the case above in which I make a report of a distant past episode on Earth after having spent many years on Twin Earth: it is necessarily in error. By hypothesis, on the representationist view, color experiences change their phenomenal character with a change in represented color. When I now say, after a long time on Inverted Earth, 'Grass looks green to me now, just as it did five, ten, and twenty years ago,' I am wrong. 'Green' (in Inverted English) means red; and grass did not look red to me twenty years ago. My memory has led me astray. (466)

According to Block, this memory reply simply begs the question, by assuming externalism about phenomenal memory without independent argument (Block 1996: 44-45). According to Tye, Block has given no argument for this claim (Tye 1998: 470). I will not try to settle this dispute by deciding who is begging the question and who is giving an argument; instead, I will try to offer some positive considerations that weigh against the memory reply. These considerations also apply to Dretske's argument and I hope to show how.

Cross-Modal Matching and Inverted Earth

Georges Rey has suggested that perceptual psychology provides some of the strongest considerations in favor of the intuition that phenomenal character stays the same for molecular-duplicate twins, regardless of inversion in intentional content, as with Inverted Earth:

> A great deal is known about the idiosyncrasies of particularly the visual system, and, although the relevant "Inverted Earth" experiments have, of course, never been performed, it would seem incredible to suppose that none of the various laws

> that apply to Earthlings wouldn't apply to double inverted individuals (on Inverted Earth with inverting lenses). For example, reds are experienced as "exciting" and "advancing," greens as more tranquil and receding (Hardin 1993). We have no reason to suppose that, after Ned himself resides (unwittingly) on Inverted Earth long enough for his semantics to change, that the laws wouldn't continue to apply to him, and that consequently things would look the same to him, despite his eventual wide semantic changes. Thus, despite his word 'green' meaning (unbeknownst to him) *red*, the experience he describes as "looking green" would continue to be, e.g. "cool" and "receding." (Rey 1998: 445)

An especially clear-cut example of such perceptual laws comes from research into a psychological phenomenon known as cross-modal correspondence. When subjects report a sense of correspondence between particular sensations in one sensory modality, like vision, and particular sensations in another sensory modality, like olfaction, psychologists speak of a cross-modal correspondence between those sensations. As it turns out, there is statistically significant agreement among subjects about which sensations correspond cross-modally. For example, consider cross-modal correspondence involving colors and odors: one study used an array of twenty odors obtained from the perfume industry and asked participants to "find the one color chip that best represented [each] odor" (Gilbert *et al.*: 344). The results were that "thirteen odors had significant *r* values [on Rayleigh's test of significance], indicating a substantial normative agreement among subjects when selecting hues to describe these odors" (345). For instance, subjects regularly matched red with the odor of *cinnamic aldehyde* and green with the odor of *galbanum oil* (347). Not only did statistically significant matching regularities obtain across subjects, but matching regularities held intrapersonally over time: "Our call-back experiment showed that odor-color correspondences were stable for a group of subjects over a period of 2 years, a consistency of result equal to that observed in a similar, but nonolfactory, task" (349).

Imagine giving Ned and forty-nine other subjects (the sample size of the Gilbert *et al*. study was fifty) this cross-modal matching test on Earth. Since subjects in the Gilbert *et al*. study regularly matched the color red with the odor of cinnamic aldehyde, it is fair to suppose that Ned and his friends would perform similarly. Next, imagine Ned and his friends repeating the test some years later when they are all living on Inverted Earth (having all been kidnapped, outfitted with inverting lenses, etc.). For this IE call-back study, there is no reason to believe the results would be any different from the call-back study performed two years after the real experiment by Gilbert *et al*. Ned and his friends on IE, all of whom wear inverting lenses, would still match 'red' with cinnamic aldehyde and they would do so despite the putative changes in the intentional contents of their experiences and despite that 'red' now means *green* for them. This is a crucial fact: the intentional contents of experience will have inverted and thus, according to phenomenal externalism, the phenomenal character of experience will also have inverted; and yet, the cross-modal matching behavior will be *the same*. How is this possible?

Of course, the externalist might reply that it is obvious that the cross-modal matching behavior would be the same, since we are imagining that Ned and his friends would receive the same retinal stimulation from the color chips on IE that they received from the color chips on Earth, thanks to their new inverting lenses, and since there is no difference in odor between Earth and Inverted Earth; when Ned and his friends on IE are presented with a 'red' [green] color chip and a whiff of cinnamic aldehyde, their brain states will be the same as when they were presented on Earth with a 'red' [red] color chip and a whiff of cinnamic aldehyde. Of course, if their brain states are the same, then their matching behavior will obviously be the same, since there can be no difference in behavior without a difference in the brain. But this perhaps obvious fact entails the following less obvious one: the externalist here must hold that the phenomenal character of experience is *causally irrelevant* to the matching behavior. After all, the externalist must hold that the phenomenal character will change, due to the change in intentional content, and yet the cross-modal match-

ing behavior will clearly not change. This leads to two problems. First, according to phenomenal externalism, the phenomenal character does not affect the cross-modal matching behavior, even though subjects will likely say that the very thing they are attending to when performing the matching behavior is the phenomenal character of their experiences! Second, if the phenomenal character is in this way *epiphenomenal*—if it makes no causal difference in terms of effects on behavior—then why do phenomenal externalists go to all the trouble of arguing for its identification with (wide) intentional content?

Externalism and Epiphenomenalism

Dretske has dealt with the charge of epiphenomenalism before, though not from the same angle (as I will explain shortly). Here is what he has to say about epiphenomenalism:

> Once the difference between behavior and bodily movement is clear, it also becomes clear that externalist theories of the mind are *not* threatened by epiphenomenalism. Mental content can explain behavior without supervening on the neurophysiological events and processes that cause bodily movement. The mental is *not* robbed of its explanatory relevance by being extrinsic. (Dretske 1995a: 152)

Dretske is responding here to the charge that, "If the mental is extrinsic, if it does not supervene on the present physical constitution of Fred and Twin Fred, as it clearly doesn't on a representational theory of experience and thought, then it will be irrelevant to why Fred and Twin Fred behave the way they do" (152). Against this objection, Dretske replies that in the same way that, for example, extrinsic factors (like evolution) can explain *why* a plant seasonally changes its pigment without explaining the *actual process* of seasonal pigment change (which depends on intrinsic chemical reactions), extrinsic factors can also explain why humans behave the way they do without explaining the *actual processes* underlying bodily move-

ment (which depends on intrinsic neurophysiology). This is all fine. The present charge of epiphenomenalism need not dispute it.

The present charge of epiphenomenalism is simpler and more problematic for the externalist: comparing the initial Earth-based cross-modal correspondence study and the later IE call-back study, the Bullet Biting phenomenal externalist must hold that the phenomenal character of experience has inverted, due to an inversion in intentional content; and yet, the cross-modal matching behavior has not changed. It follows that phenomenal character must not affect the cross-modal matching behavior, even though subjects will likely say that the very thing they are attending to when performing the matching behavior is the phenomenal character of their experiences. Two options follow: either subjects *are* using the phenomenal character of their experience to perform the cross-modal matching behavior, in which case the Bullet Biting externalist replies of Dretske and Tye above cannot be correct to say that the phenomenal character could change from Earth to IE (unless they can explain why subjects match colors-to-odors in the same manner despite the fact that the colors they are matching have phenomenally inverted with respect to the odors!), or, subjects are *not* really using the phenomenal character of their experience to perform the cross-modal matching behavior, despite the fact that they *say* they are, in which case the phenomenal character of experience is causally irrelevant in what seems like an ideal case for the phenomenal character to do real work.

I suggest that the externalist pursue the first option and give up Dretske's and Tye's arguments, rather than biting the epiphenomenal bullet, for that bullet is just too difficult to swallow. Recall that Dretske's strategy was to raise questions about how Twin Ned could know that the phenomenal character of his experience was the same as Ned's and that Tye's strategy was to raise questions about how Ned could know that the phenomenal character of his experience on IE is the same as it was back on Earth. The cross-modal matching test seems to provide an answer: if you want to know if the experiences of Ned and Twin Ned (or Ned on IE and Ned back on Earth) have the same phenomenal characters, give them the cross-modal matching test; see if they both match 'red' with cinnamic aldehyde or if the one on

IE matches 'red' with *galbanum oil* (the odor matched with green in the Gilbert study), which is what we might expect if one's experience of 'red' on IE were really phenomenally green. This may seem like cheating, since of course molecular duplicates will have the same matching behavior—red with cinnamic aldehyde—*but that is just the point*. If the matching behavior depends on the phenomenal character, then the same matching behavior will, other things equal, indicate the same phenomenal character. (If it does not, we have epiphenomenalism for phenomenal character.) We have answered Dretske's question of how Ned and Twin Ned could know they have experiences with the same phenomenal character, irrespective of the $look_p$/$look_d$ distinction, and Tye's question of how Ned on IE could know he was having an experience of the same phenomenal character as he had back on Earth, irrespective of debates about externalism for phenomenal memory. The answer is in the form of a perfectly empirical test and it vindicates the "Internalist Intuition" that phenomenal character is the same between molecular duplicates on Earth and IE.

Saving Face on Inverted Earth

At this point, I think the externalist ought to give up the Bullet Biting reply to Inverted Earth and instead pursue a "Saving Face" reply: argue against Block that Ned's (visual) intentional content does not in fact invert when he goes to IE, though his semantics may. This would then explain why Block thinks (correctly) that phenomenal character does not invert and why it makes sense that cross-modal matching behavior stays the same. One externalist reply along these lines appeals to the *teleology of tracking*: on an evolutionary-selectional view of representation like Dretske's, intentional content does not invert because, even with the change of worlds and inverting lenses, the perceptual state produced in Ned when he looks at the yellow sky on IE was "designed" by nature (through natural selection) to provide information about *blue* skies on Earth and, thus, the state represents the IE sky as blue too. Though *caused* by yellowness on IE, that state was hard-wired by evolution to track blueness and so blueness it will represent:

> Once an indicator system is [naturally] selected to provide the needed information it has the function of providing it. The states these systems produce by way of performing their informational duties then become representations of the conditions they have the (systemic) function of informing about. (Dretske 1995a: 164)

Ned's state was, in evolutionary history, selected to provide information about the blueness of skies, not yellowness. Thus, since this state has the (systemic) function of informing about blueness, it represents blueness, even if it is caused by yellow skies on IE.

Considerable difficulty has been raised for such teleological views by various "Swampman" examples: a molecular duplicate of a person materializes by chance when lightning strikes a swamp; this "Swampman" has, *by hypothesis*, no intentional content—for he has no teleological history—and yet, musn't he have the same sort of phenomenal consciousness as his non-Swamp twin, when they are both looking at the sky? Dretske valiantly defies this Internalist Intuition and says 'no,' because on his view, a state is a conscious state in virtue of having the function of providing information, but Swampman's states do not have any functions at all, for states acquire functions by being naturally selected for in evolution and Swampman is not a creature with an evolutionary history (Dretske 1995a: 148-149).

So Dretske admits that on his version of externalism, everything goes on "in the dark" (143) for the Swampman—he has no phenomenal consciousness at all—though he is a molecular duplicate of some other person *who is* conscious. Now, suppose we administer the cross-modal matching test to the Swampman. There can be no doubt that he would perform the same way on the test as his molecular-duplicate, human counterpart, matching red with cinnamic aldehyde, for example. But then it seems that the externalist must again hold that phenomenal experience is causally irrelevant to the cross-modal matching behavior. Even though subjects in a cross-modal correspondence test probably cannot imagine performing the matching behavior without any phenomenal characteristics of olfactory and visual experiences to compare, Dretske must say that the Swampman does just this:

he performs the same way as the humans do, but with no phenomenal experiences at all! Incredible as this is, it may actually be that full epiphenomenalism does not follow, as John Perry points out a delicate issue here (from an unrelated discussion of zombies): It could be that the experiences in our world are redundant. They have effects, but for each and every experience in our world that has a physical effect there is some other state that would have brought about that effect if the experience hadn't. This hypothesis, inspired by recherché considerations in the literature on counterfactual analyses of causation, may be even less plausible than epiphenomenalism. (Perry: 218)

On this picture, though the phenomenal character of experience is causally relevant when we perform the matching behavior, when Swampman performs the matching behavior "in the dark," some other, *non-phenomenal* properties apparently work just as well. As implausible as this seems, even if phenomenal characters were not fully epiphenomenal but rather just causally unnecessary in this way, this would still be bad enough to pose a dilemma. If the Swampman does have phenomenal experiences, Dretske's evolutionary-selectional theory is undermined. If the Swampman does not have phenomenal experiences, Dretske's theory has rendered phenomenal experience causally unnecessary, even for cross-modal matching, where phenomenal experience certainly seems to be playing a vital role. Without independent reason to believe that phenomenal experience *is* causally unnecessary, this consequence of phenomenal externalism surely counts as a serious cost of accepting the theory.

"The Scylla of Inverted Earth and the Charybdis of Swampman"

But perhaps there can be an externalist, strong representationist theory that does not strip the phenomenal of its causal importance. Tye gives an account of representation, similar to that with which we began, that does not appeal as directly to evolution and thus might avoid the Swampman problems. According to this account, "S represents that *P* =df If optimal conditions obtain, S is tokened in *x* if and only if *P* and because *P*" (Tye 1995:

101). The key here is the notion of *optimal conditions*, which offers another Saving Face reply to Inverted Earth, one that denies any inversion in intentional content. On Tye's view of representation (what he calls "ideal" causal covariation), intentional content does not invert because, given the change of worlds and inverting lenses, optimal conditions for causal-covariational tracking—the kind of tracking required for representation—do not obtain: "The insertion of the lenses interferes with the operation of the sensory transducers. Accordingly, the transduction process is not in itself normal or optimal" (Tye 1998: 472). Tye goes on to say:

> Intuitively, then, it is true of the traveler's sensory state, as he looks at the clear sky on Inverted Earth (after however many years), that *had* there been no interference, that phenomenal state *would have been* causally correlated (in him) with blue things. Accordingly, by the causal covariation proposal, the traveler's sensory state continues to represent the clear sky as *blue*. (473)

So Tye's ideal causal-covariational view of representation would seem to be compatible with the consistency of cross-modal matching between Earth and IE, since interference in optimal conditions precludes the inversion in intentional content and thus the inversion in phenomenal character that has led to the problems of epiphenomenalism we have seen.

However, this way out of the cross-modal problem faces a different difficulty: imagine (as Tye himself does) that Swampman travels to Inverted Earth. Is there now any sense of *interference* in the Swampman's visual system—interference that would preclude the problematic inversion in intentional content—given that Swampman has no history and thus no "design"? Tye thinks that "it is surely pre-theoretically correct to say that his transducers have been interfered with" (Ibid.: 473). The problem is that the notion of interference only makes sense in the context of some deviation from the normal design or function of the Swampman's systems. But as Dretske illustrates with his clever examples involving swamp duplicates of artifacts, systems in swamp duplicates have no normal design or function,

because they came together haphazardly from lighting explosions! The only basis one has for saying that the lenses constitute interference is the resemblance of the swamp duplicate's systems to ours. Dretske urges us to resist this superficial resemblance:

> . . . how unreliable ordinary intuitions are about miraculous materializations and instantaneous replacements. Our judgments about what it makes sense to say—what it would be true to say—about such bizarre cases are influenced by factors that, on deeper reflection, we see to be quite irrelevant. We are, for instance, influenced by a striking resemblance in appearance and placement of parts. . . . When asked to render judgments about more complex [swamp] objects—automobiles, for example—we blithely ignore the fact that the resemblance in both appearance and placement is (by hypothesis) completely fortuitous and thus, irrelevant to determining the function of parts. We ignore this and proceed to assign functions on the basis of resemblance and placement anyway. We seem driven by what I call the Paley Syndrome—an irresistible tendency to use resemblance and placement as a basis for attributing purpose and design. (Dretske 1995a: 146)

The point, then, is that it is unclear how Tye's notion of interference has determinate application in the case of Swampman, except by his resemblance to us (and what would constitute interference in our case), which Dretske rightly urges us to discount. So, while invoking optimal conditions and "interference" at first seemed to allow the phenomenal externalist to account for the consistency of our cross-modal matching behavior between Earth and IE, it does not seem adequate to account for Swampman in the same way, to whom such terms as "interference" do not clearly apply. But if the optimal conditions and interference account is not generally applicable, then without further explanation of why this is so, it seems *ad hoc* to claim it is still what accounts for the standard human cases. I think this situation calls for skepticism about Tye's claim that "a safe path exists, for the rep-

resentationist, between the Scylla of Inverted Earth and the Charybdis of Swampman" (Tye 1998: 460).

Then again, I do not claim to have decisively closed off this "optimal conditions" escape in the vein of Tye's Saving Face reply to Inverted Earth, a reply that promises to account for the cross-modal matching cases on behalf of phenomenal externalism. The critic of phenomenal externalism does have something closer to a knockdown argument against the other externalist replies to Inverted Earth, however, for these replies run into serious trouble with their denial of the natural supervenience of phenomenal character on the brain. I have tried to show that the Bullet Biting replies of the externalist render the phenomenal character of experience epiphenomenal (and Dretske's Saving Face reply renders it at least causally unnecessary), which strongly conflicts with our intuition that the phenomenal character of experience is causally efficacious, especially in the case of cross-modal color-to-odor matching. Indeed, as Rey puts it, "given at least the ordinary interest we have in classifying people's experience *as it seems to them*, there surely is a strong presumption that we should take them at their (sincere) word. . . the burden is on the doubter of a person's sincere introspections to show good reason to think those introspections mistakes" (Rey 1998: 443). And if people are in fact mistaken that the phenomenal character of their experience is causally efficacious, why do we need a detailed externalist theory of it? Perhaps it is time we return to the view that phenomenology is in the head and think more carefully about how to go about finding it there.

Works Cited

Block, Ned. “Inverted Earth.” *Philosophical Perspectives* 4 (1990): 53-79.

———. “Mental Paint and Mental Latex.” *Philosophical Issues* 7 (1995): 19-49.

Dretske, Fred. *Naturalizing the Mind*. Cambridge: MIT Press, 1995a.

———. “Phenomenal Externalism or If Meanings Ain’t in the Head, Where are Qualia?” *Philosophical Issues* 7 (1995b): 143-158.

Gilbert, Avery N., Robyn Martin & Sarah E. Kemp, “Cross-Modal Correspondence Between Vision and Olfaction: The Color of Smells.” *American Journal of Psychology* 109.3 (1996): 335-351.

Levine, Joseph. *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press, 2001.

Perry, John. “Replies.” *Philosophy and Phenomenological Research* 68.1 (2004): 207-229.

Putnam, Hilary. “Meaning and Reference.” *Journal of Philosophy* 70.19 (1973): 699-711.

Rey, Georges. *Contemporary Philosophy of Mind*. Cambridge: Blackwell Publishers Inc., 1997.

———. “A Narrow Representationist Account of Qualitative Experience.” *Nous* 32 (1998): 435-457.

Tye, Michael. *Ten Problems of Consciousness: a Representational Theory of the Phenomenal Mind*. Cambridge: MIT Press, 1995.

———. “Inverted Earth, Swampman, and Representationism.” *Nous* 32 (1998): 459-477.

———. *Constructions of Reason*. Cambridge: Cambridge University Press, 1989.

Interview With Richard Rorty, Stanford University

Richard Rorty is professor emeritus of Comparative Literature and Philosophy at Stanford University. Among his many publications are *The Linguistic Turn* (1967), *Philosophy and the Mirror of Nature* (1979), *Consequences of Pragmatism* (1982), and *Contingency, Irony, and Solidarity* (1988). Professor Rorty is famous for his distinct views on pragmatism, epistemology, and the fate of analytic philosophy. This interview was conducted via email over the week of March 20, 2006.

THE YALE PHILOSOPHY REVIEW: *Over the years, your writings have embraced a wide array of philosophical areas and the work of vastly different thinkers. What directions has your work taken since the publishing of* Philosophy and the Mirror of Nature *in 1979 and what current projects are you working on?*

RICHARD RORTY: *Philosophy and the Mirror of Nature* was intended as a contribution to analytic philosophy, building on the work of Quine and Sellars. After I finished working on that book, I wanted to do something different. So I started giving seminars on such figures as Nietzsche, Heidegger, Gadamer and Habermas. This eventually led to my writing *Contingency, Irony, and Solidarity* and *Essays on Heidegger and Others*. Over the last couple of decades, I have tacked back and forth between taking part in controversies within analytic philosophy and writing on other topics. These days I am thinking of revising and supplementing some recently published pieces with an eye to publishing a short book called *Philosophy as Cultural Politics*.

YPR: *What current American philosophers excite you?*

RR: The two philosophy books from which, in recent years, I have learned most are Robert Brandom's *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality* and Robert Pippin's *The Persistence of Subjectivity: On the Kantian Aftermath*. Brandom and Pippin combine vast learning with genuine originality. Their accounts of what has been going on in philosophy since Kant's time are very persuasive. Both are trying to get us to take Hegel more seriously than most Anglophone philosophers are presently willing to take him. I hope that they succeed.

YPR: *How would you respond to Habermas's critique that both you and Derrida fail to differentiate between the* communicative, *problem-solving discourses of ordinary language and the complex,* world-disclosing *discourses of art and literature? Can all philosophy be read as little more than good or bad literature?*

RR: I do not think that either Derrida or I fail to recognize the difference between these two kinds of discourse. Like other fans of Thomas Kuhn, I routinely invoke his distinction between normal and revolutionary science; that is one species of the generic distinction that Habermas draws. But world-disclosure is not something that can be stuck in a pigeon-hole called "art and literature." It is what geniuses in every area accomplish. Plato, Galileo, Darwin, Hegel, Freud, Marx, Nietzsche, Whitman, Shelley, Giotto, St. Paul, and the Buddha all disclosed new worlds. I do not know what it would be to read any of these authors "as literature" or "as philosophy." One simply reads them, and lets the result of doing so interact with the results of having read various other books.

YPR: *You have published widely on political issues and on the relationship between a pragmatist epistemology and liberalism. What significance can a theory of knowledge have for politics?*

RR: Not much. I have argued that there is nothing self-contradictory about being a pragmatist in one's views of knowledge, truth and inquiry and being a reactionary Republican, or even a Nazi. Still, it might be that if pragmatist views became universal, we would be somewhat less susceptible to appeals to authority. I have suggested this in an essay called "Pragmatism as Anti-Authoritarianism."

YPR: *Explain briefly what you have in mind when you refer to the "Good Global Society."*

RR: It is a society in which Rawls' "Difference Principle" is applied, and thus one in which Marx's program "From each according to his ability, to each according to his need" is, as far as proves possible, carried out. I take it that democratic institutions are necessary for this goal to be achieved. So I think of a global utopia as one in which these institutions are found in every country, and in which nation-states cede considerable authority to supernational entities such as the UN so as to permit the creation of a global police force, a global disaster relief agency, and the like.

YPR: *Does belief in this kind of liberal utopia lead you to support democracy promotion?*

RR: I certainly believe that democratic nations should unite to foster democratic institutions in countries in which they do not exist. Sometimes this must be done by using military force, as in the struggles against Hitlerism and Stalinism, and in the cases of Bosnia and Kosovo. Sometimes it can best be accomplished by non-military means.

YPR: *You have mentioned that the critical philosophy of thinkers like Heidegger or Foucault is perfectly compatible with an old-fashioned liberal humanism, even though the two have traditionally been at odds. Imagine you were sitting in the audience with your teacher Carnap at the famous Davos Debate of 1929 between Heidegger and Cassirer. Between the three of these*

thinkers, where would your sympathies have lay, both philosophically and politically? Could Cassirer's liberal humanism have been reconciled with Heidegger's critical, existential reading of Kant?

RR: I do not think that one's reading of Kant has much to do with one's attitude toward liberal humanism. Heidegger had a screwy interpretation of Kant and a jaundiced attitude toward democracy. Cassirer had both a more plausible interpretation of Kant and more reasonable political views. One can imagine a third figure who was a better reader of Kant than either but whose politics were even more repellent than Heidegger's.

Special thanks to the Yale Philosophy Department.

This issue of THE YALE PHILOSOPHY REVIEW was produced using MiKTEX <http://www.miktex.org/> and TEXnicCenter <http://www.texniccenter.org>, and was printed by the Yale Reprographics and Imaging Services of New Haven, Connecticut. The typeface is Times.