

Volume 6 Number 2
2011

ISSN 1807-9792

abstracta

Linguagem, Mente & Ação

Concerning the Possibility of Exactly Similar Tropes
M. A. Istvan Jr.

**New Criteria for Pain:
Ordinary Language, Other Minds, and the Grammar of Sensation**
Kieran Cashell

Multiple Realizability and Novel Causal Powers
Ricardo Restrepo

**Phenomenal Character as the Mode of Presentation
of Environmental Properties**
Roberto Horácio de Sá Pereira

**Nietzsche ou MacIntyre:
Duas Alternativas à Moralidade Moderna?**
Helder Buenos Aires de Carvalho & José Elielton de Sousa

ABSTRACTA
Linguagem, Mente e Ação

ISSN 1807-9792

Volume 6 Number 2
2011

Editors

André Abath

Leonardo de Mello Ribeiro

Executive Editor

Gottfried Vosgerau

TABLE OF CONTENTS

Concerning the Possibility of Exactly Similar Tropes	158
M. A. Istvan Jr. (<i>Texas A&M University</i>)	
 New Criteria for Pain:	
Ordinary Language, Other Minds, and the Grammar of Sensation	178
Kieran Cashell (<i>Limerick Institute of Technology</i>)	
 Multiple Realizability and Novel Causal Powers	
	216
Ricardo Restrepo (<i>Instituto de Altos Estudios Nacionales, Ecuador</i>)	
 Phenomenal Character as the Mode of Presentation of Environmental Properties	
	231
Roberto Horácio de Sá Pereira (<i>Universidade Federal do Rio de Janeiro</i>)	
 Nietzsche ou MacIntyre: Duas Alternativas à Moralidade Moderna?	
	252
Helder Buenos Aires de Carvalho (<i>Universidade Federal do Piauí</i>) & José Elielton de Sousa (<i>Universidade Federal do Piauí</i>)	

CONCERNING THE POSSIBILITY OF EXACTLY SIMILAR TROPEs

M. A. Istvan Jr.

Abstract

In this paper I attempt to show, against certain versions of trope theory, that properties with analyzable particularity cannot be *merely* exactly similar: such properties are either particularized properties (tropes) that are dissimilar to every any other trope, or else universalized properties (universals). I argue that each of the most viable standard and nonstandard particularizers that can be employed to secure the numerical difference between exactly similar properties can only succeed in grounding the particularity of properties, that is, in having properties be tropes, at the expense of ruling out the possibility of their exact similarity. Here are the four nonstandard particularizers that I examine: the genealogy of a property, the history of a property, the causal effects of a property, and the duration of a property. And here are the two standard particularizers that I examine: the bearer of a property, by which I mean either a bare particular or a spatiotemporal location, and the property itself, by which I mean that the property is self-particularized. In my concluding remarks, I explain that the only remaining hope for preserving the possibility of exactly similar tropes is regarding properties as primitively particular, and that this must mean not that properties are self-particularized but that they are particularized due to nothing. I close by arguing that this may not help trope theory after all.

0. Introductory remarks

A common practice of trope theory, the view that regards properties as particulars and so as lacking the universal's "promiscuous" capacity (Campbell 1990: 53) for being a constituent of multiple nonconcurrent entities at one and the same time, is to replace fully determinate universals with sets of exactly similar, that is, qualitatively indiscernible, properties—sets that, without being repeatable, provide many of the same services as universals do (such as being that to which abstract singular terms refer) (see Williams 1966: 81-82; Campbell 1990: 74). For me at least, the admission of properties, whatnesses, that are exactly similar and yet in no way whatsoever identical—in short, properties that are *merely* exactly similar—is somewhat perplexing. After all, each of these fully determinate and exactly similar yellowness properties occupy and express one and the same position on the yellowness spectrum; each of these fully determinate and exactly similar mass properties occupy and express one and the same position on the mass spectrum; and so on. That this is tough for me to get my head around seems to be why, despite the many interesting questions related to trope theory, I find there to be one question in particular whose calling drowns out all the rest—one, to which I will

now attend, that concerns the very possibility of trope theory as commonly understood. Is there any viable way to individuate, where by “individuate” I mean numerically differentiate, exactly similar particularized properties without undermining either their particularity or their exact similarity? Is there, in short, any successful particularizer for exactly similar tropes?

In this paper, I endorse an answer of “no” to this question. I will argue that although it might be true that properties are particulars, that is, are tropes rather than universals, it is impossible for there to be tropes that are exactly similar if their particularity is analyzable, explainable. After a brief section where I will argue that the relation of being exactly similar in question must be internal to the tropes that are purportedly exactly similar (section 1), I will argue that even the best among the various nonstandard (section 2) and standard (section 3) particularizers that trope theory might employ to secure the numerical difference between exactly similar tropes can only succeed in grounding the particularity of properties, that is, in having properties be tropes, at the expense of ruling out the possibility of their exact similarity. This I intend to accomplish by disclosing the various ways that a trope theorist might employ each candidate particularizer to perform its particularizing job and by then proceeding to show that, in some of these ways, the particularizer fails to have properties be tropes and that, in the rest of these ways, the particularizer succeeds in having properties be tropes at the expense of ruling out the possibility of exact similarity. Since I show that no particularizer (whether it be a property’s genealogy, history, causal effects, duration, bearer, or even a property’s very own self) can both particularize a property and preserve exact resemblance between properties, by the end I hope it will be clear why, as is now being acknowledged in the literature, the only way for trope theory to guarantee both particularity and exact resemblance is to hold that property particularity is a brute fact, that it is due to nothing—neither something other than the property nor even (despite what we may think) the property itself. I do raise a problem for this view in my concluding remarks, however.

1. The relation of being exactly similar is internal

When I am worrying about how monadic properties x_1 and x_2 can be individuated despite standing in the relation of being exactly similar, am I taking this relation to be

external or internal to x_1 and x_2 ? As most contemporary trope theorists agree (take, for instance, Campbell 1990: 37, 90), this relation is internal, that is, entirely a function of the relata, such that the mere existence of x_1 and x_2 entails their being exactly similar. Aside from problems such as the fact that the infamous Bradley regress appears to sweep up any view that regards this relation as external, it just seems plain wrong to say the opposite: that it is a contingent matter whether the two *being two kilotons* properties, x_1 and x_2 , are exactly similar, such that instead of being exactly similar to x_2 x_1 could have been exactly similar to y_1 , *being three kilotons*, or even perhaps to g_1 , *being a creeping motion*.

Now, if x_1 and x_2 were ordinary objects such as houses, then it might not be the case that the mere existence of the two entails their being exactly similar. For on the at least somewhat reasonable assumption that the two houses would retain their identities through such minor changes as patio additions, by only house1 gaining a patio and not so too house2 (which was exactly similar to house1 before the patio addition), it will be the case that their being exactly similar does not follow from the mere fact that they both exist. The thing is, because the two exactly similar *whatnesses*, x_1 and x_2 , indubitably cannot in any intrinsic way change without ceasing to exist, saying that they are not guaranteed to be exactly similar is tantamount not to saying that the existence of the two houses does not guarantee their being exactly similar, but rather to saying that the houses being exactly as they are when they were objectively exactly similar does not guarantee that they are objectively exactly similar. This seems absurd.

2. Nonstandard particularizers for exactly similar properties

Let me move on now to interrogating some candidate particularizers for properties that are exactly similar. I will start with nonstandard particularizers. I call them “nonstandard” because they occur, from what I gather, rarely or not at all in the literature. There are four that I want to examine, each of which can be employed in a variety of ways: (1) a property’s genealogy, (2) its history, (3) its causal effects, and (4) its duration. I will argue that even in the cases where these particularizers are employed in ways that succeed in having properties be tropes, that is, that succeed in ruling out

any element of universality in properties, this success comes at the expense of ruling out the possibility of exact similarity between tropes.¹

2.1. *Genealogy as particularizer*

Can the genealogies of x_1 and x_2 , that is, the causal chains that lead to their coming into being, secure their being *merely* exactly similar?² On the quite reasonable assumption that the genealogies are external to x_1 and x_2 , which is to say that x_1 and x_2 are each something more than their genealogies, the possibility of x_1 and x_2 being numerically distinct simply in virtue of differing genealogies appears to be ruled out right away. The basis for me saying this is not that it is possible for two merely exactly similar properties to have an identical genealogy or at least exactly similar genealogies. For the sake of the argument, I am willing to grant that this is in fact impossible. So why do I still say that genealogy is ruled out as being a particularizer? Well, even assuming that the two natures have radically divergent genealogies, when we attend only to the two natures themselves—in effect bracketing the particularizing genealogies—before the mind's eye can only be but one nature. This follows, of course, because bracketed was what here is assumed to be the *only* things securing the numerical difference between x_1 and x_2 : the differing genealogies.

Am I to move on so soon to another candidate particularizer, then? There is, I guess, one way in which the genealogies can be external to the natures and yet when we peel those genealogies away there remain two natures. We can say that each nature in itself has an intrinsic directedness, an orientation grounded in each whatness itself, towards its specific genealogy. There are two problems with saying this, though. First, the genealogy is no longer the particularizer. The particularizer seems to be, rather, the trope itself. But in this case, and this is the second point, x_1 and x_2 cannot be exactly

¹ Note that what is most important in this section is not so much the catalogue of candidate particularizers, but rather the catalogue of options for how a particularizer can be employed to particularize. First of all, this latter catalogue covers, I believe, all the possible options. So by including it, readers will be able to see the parameters and, hopefully for the discipline, see if I missed something. Second, having this catalogue is helpful for those beginning to think about these issues and are considering taking one of the options that experts would regard as unviable for the reasons that I will provide. Finally, by going through these options now, when I get to the more interesting particularizers, the standard ones, I can focus on what really matters—namely, if they work—without having to go through the legwork of fleshing out the various ways that they might work, which would, I think, bog down the conversation.

² Genealogy as a possible particularizer, at least for events, has been stated, for example, by Van Inwagen (1983: 169) and Donald Davidson (1980: 306).

similar anymore. To be intrinsically oriented toward dissimilar genealogies is to be intrinsically dissimilar.

Since the goal here is to keep genealogies as the particularizers, and since when the genealogies particularize the natures from without, that is, externally, they can simply be bracketed, it seems that the only move we have left is to make the genealogies somehow intrinsic to the properties so that when we attend to the natures in themselves they are present. The first option for doing this that comes to mind is to have the genealogies make up some portion of x_1 and x_2 . I say “portion” on the strength that if x_1 and x_2 were nothing but their divergent genealogies, then we would by hypothesis give up on x_1 and x_2 being exactly similar. Right? However, would not the preclusion of being exactly similar follow even when we make the genealogies but a *portion* of the natures? Considered in their entirety, yes. But since x_1 and x_2 were exactly similar before making the genealogies intrinsic to them, it follows that at least whatever about x_1 and x_2 that exceeds the genealogy-portions will be exactly similar to each other. There is still a sense, in this case, to saying that x_1 and x_2 are exactly similar. They are exactly similar qua the portions of themselves exceeding their genealogies.

Alas, this glimmer of good news is short-lived. Even when I bracket the fact that this suggestion seems to violate the requirement that tropes be ontologically simple (see Morganti 2009: 190; Maurin 2002: 15; Campbell 1990: 20), that is, that they not have more than one constituent, it turns out that we are in the same predicament as when the particularizing genealogies were external to x_1 and x_2 . When we turn our attention only to x_1 and x_2 qua the portions of themselves exceeding their genealogies—in effect yet again managing to bracket the particularizers—reposing before the mind’s eye can only be what for the bluenosed scorer of universals is an anathema in this situation: one entity as opposed to many. This follows, of course, because bracketed was the *only* thing that made x_1 and x_2 —well, x_1 and x_2 qua the portions of themselves exceeding their genealogies—distinct: the genealogies.

Now, to try to overcome this problem that comes about when we make the genealogies portions or parts of the natures, I might say that the natures each possess their genealogies as *features*. Even here, however, there is something in excess to the genealogies. What is that? Well, that which has the features: the natures— x_1 and x_2 . When we attend only to the natures, then, we attend only to one.

It is now clear that the only way to prevent the mind's eye from happening upon any universal element in x_1 and x_2 is by doing what I noted would foil all hopes of x_1 and x_2 being—at least in some way—exactly similar: making x_1 and x_2 nothing but their divergent genealogies. Is there any way to make these natures nothing but their genealogies without undermining their being exactly similar? The only option is to make there be but a distinction of reason between the nature and the genealogy, in which case the nature and genealogy differ only as different ways of thinking or speaking about one and the same entity. This way trope theory can say that *qua nature* x_1 and x_2 are exactly similar whereas *qua genealogy* they are distinct.

There are several options as to what we could mean when we say that there is a mere conceptual distinction between the nature and the genealogy. We must decide what exists outside of the classifying mind when we say that there is such a distinction of reason between the nature and its particularizing genealogy. We must also decide whether there is a basis for the distinction of reason in the reality independent of the classifying mind.

There are four options for what exists independent of the classifying mind when we say that there is a mere distinction of reason (that is, a distinction related to our conception) between the particularizer, which in this case is the genealogy, and the nature. (1) Outside of the intellect there is both the nature and the genealogy, but it is just that each cannot exist without each other the way that entities that are “really distinct” can. (2) Outside of the intellect there is neither the nature nor the genealogy, only that entity of which the nature and the genealogy is predicated, that entity that is perceived or considered as being the nature and yet the genealogy. (3) Outside of the intellect there is the genealogy but not the nature, and the genealogy is that of which the nature is predicated, is that which is perceived or considered as being the nature. (4) Outside of the intellect there is the nature but not the genealogy, and the nature is that of which the genealogy is predicated, is that which is perceived or considered as being the genealogy.

Now, when deciding whether there is a basis for the distinction of reason we have only two options. (A) The distinction between the two has a basis in the reality independent of the classifying mind, that is, the distinction between the two “arises not entirely from the sheer operation of the intellect, but from the occasion offered by the

thing itself on which the mind is reflecting” (Suarez 1947: 18). (B) The distinction between the two is generated merely by reason, that is, the distinction between the two “has no foundation in reality and arises exclusively from the reflection and activity of the intellect” (18), and as such is like the distinction between P and P that arises in the identity statement “P is P.”

Holding A, that there is a basis for the conceptual distinction in the reality independent of the classifying mind, proves unhelpful for trope theory when it comes to each of the four aforementioned options concerning what exists independent of the classifying mind when we say that there is a distinction of reason. I will go through each of these options and explain why. (1A) Trope theory does not want to say that both the nature and the genealogy have reality outside of the classifying mind and that, even though they cannot mutually exist without each other, there is some sort of basis for their distinction in the reality outside the classifying mind. For even though they cannot mutually exist without each other, that would not stop there from being the particularizing genealogy, on the one hand, and the *universal* nature, on the other, since there is a basis for their distinction outside of the classifying mind. (2A) Trope theory cannot say that neither the nature nor the genealogy exists independent of the classifying mind—only that entity which is perceived or considered as the nature and the genealogy—and yet say that there is a basis for the distinction between the two in the entity independent of the classifying mind. For if there is a basis for the distinction in the mind-independent entity, then the genealogy and the nature would be mind-independent as well (would be *aspects* of the entity, we might say). Besides, contrary to traditional forms of nominalism (predicate nominalism, resemblance nominalism, and so on), trope theory believes that there are such things as natures in mind-independent reality. (3A) Trope theory cannot say that the genealogy exists independent of the classifying mind but the nature does not and yet say that there is a basis for the distinction between the two in the genealogy. For if there is a basis for the distinction in the mind-independent genealogy, then the nature would be mind-independent as well (would be some aspect of the genealogy). Besides, since trope theory believes that there are such things as natures, it does not want to say that the genealogy exists but the nature does not. (4A) Trope theory cannot say that the nature exists independent of the classifying mind but the genealogy does not and yet say that there is a basis for the

distinction between the two in the nature. For if there is a basis for the distinction in the mind-independent nature, then the genealogy would be mind-independent as well (would be some aspect of the nature). Besides, even if one could somehow make this work, it would still be the case that the nature is universal since it is distinct to some extent from what particularizes it.

Holding B, that there is no basis for the conceptual distinction in the reality independent of the classifying mind, also proves unhelpful for trope theory when it comes to each of the four aforementioned options concerning what exists independent of the classifying mind when we say that there is a distinction of reason. I will go through each of these options and explain why. (1B) Trope theory cannot say that both the nature and the genealogy have reality outside of the classifying mind and yet say that there is no basis for their distinction. There must be a basis if they are *both* outside of the classifying mind. (2B) Trope theory does not want to say that neither the nature nor the genealogy exists independent of the classifying mind even if they also hold that there is no basis independent of the classifying mind for the distinction made between them. For trope theory believes that there are such things as natures. (3B) Trope theory does not want to say that only the genealogy exists independent of the classifying mind, even if trope theory also says that there is no basis independent of the classifying mind for the distinction. Again, trope theory believes that there are such things as natures. (4B) Trope theory does not want to say that only the nature exists independent of the classifying mind (and so not as well the genealogy), even if trope theory also says that there is no basis for the distinction between the nature and the genealogy in the reality outside of the classifying mind, namely, in the nature. This would undermine the purported exactly similar between natures x_1 and x_2 . Here is why. First, the genealogies of natures x_1 and x_2 are assumed to be dissimilar. Second, the conception of the particularizing genealogy is nothing but a conception of the nature it particularizes (lest the particularizer have no efficacy to particularize according to the parameters of this 4B option where, on the one hand, there is no genealogy independent of the classifying mind and, on the other hand, there is no basis for a distinction between the genealogy and the nature in the reality independent of the classifying mind). To put this second premise in a different way, the conception of the particularizing genealogy is nothing but the conception of the nature, the way that the conception of the apple's redness is,

for the austere nominalist, nothing but the conception of the apple itself (rather than of some ontological constituent of the apple: redness). From these two points it follows that natures x_1 and x_2 alone account for their dissimilarity, in which case x_1 and x_2 are not exactly similar.

It is appropriate at this point to move on to the next candidate particularizer. But first, here is a graph of all the options as to what we could mean when we say that there is merely a conceptual distinction between the nature and its particularizing genealogy. Included is a major reason why each option fails.

	A. There is a basis for the distinction	B. There is no basis for the distinction
1. Both the nature and the genealogy are outside of the mind.	Does not work because the nature itself would be universal.	Does not work because there must be a basis if they are <i>both</i> independent of the intellect.
2. Neither the nature nor the genealogy is outside of the mind.	Does not work because trope theory believes in natures independent of the intellect.	Does not work because trope theory believes in natures independent of the intellect.
3. The genealogy is outside of the mind, but the nature is not.	Does not work because trope theory believes in natures independent of the intellect.	Does not work because trope theory believes in natures independent of the intellect.
4. The nature is outside of the mind but the genealogy is not.	Does not work because the nature itself would be universal.	Does not work because the exact similarity of natures is thus disallowed.

Note that in subsequent sections of this paper, where I test other particularizers, I will only refer to option 4B when I suggest that trope theory might say there is distinction of reason between the nature and its purported particularizer. First, 2 and 3 straightaway deny what trope theory believes: that there actually are mind-independent natures. Second, A straightaway denies what trope theory believes: that natures are not universals. Third, 1B is contradictory: the nature and its particularizer cannot *both* exist in the reality independent of the mind and yet there be no distinction between the two. 4B is the only option that is not inconsistent or in violation of the trope theory view that natures are particulars or in violation of the trope theory view that there are mind-

independent natures; it is the only option, in other words, that supports the trope theory view that natures are particulars, and does so without any directly apparent contradiction and without violating the trope theory view that there are mind-independent natures.³

2.2. *History as particularizer*

Can the histories of x_1 and x_2 secure their being merely exactly similar?⁴ Even aside from the fact that it may be that two exactly similar tropes could have an identical history or at least exactly similar histories, in which case their histories would not in fact be able to individuate them, the possibility of the histories of x_1 and x_2 grounding their numerical distinctness is ruled out for the same reasons why the genealogies of x_1 and x_2 were ruled out as being able to ground their numerical distinctness. Let me briefly recount the previous movement, this time as it occurs on the assumption that the histories of x_1 and x_2 — x_1 and x_2 at time t_1 —are the only particularizers.

First, if the particularizers are extrinsic to the natures (option α), then when we attend only to x_1 and x_2 , thus bracketing even the most divergent histories of the natures in question, we are thereby attending to one nature. Second, if we make the natures have an essential directedness towards the divergent histories (option β), then we rule out the histories themselves as the particularizers and we disallow the exact similarity between x_1 and x_2 . Third, even if we make the histories intrinsic to the natures (option γ), whether by making them portions of the natures or features of the natures, there still remains an excessive element of universality. Fourth, for trope theory

³ This understanding about what a trope theorist is going to mean by there being a distinction of reason between a nature and its particularizer is fairer to trope theory than, for example, the understanding that Moreland has of it in his work against trope theory (see Moreland 2001: 59; 1989: 393-394). Moreland thinks that by making there be just a distinction of reason between the property and the particularizer, trope theory is necessarily faced with an inescapable dilemma: either the nature is nothing but a universal nature or it is not a nature at all, but rather is nothing but its particularizer (2001b: 64). My 4B option shows that trope theory need not be faced with this dilemma. As trope theory sees things according to 4B, the particularizer of this nature is to this nature what, according to austere nominalism, the redness of this apple is to this apple. In this case, it is true to conceive of the nature as the particularizer but the truthmaker for this correct conception is nothing but the nature itself. Because the particularizer is not an ontological entity, one horn of the dilemma is gone: the nature cannot in truth be nothing but its particularizer. And yet because the particularizer is truly predicated of the nature that *itself* serves as the truthmaker for that predication, it is the nature *itself* that secures its particularity, thus making it impossible for the other horn to arise.

⁴ I do not know how serious it is taken, but you do see such a view that history can serve to differentiate things. Simons, for example, points out that, as Hughes would have it, two things, such as ships, can be numerically differentiated by having different histories (1997: 762).

to make each nature merely conceptually distinct from its history (option δ) can only be for trope theory to say, for reasons I gave in the previous section, (4B) that outside of the intellect there is the nature but not the history, and there is no basis for a distinction between the nature and the history in the reality outside of the intellect, namely, in the nature. This δ option for how the particularizing history is employed to perform its particularizing job undermines the purported exact similarity between natures x_1 and x_2 for the same reasons that this δ option did in the previous section where genealogy was the particularizer. Since the histories of natures x_1 and x_2 are assumed to be dissimilar, and since the conception of the particularizing history is nothing but a conception of the nature it particularizes (lest the particularizer have no efficacy to particularize according to the parameters of this 4B option), then natures x_1 and x_2 must be dissimilar, must not be exactly similar.

2.3. Chain of causal effects as particularizer

Can the chains of causal effects of x_1 and x_2 secure their being merely exactly similar?⁵
No, and for the same reasons as above.

2.4. Duration as particularizer

Can the durations of x_1 and x_2 secure their being merely exactly similar?⁶ No, and for the same reasons as above.

3. Standard particularizers for exactly similar properties

Now let me move on to the standard particularizers for exactly similar properties. I call them “standard” because these are the ones that appear most often in the literature. There are two main candidate particularizers that I want to examine: (1) the property’s bearer or (2) the property itself.

⁵ Davidson suggests that the numerical distinctness of some things—he has in mind events—might be explained by a difference in their causal effects (1980: 306). When it comes to events, he personally thinks that the causal effects plus the genealogies is the particularizer that works. Note, by the way, that by the chains of causal effects I do not mean the causal powers of the natures, which will be exactly similar between the two exactly similar natures. I mean, rather, everything that each nature is complicit in bringing about, which can of course differ drastically.

⁶ Davidson offers the suggestion that the numerical distinctness of some things might be explained by a difference in their duration (see 1980: 305). He does not agree, however, that this is a good particularizer.

3.1. *Bearer as particularizer*

Two things are commonly meant by “bearer of a property.” Going from least to most popular, by “property bearer” we are going to mean either a bearer that is not a spatiotemporal location or a bearer that is a spatiotemporal location. So let us look at both of these.

3.1.1. *Non-spatiotemporal location bearer*

By “non-spatiotemporal location bearer” it seems that we are going to mean either a trope-bundle bearer, whether consisting of many tropes or just one, or else simply a bare particular bearer, where by “bare particular” I mean that in which properties inhere (but in itself is propertyless) and whose particularity is unanalyzable or else is due to itself alone (which presumably is something different from being unanalyzable).⁷ It is clear that the trope-bundle bearer would not be a good particularizer. For just as trope theorists, at least the ones with which I am dealing, find it uncontroversial that there can be exactly similar properties, trope theorists also find it uncontroversial that there can be exactly similar bundles, whether these bundles consist of one property or several. Because the question would arise about what particularizes the bundles in the case where two are exactly similar, and because trope theory holds that the particularity of the properties of the bundles explains the particularity of these exactly similar bundles (explains why they are numerically distinct), this option is of no use until it is established what particularizes tropes, which is precisely what we are now trying to do. In light of this, let us just consider the option where the bearer is a bare particular.

Can the bare particular bearers of x_1 and x_2 secure their being merely exactly similar?⁸ First, it is reasonable to say that x_1 is distinct from its bare particular (option α). After all, a bare particular is not itself a property. By x_1 being particularized by the bare particular, then, we must mean that x_1 is particularized insofar as it is tied to something that is distinct from itself. Now, since x_1 is of course not nothing, we are attending to something when we attend only to x_1 . The thing is, when we attend only to

⁷ C. B. Martin (1980) is the famous advocate of the Lockean view that bare particulars are the bearers of particularized properties.

⁸ Rojek thinks that bare particular bearers particularize tropes (2008: 364). Leftow also thinks this, claiming for example that “Cain’s humanity is distinct from Abel’s just because it is Cain’s, not Abel’s” (1999: 203). Kim seems to suggest that Denkel individuates tropes by their bearers (2000: 159) and he says that Campbell never considers this option (149). Although they do not agree that it is successful, Levinson (2006: 578) and Trettin (2002: 509) suggest this as being one of the particularizers of a trope.

x1 in this case, we are attending to that something as it is independent of being particularized. We are, in short, attending to a universal.

Second, if we make x1 have an essential directedness towards its bare particular (option β), then we rule out both the bare particular as the particularizer and the possibility of x1 being exactly similar to x2. For you see, x2 is by hypothesis intrinsically oriented to some *other* bare particular than x1 is, which thus makes the nature that it is dissimilar to the nature that x1 is.

Third, making the bare particular intrinsic to x1 (option γ) is doomed to fail as well. For if we make the bare particular a portion of x1, whatever exceeds that portion will be universal. And because a bare particular is not a feature, we surely cannot make it a feature of x1. Even if we could, when we attend solely to that which has the particularizing feature we will be attending to a universal.

Fourth, for trope theory to make each nature merely conceptually distinct from its bare particular (option δ) can only be for trope theory to say (4B) that outside of the intellect there is the nature but not the bare particular, and there is no basis for a distinction between the nature and the bare particular in the reality outside of the intellect, namely, in the nature. But since the bare particular bearers of natures x1 and x2 are different, and since the conception of the particularizing bare particular bearer is nothing but a conception of the nature it particularizes, natures x1 and x2 must be different. And since they must be different due to nothing but themselves alone, that is, since they must be different *qua nature, qua qualitative content*, then this can only entail that they are not exactly similar (as I will explain in further detail in section 3.2, where I explicitly consider the view that a property is self-particularized).⁹

3.1.2. Spatiotemporal location bearer

Let us move on now to spatiotemporal locations as the particularizers of tropes, where this means that two exactly similar properties are numerically distinct because they are at different spatiotemporal locations, that is, are at a non-zero distance from each other in time and space. So, can the spatiotemporal locations of x1 and x2 secure their being

⁹ Besides, it seems impossible for a property to be the truthmaker for its truly being said to be a bare particular. Since a bare particular is a non-property, and since the property is indeed a property, this predication cannot be true.

merely exactly similar?¹⁰ Whether we take a substantival or a relational view of spacetime,¹¹ the answer is no for pretty much the same reasons that we already saw.¹²

First, if the locations are extrinsic to the natures (option α),¹³ then when we attend only to x_1 and x_2 we are thereby attending to one nature. Second, if we make the natures have an essential directedness towards their different locations (option β), then we rule out both the locations themselves as the particularizers and we disallow the exact similarity between x_1 and x_2 . For you see, x_1 and x_2 are thus intrinsically oriented to different locations, which makes them dissimilar in a certain way. Third, even if we make the locations intrinsic to the natures (option γ), whether by making them portions of the natures or features of the natures, there still remains an excessive element of universality. Fourth, for trope theory to make each nature merely conceptually distinct from its location (option δ) can only be for trope theory to say (4B) that outside of the intellect there is the nature but not the location, and that there is no basis for a distinction between the nature and the location in the reality outside of the intellect, namely, in the nature. But since the locations of natures x_1 and x_2 are different, and since the conception of the particularizing location is nothing but a conception of the nature it particularizes, then natures x_1 and x_2 must be different. Just as we saw with option β , since they must be different due to nothing but themselves alone (that is, since they must be different *qua nature*, *qua qualitative content*), it would be incorrect to say that x_1 and x_2 are exactly similar (as I will explain in more detail right now).

3.2. *Property itself as the particularizer*

Let us move on now to the view that natures are self-particularized. So, can the very being of x_1 and x_2 secure their being merely exactly similar? When we say that the particularity of a nature is a function of the nature alone, that is clearly not going to

¹⁰ Stout supposedly held this view (see Moreland 2001: 51). Famous for this view of course is the early Campbell (1997: 136). Harré seems to espouse this particularizer too, speaking of this redness-here-now (2009: 98). Davidson suggests, but does not agree, that the numerical distinctness of events might be explained by a difference in their spatiotemporal locations (1980: 306). Although he does not think that it would be helpful to trope theory, Levinson also suggests this as a particularizer (2006: 578-579).

¹¹ For the distinction between the relational and the substantival theory of spacetime in regards to individuating tropes by location, see Schaffer (2001: 251).

¹² Note that, for the sake of ease, throughout the rest of this section I will focus merely on *spatial* location.

¹³ Kim seems to think that this is the only option. He complains that it is wrong to have tropes be individuated by location because location is extrinsic to the trope (2000: 177).

mean that something external to the nature has particularity (option α) or that merely some portion or feature of the nature has particularity (option γ). Moreover, we would not take this to mean that the property has an essential directedness towards something else that has particularity (option β) (although this in effect does seem to entail that the property is self-particularized). What we are left with, in effect, is that there is a distinction of reason between a property (a nature, a qualitative content) and its particularity, where this is understood in terms of option 4B: outside of the intellect there is the nature but not in addition some particularity, and there is also no basis for a distinction between the nature and the particularity in the mind-independent nature.¹⁴

But there is still a problem even here. Since the particularities of natures x_1 and x_2 are different, and since the conception of the particularity is nothing but a conception of the nature it particularizes, then natures x_1 and x_2 must be different. That is fine for the trope theory—indeed, it is precisely what trope theory wanted: ontological individuation. But since, in other words, x_1 and x_2 must be different due to nothing but themselves alone (that is, since they must be different *qua nature, qua qualitative content*), that can mean nothing else but that they are dissimilar *qua nature*, in which case it would be incorrect to say that x_1 and x_2 are exactly similar.

Am I right about this, as I have been taking for granted? Well, I cannot see why their qualitative difference, which is due to nothing but the *mere* qualities that they themselves are, would not guarantee their qualitative dissimilarity. Any gap between difference and dissimilarity closes at this point, no? If this orangeness is different than that orangeness due to nothing else but the orangenesses themselves, then that can only mean, it seems, that they themselves are not exactly similar, whatever might have been assumed and however resembling they may seem. This is not an a posteriori affair. The two whatnesses are two on this view due to nothing other than the whatnesses themselves—not due to inhering in different substrata or being in different locations. In other words, each of these properties are particular due to nothing other than themselves alone. But what can it be about mere orangeness, orangeness alone (not its location or whatever), that makes it different from some other orangeness? It can only be that it is a different “shade” of orangeness than the other orangeness. Put it this way. If the *mere simple quality itself* is sole ground for its particularity (in short, if the quality is self-

¹⁴ It may very well be that this is what Campbell, in his revised trope theory, takes to be the right view (see Moreland 2001: 60).

particularized), then the ground for the particularity can be nothing else but qualitative. But if the ground for the particularity of a mere simple quality is qualitative, then the ground for the numerical difference of purportedly exactly similar qualities is of course qualitative. But if the ground for the numerical difference of purportedly exactly similar qualities is qualitative, then that means that there is a qualitative difference, and thus qualitative dissimilarity, between them. So while having a property be self-particularized guarantees its particularity, the expense is that the possibility of exact resemblance between such properties is ruled out.

4. Concluding remarks

I have not argued that the concept of a particularized property, a trope, is at bottom inconsistent. I have not ruled out the possibility that properties are particulars rather than universals. Aside from delineating each of the ways that property-particularizers might be employed to perform their jobs, my aim in this paper was to explain why I think that in the few ways to employ a particularizer that actually succeed in having the property be particularized, that is, be a trope, the possibility of exact similarity between tropes is disallowed. My aim, in other words, was to explain why I think that a particularizer's success in getting rid of any universal element of the property results in the impossibility of exact similarity between the properties.

Might the realist be faced with a problem analogous to that faced by the trope theorist who regards properties as analyzably particular? If we are going to ask the trope theorist to explain how exactly similar properties are not in truth simply one property, should we not also ask the realist to explain how multiple property instances are not in truth simply multiple properties? We might wonder, in effect, whether there can be a successful universalizer for a property. This question may be just as well motivated as my question as to whether there can be a successful particularizer for exactly similar tropes. For whereas I find it odd that mere whatnesses can be indiscernible and yet not identical (even though, for example, indiscernible mass properties occupy and express one and the same position on the mass spectrum), the trope theorist might find it odd that one thing can be wholly expressed through two nonconcurrent entities at once. The realist's answer to this question is going to be that the property is its own universalizer; it is self-universalized. At the end of the previous section I discussed the problem that

trope theory would face if it regarded properties as self-particularized. The problem was that self-particularization secures the ontological distinctness of a property at the expense of ruling out the possibility of a property being exactly similar to any other property. Does the realist face an analogous problem by regarding properties as self-universalized? Self-universalization secures the property's ability to be had by multiple entities at one and the same time (such that the entities with this property will be strictly identical in terms of it), but there is no analogous expense from what I can see.¹⁵ The disanalogy seems telling. We would expect self-particularization to do the job for trope theory if it is going to be a viable alternative to realism. But if demanding that properties be considered self-particularized was not bad enough (well, bad at least from the perspective of the realist), trope theory must demand even more if it is going to preserve the possibility of exactly similar tropes.

What more is that? The only option left, as should now be clear, is for trope theorists to regard properties as *unanalyzably* particular, particular due to nothing—not even themselves.¹⁶ Such a move is considered the last bastion of hope for trope theory (mainly in light of the problems that I have consolidated in this paper). One may be prone to think, however, that pulling the brute-fact card here is in fact no help for the trope theorist (and not only for the reason that, as I just pointed out, we would expect properties to be *self*-particularized—not *brute*-particularized—if trope theory is to be a viable alternative to realism). Here is why. Since a nature being primitively particular entails that it is wrong to hold even that the nature itself provides for its own

¹⁵ One may say that there are expenses that come with holding the realist view in general, though. For example, the redness here in this apple is spinning (insofar as this apple is spinning) whereas the redness there in that apple is stationary (insofar as that apple is stationary), such that redness is both moving and not moving. There are several replies that can be made to this. Just as we say that Descartes insofar as he is 25-years-old-and-a-day is bearded whereas Descartes insofar as he is 26-years-old-and-a-day is shaved, we might say that property *x* insofar as it is over here is spinning whereas *x* insofar as it is over there is stationary. There is nothing odd about me saying “See that clean-shaven man over there, he was bearded a year ago.” Likewise there is nothing odd about me saying “See property *x* spinning over there, it is stationary over here.” When I say “*x* is spinning at place *p1*” I mean that *x* has the property *spinning at p1*, and when I say that “*x* is stationary at place *p2*” I mean that *x* has the property *being stationary at p2*. The properties of *x* are, in short, place-and-time indexed. We clearly have one and the same *x* in the two places at the same time, and yet there is no contradiction. There would only be contradiction if *x* was both spinning and not spinning *at the same place and time*. This could never happen, though. There cannot be two instances of *x* at the same place and time: a purported two instances in the same place and time would be one instance according to realism.—Or perhaps one might take the Moreland-Wolterstorff line and say that universals are not really located, in which case it would not be true that it is both spinning and stationary.

¹⁶ D. C. Williams famously holds the view that the particularity of tropes is unanalyzable (1966). Maurin too takes particularity to be an unanalyzable fact about tropes (see 2002: 16-21). Trettin (2002: 509) and Livanios (2007: 365) suggest this as an option.

particularity (an easy point to forget), it follows that a nature being primitively particular entails that it is right to hold that the nature itself does *not* provide for its own particularity. But if a nature itself does not provide for its own particularity, then that nature—considered in itself—just is a universal! So not only is trope theory saddled with the odd position that nothing, not even the property itself, provides for the particularity of a property (which is a problem not faced analogously by realists since universals are self-universalized), this position seems to undermine trope theory.

Does this argument involve slippery reasoning? It may be motivated by an explanatory rationalist's scorn for brute facts, and I admit that such a temperament is out of touch with how disputes in metaphysics work these days (where everyone grants that brute-fact cards will be pulled by each competing theory). Yet I cannot shake being compelled by the argument. If the property is neither self-particularized nor other-particularized, then surely it is appropriate to say that the property in itself is *not* particularized.

Even if I am wrong about this, it does seem odd for the brute-fact card to be pulled right on the very contended issue as to whether properties are particular or universal. Perhaps it is only odd from the realist perspective. This I am self-critical enough to admit. Nevertheless, one advantage realism has over trope theory in this regard is that it does not do the same: properties for it are *self*-universalized, not *brute*-universalized. Explanation has to stop somewhere, yes. And that is why I think we are so self-consciously open nowadays to pulling the brute-fact card. Of course, the point where an explanation stops need not be a brute fact. Instead of stopping at the *unexplained*, explanation can stop at the *self*-explained. Not all endpoints to explanation are created equal, I think. The more “virtuous” buckstopping point is the self-explained rather than the non-explained (or, more accurately, the nothing). Both stopping points may leave people dissatisfied, to be sure. But since it is so repugnant to say that there is something that is explained neither by itself nor by any other,¹⁷ we ought to be more dissatisfied with the strictly brute stopping point. So I take it as a major mark against trope theory that it cannot take the more virtuous path of having properties be self-particularized without thereby undermining (for reasons I explained in the previous section) the very possibility of exact similarity between particularized properties. When

¹⁷ According to the Copenhagen interpretation of quantum mechanics, there seem to be cases where commonsense is downright wrong in this regard. So perhaps I am speaking a bit recklessly.

we ultimately measure the two theories, realism and trope theory, against each other (at the forever-to-come time where we will consider which has more explanatory power, satisfies more desiderata, is more economical, pulls less primitive-fact cards, and so on), we should perhaps note not merely how many endpoints to explanation are posited by a theory, but what those endpoints are. Are they self-explained or non-explained? Self-explained termini count much less against a theory than unexplained, that is, brute, termini.¹⁸

M. A. Istvan Jr.

Texas A&M University

maistvanjr@gmail.com

References

- Campbell, K. (1997) 'The Metaphysics of Abstract Particulars', IN *Properties*, ed. D.H. Mellor and A. Oliver, Oxford: Oxford University Press.
- _____. (1990) *Abstract Particulars*, Oxford: Blackwell.
- Davidson, D. (1980) 'The Individuation of Events', IN *Essays on Actions and Events*, Oxford: Clarendon Press.
- Harré R. (2009) 'Trope Theory and the Ontology of Chemistry', *Foundations of Chemistry*, 11.2: 93-103.
- Kim, H-J. (2000) *On Trope Theory*, Dissertation (PhD)—University of Iowa.
- Leftow, B. (1999) 'Anti Social Trinitarianism', IN *The Trinity: An Interdisciplinary Symposium on the Trinity*, ed. S. Davis, D. Kendall, & G. O'Collins, Oxford: Oxford University Press.
- Levinson, J. (2006) 'Why There Are No Tropes', *Philosophy* 81.4: 563-580.
- Livanios, V. (2007) 'Tropes, Particularity, and Space-Time', *Journal for General Philosophy of Science* 38.2: 357-368.
- Martin, C. B. (1980) 'Substance Substantiated', *Australasian Journal of Philosophy* 58.1: 3-10.
- Maurin, A-S. (2002) *If Tropes*, Synthese library, v. 308, Dordrecht: Kluwer Academic.
- Moreland, J.P. (1989) 'Keith Campbell and the Trope View of Predication', *Australasian Journal of Philosophy* 67: 379-93.
- _____. (2001) 'Moderate Nominalism and Properties', IN *Universals*, Montreal: McGill-Queen's University Press.

¹⁸ I thank Robert Garcia, Jonathan Schaffer, an anonymous referee for *Abstracta*, and especially Alesha Istvan for their help on this project.

- Morganti, M. (2009) 'Tropes and Physics', *Grazer Philosophische Studien* 78: 185–205.
- Rojek, P. (2008) 'Three Trope Theories', *Axiomathes* 18.3: 359-377.
- Trettin, K. (2002) 'Tropes and Time', IN *Argument und analyse*, ed. Beckermann, Ansgar, and Christian Nimtz, Paderborn: mentis: 506-515.
- Schaffer, J. (2001) 'The Individuation of Tropes', *Australasian Journal of Philosophy* 79.2: 247-257.
- Simons, P. (1997) 'On Being the Same Ship(s)—Or Electron(s): Reply to Hughes', *Mind* 106.424: 761-767.
- Suarez, F. (1947) *Metaphysical Disputations 7: On the Various Kinds of Distinctions*. Ed. Cyril Vollert. Milwaukee: Marquette University Press.
- Van Inwagen, P. (1983) *An Essay on Free Will*. Oxford: Clarendon Press.
- Williams, D. C. (1966) 'The Elements of Being I', IN *Principles of Empirical Realism: Philosophical Essay*, Springfield, Ill: C.C. Thomas.

**NEW CRITERIA FOR PAIN:
ORDINARY LANGUAGE, OTHER MINDS, AND THE GRAMMAR OF
SENSATION**

Kieran Cashell

Abstract

What does ordinary language philosophy contribute to the solution of the problems it diagnoses as violations of linguistic use? One of its biggest challenges has been to account for the epistemic asymmetry of mental states experienced by the subject of those states and the application of psychological properties to others. The epistemology of other minds appears far from resolved with reference to how sensation words are used in everyday language. In this paper, I revisit the Wittgensteinian arguments and show how they engage the ordinary language method (in the modified form of grammatical investigation) to ‘dissolve’ the problem. Several important results are generated by way of this reconstruction. An expressive view of the vocabulary of sensation is defended which facilitates a discussion of sensation discourse emphasising the normative grammatical conditions for the communication of psychological states. This motivates a reassessment of criterial justification for the ascription of psychological concepts in the third person. In the final sections, I mobilise a normative approach to expose the moral relevance of the epistemology of other minds. Even if it is conceded that belief in other minds lacks warrant from an epistemological standpoint, this does not justify adopting the skeptical attitude from an ethical standpoint. In light of this, a normative justification for the a priori belief that others are subjects of consciousness is defended.

It wasn't the exotic I was after, but the
ordinary, that strangest and most elusive of
enigmas.
- John Banville

ὁ ἀναξ, οὐδὲ τὸ μαντεῖόν ἐστι τὸ ἐν Δελφοῖς,
οὔτε λέγει οὔτε κρύπτει, ἀλλὰ σημαίνει.
- Heraclitus (DK 93)¹

Ordinary Language Philosophy (OLP) is a critical method that treats philosophical problems as a consequence of alienating language from its communicative environment. Words with perfectly adequate colloquial meanings, when inducted into the philosophical context, acquire enigmatic, precisely *extraordinary* qualities. The objective of OLP is to deflate the bogus profundity produced by this alienation-effect by insisting that the meaning of a word is inextricable from its everyday communicative

¹ “The lord whose oracle is in Delphi neither speaks out nor conceals, but gives a sign.” (in Kirk and Raven's translation, 1964: 211). Diels Kranz fragment 93.

milieu. Although associated with post-war Oxford philosophers (in particular Ryle and Austin) the characteristic features of OLP arguably originated in Wittgenstein's seminars at Cambridge in the 1930s. In the *Blue Book*, for instance, there is the following advice: 'The thing to do in such cases is always to look at how the words in question *are actually used in our language*'.² Philosophic perplexity is attributed to a peculiar disposition to view concepts as inherently problematic. Yet Wittgenstein cautions: 'We are in all such cases thinking of a use different from that which our ordinary language makes of the words ... a use which just then for some reason strongly recommends itself to us'.³ Stanley Cavell has consistently emphasised the role of the ordinary in Wittgenstein's later philosophy: 'The ordinary occurs in *Philosophical Investigations*' he recently stated, 'as what skepticism denies, and metaphysics transcends'.⁴ Even if it remains controversial to identify Wittgenstein exclusively with the approach, Cavell's emphasis is, I believe, correct. Wittgenstein's appeal to utility is identified as an attempt to retrieve words from their philosophic alienation and repatriate them in their lay 'habitat'.⁵

When philosophers use a word – “knowledge”, “being”, “object”, “I”, “proposition”, “name” – and try to grasp the *essence* of the thing, one must always ask oneself: is the word ever actually used in this way in the language-game which is its original home? (PI §116)⁶

'What *we* do' the paragraph concludes, 'is to bring words back from their metaphysical to their everyday use'. According to this motive, the method constitutes, in Cavell's words, 'the welcome idea of returning words to the circulation of language ... rather than keeping them fixated in some imaginary service'.⁷

Returning words to linguistic circulation is fine: but what does OLP contribute to the relief of the problems it diagnoses as transgressions of ordinary use? One of its biggest challenges has been to settle the perennial conflict between 'private' more or

² Wittgenstein (1958: 6).

³ Ibid.: 56. See also Wittgenstein (1953: §89) (paragraph numbers in text are 1953 unless otherwise indicated).

⁴ Cavell (2005: 195).

⁵ For this translation of *Heimat*, see Nielsen (1958: 119).

⁶ Cavell draws attention to connotations of the German words '*Sprache*' and '*Heimat*' lost in translation to the technical 'language-game' and 'original home' respectively. The image evoked when we attend to these words in context is not one of a philosopher who refuses to play 'the game of the ordinary' but rather someone who casts 'words into exile'. Cavell (2005: 197-98).

⁷ Ibid.: 199.

less completely hidden psychological states and public behavioural manifestations of agency. How can it be claimed, for instance, that how I feel about him, when it seems so easy to conceal, is completely “open to view” (§435)? Such claims appear counter-intuitive, contrary to commonsense and not just common linguistic usage. The problem of other minds, as encapsulated by the epistemic asymmetry between mental states experienced by the subject of those states and the application of psychological properties to others, appears far from resolved with reference to how sensation words are employed in colloquial language. Yet Wittgenstein (according to some of his most dedicated commentators) accomplished a convincing resolution of this problem. Apropos the problem of other minds, I survey the Wittgensteinian argument and, with reference to the key commentaries, show how it engages the ordinary language method (albeit with important modifications) to ‘dissolve’ the problem. Wittgenstein’s emphasis on the grammatical conventions that determine how concepts are employed in ordinary contexts contains a strong, if somewhat incipient, normative character. In the final sections, motivated by Cavell’s re-negotiation of philosophical fields (itself inspired by his reading of Wittgenstein’s *Investigations*), I amplify this normative character in a way intended to reveal the moral relevance of the problem. Even if it is conceded that belief in other minds is ultimately unwarranted, I argue that this does not prevent alternative justification being sought outside the epistemological domain. To this end, a normative solution to the problem of other minds is proposed.

Other Minds and the Argument from Analogy

Recently described as the ‘most challenging of problems about consciousness,’⁸ the epistemology of other minds seeks justification for the folk-psychological conviction that non-autobiographical mental states exist. The subjective perspective that affords conscious awareness of my own psychological processes, being unique, is ruled out a priori for other people: where my intentions, emotions and sensations are accompanied by an implicit “I know ...” I am forced to infer to the best explanation⁹ on the evidence of ambiguous physiological data that other subjects experience states of consciousness. Thus it is my restricted experiential (and hence epistemic) scope that facilitates the suspension of belief concerning other minds. April, the skeptic informs me, may turn

⁸ Noë (2009: 25).

⁹ Hacker (1997: 32).

out to be a sophisticated android. Such sci-fi hypotheses derive their persuasive power in philosophical discussion by presupposing that the only valid knowledge of conscious experience is secured by introspection. This, it seems, motivates the doubt that psychological predicates, strictly speaking, apply to others. If I, logically, begin to doubt that you cannot experience what I experience then my tendency to ascribe intentionality, emotion and sensation to you may begin to appear, at best, a fiduciary inclination that does not (and cannot) be justified evidentially, but rather has, at most, the support of doxological compulsion. ‘The reason why I cannot directly know the experiences of another’ A. J. Ayer observes ‘is simply that I cannot have them’.¹⁰ Others are compelled to hypothesise, on the basis of circumstantial evidence, what is available to me *a priori*. As such, any proposition about my own mental state, uttered by me, as I am subject to it, is uncontroversial, any proposition about another person’s state (by me or by them *of me*), on the other hand, is epistemically controversial. Hence the epistemological asymmetry (identified by Stuart Hampshire) of autobiographical statements (author-subject identity) and ‘heterobiographical’ statements (author-subject heterogeneity)¹¹ distinguishes cogito-type certainty in the first instance from inferential and thus inherently dubitable conjecture in the other. One and only one subject, it seems, is in the position to know the truth conditions of sensation predicates; and that subject is, of course, me: private states, amenable to verification only via the process of introspection, *ex hypothesi*, logically exclude verification in cases that transcend mine (i.e., *all* other cases).

What is it not possible to achieve here? Omniscience? The capacity, that is, to directly experience, as in a realist novel or movie, the private psychological life of someone else (as I would be in a position to do if I shared their perspectival subjectivity)? This kind of phenomenological privilege, in practice, is excluded not only by logic but by physiological facticity: embodiment restricts subjectivity to a single perspective-point of consciousness. Because necessarily alternative to *my* viewpoint, I can never know (i.e., be certain about) the private thoughts, intentions, sensations or emotions of another.¹² ‘The idea [of this incapability] is’, Wittgenstein says in the *Blue Book*, that even if ‘the same object may be before his eyes and mine ...

¹⁰ Ayer (1954: 194).

¹¹ Hampshire (1952: 2).

¹² Malcolm (1977: 135-6).

I can't stick my head into his (or my mind into his, which comes to the same) so that the *real* and *immediate* object of his vision becomes the real and immediate object of my vision too'.¹³ Skepticism about other minds originates in the suspicion that others are radically 'closed off from me (within, as it were, *their* own experience).'¹⁴ I am, it seems, destined, like Sylvia Plath's Esther, to remain forever shut inside the bell jar of subjectivity, 'stewing in my own sour air.'¹⁵

Attributed to Mill and later defended by Russell (and Ayer),¹⁶ the Argument from Analogy suggests a way out of the bell jar of subjectivity. The argument renders intuitions about other minds conceptually respectable by proposing to logically extend what I know of my own experience to others. Drawing on the evidence of common existential features, it seems reasonable to infer that other people, like me, in all probability, are 'animated' by consciousness. Thus by way of this analogy, the problem of other minds becomes more tractable: we can take what we know from our own experience (of the causal correlation between psychological cause and behavioural effect) and extrapolate from this to the probable existence of other minds. So although I may not be in a position to immediately observe my various alter-egos' psychological states (by telepathy?), I can nevertheless read intentional motivation back into their overt behaviour and extrapolate to the causal mental agency underlying it. It seems *prima facie* reasonable, that is, to compensate for the lack of phenomenological verification for mental states other than our own by transferring what we *know* a priori (regarding the role of intentionality, sensation and emotion in determining my consequent behaviour) to explain another person's observed agency, and establish (even if I can never *ultimately know*) that, like mine, April's behaviour must have consciousness as its causal (if unobserved) antecedent. The argument from analogy, Russell concludes, thus logically justifies the inference to other minds.

Not quite. It is precisely the deficiencies of this inference that the Wittgensteinian critique undermines. Indeed, *Philosophical Investigations* has become a *locus classicus* of skepticism about the predication of psychological states to others by analogy. Wittgenstein shatters our confidence that we identify mental states by

¹³ Wittgenstein (1958: 61).

¹⁴ Cavell (1979: 161); quoted in Tanesini (2001: 14).

¹⁵ Plath (1963: 178). For a similar sentiment see Kołakowski (1988: 61).

¹⁶ Mill (1865); Russell (1948); Ayer (1953); see also Hampshire (1952) and Price (1938).

extending previously-identified private experiences to others. He drives in the skeptical wedge as follows:

If one has to imagine someone else's pain on the model of one's own, this is none too easy a thing to do: for I have to imagine pain which I *do not feel* on the model of the pain which I *do feel*. That is, what I have to do is not simply to make a transition in imagination from one place of pain to another. (§302)

Wittgenstein's caveat suggests that 'abstraction from the paradigm' fails in this instance because the counterfactual move involved, rather than establishing the presence of sensation via something I *do* experience to something I don't (as from one painful somatic location to another), infers, from the imaginary idea of a sensation (which *qua* idea is not *felt*), another instance of unfelt sensation (the other person's pain). This is the difficulty identified in paragraph 302. Although the inference is assumed to employ sensation previously experienced (and felt) – but not *now* experienced or felt – as paradigm, this is not merely an application of past experience of pain to present instance of pain (application of prior acquired concept to new, but still radically different, case) (§§448-449). Rather what is involved here seems more intractable because it presupposes the Cartesian picture of mind as a hermetic enclosure containing a cache of 'objects' accessible only to the subject in whose body the enclosure is metaphysically embedded. The subject, searching inwardly, identifies the relevant object and retrieves it in order to identify, by comparison, the presence of the *same* sensation in the other person whose odd behaviour must otherwise appear a stylised and unintelligible mime.

The analogical inference from my concept to another person's sensation is inadequate to its target not only because it implies what Kripke has termed a 'behaviourist ersatz for imagining the sensation of others on the model of my own'¹⁷; but rather because the inference involves the application of a type of concept (a cogito-concept) to a qualitatively different epistemic scenario (physiological behaviour) presupposed to be a token instance of the type. Yet this comparative transition is theoretically compromised by the very theory of mind that seems to support it. In other words, the logical difficulty with analogy is not merely that there is insufficient evidence for the inference to an underdetermined cause (never mind that it is a very weak sort of inductive argument based on reasoning from a single instance) it is rather

¹⁷ Kripke (1982: 125).

that one type of knowledge – non-inferential, direct and indubitable (call it Cartesian) – is prioritised as paradigmatic and partially induced out of an entirely different type – inferential, indirect, and defeasible (call it behaviourist) – according to a tacitly endorsed theoretical apparatus that has categorically separated the type (mental process) from the token (bodily behaviour). Yet the epistemic asymmetry operative here tacitly accepts that psychological states are *necessarily* private (and therefore radically inaccessible to external observers). On the premises of the Cartesian presuppositions, the conclusion that the publicly observed behaviour of others is *caused* by private conscious events necessarily involves a troublesome (and even ultimately inconsistent) leap of faith.

The difficulty about the primacy of private experience, as Wittgenstein observes, is ‘not that each person possesses his own exemplar, but that nobody knows whether other people also have *this* or something else’ (§272). Skepticism gains its intuitive plausibility by nourishing the epistemological insecurities associated with the post-Cartesian theoretical perspective according to which it is impossible not only to verify the existence of other ego cogitos (a phrase that, incidentally, doesn’t make sense in the plural)¹⁸ but also to justify the inference from something observed (surface behaviour) to something unobserved (deep consciousness). Indeed, the trouble with the analogical argument, as Malcolm observes, is the captivating picture that what is necessary for knowledge of mental states is phenomenological first-hand experience: that I learn about the existence of psychological attitudes only by introspection. This seems, he comments, to be ‘the most natural assumption for a philosopher to make and indeed seems at first to be the only possibility’.¹⁹

Yet is it credible that the correlation of “inner” sensation and “outward” behaviour is established via self-observation? ‘*I* cannot be said to learn of [my sensations],’ Wittgenstein remarks, ‘*I have* them’ (§246). If we insist that we *do* identify sensation through internal observation – by introspection (‘private ostensive definition’ or ‘pointing-into-yourself’ [§380]) – then, Wittgenstein argues, we abandon the very possibility of verification prior to the application to others that the argument from analogy demands. Private demonstration (‘this’) abolishes independent standards

¹⁸ See Kolakowski (1988: 61): ‘It is incontestable that the *Cogito* could be expressed only in the first person singular; in any other grammatical form ... it becomes an absurdity.’

¹⁹ Malcolm (1958: 974).

for distinguishing between success and failure because even if I could identify a sensation ('this') as it rises up in me (by private indexical demonstration) I could fail to recognise it as the 'same' (as 'that') sensation because a paradigm case cannot be established in isolation of criteria that transcend my subjective arbitration (necessary even for the initial identification): 'it is not possible to obey a rule "privately"' Wittgenstein observes, 'otherwise thinking one was obeying a rule would be the same thing as obeying it' (§202). What epistemic function can a *subjective* paradigm serve? Anything I decide is "right" will be right. From its seeming to me to be so, no criterion of identification follows. 'And that only means,' Wittgenstein observes, 'that here we can't talk about "right"' (§258). On the basis of introspection, it cannot but be 'a *contradiction* to speak of *another's* pain'; my pain and your pain are, as philosophers (and *only* philosophers) say, *numerically* different.

But is it an essential characteristic of sensation that *I* have it?²⁰ If the conclusion is that *I cannot refer to pain that is not my pain* then, clearly, something has gone wrong with the reasoning (because I frequently, effortlessly and without doubting, refer to pain not my own).

The Grammar of Sensation

Wittgenstein's remarks, typically, do not contain an explicit refutation of the argument from analogy (they invoke, by implication, a family of several heterogeneous, but related, targets). What is consistent about his engagement with the problem, however, is his conviction that the epistemological difficulties have a semantic genealogy that needs to be acknowledged if progress is to be made. When I predicate a sensation of myself it seems that I report an *experience* of the sensation. But "He is in pain," despite having a syntactical similarity to such reports, states an empirical proposition to the effect that (I observe that) *he is in pain* (i.e., I am saying precisely that *I* don't *feel* pain).

Can I be said to infer from epistemically equivocal behaviour that April is angry, that Jude is anxiously preparing for an exam, or that little Hans is frightened because I have already pre-identified these states in my imagination by introspection? To construe these quotidian acts of awareness as inferences to the best explanation on the basis of inconclusive evidence seems absurd. We know that the skeptic challenges

²⁰ Malcolm (1954: 538; 1977: 119).

something that seems intuitively obvious: ordinary propositions, that is, held certain and incontrovertible in practice, but when critically scrutinised, appear to lack the epistemic justification required to secure knowledge of them. But Wittgenstein has demonstrated that *knowing* the psychological attitudes of others (assuming that this is, in principle, at least possible) does not depend on analogical inference from self-observed correlations between behaviour and ontologically distinct psychological states (§417; §357) – nor does it depend on inductive reasoning “outward” from an introspectively-verified paradigm case. Because justification requires an independent (objective) criterion, and this cannot be supplied by the introspection of private objects, analogical reasoning cannot provide an adequate response to the skeptic’s challenge. It would be as absurd to analyse my response to a child’s cries of distress in terms of a pattern of inferential reasoning (from the conditional premise: *If April cries then she’s in pain*) as it would be to confirm my own involuntary reaction to an accidental burn by *modus ponens*. In rejecting the theoretical explanation that we respond sympathetically to another person in pain ‘because by analogy with our own case we believe that he too is experiencing pain’,²¹ Wittgenstein argues that the ordinary discourse of pain is never (or never in *quotidian* circumstances) informed by such reasoning. This constitutes, for him, a deep mistake of analysis. The behavioural pattern of sympathy and response to another person’s distress (or pain) is, rather, a normative extension of nonverbal, infra-propositional instincts of concern. Indeed, the ‘language-game’ of sensation is, Wittgenstein claimed, erected on a foundation of pre-sentential behaviour around which the socio-symbolic conventional discourse is ultimately articulated.

Despite their syntactical homogeneity, however, there *is* a semantic difference between sensation-statements in the first person and their third-person counterparts. If Wittgenstein acknowledges the asymmetry of propositions in the ‘autobiographical’ and ‘heterobiographical’ modes, however, he insists that this is not due to the Cartesian distinction between the necessary privacy of mental states and their corollary epistemic inaccessibility to observers. Rather the asymmetry is a function of the grammar of ordinary language which, when reflected on through a philosophical prism (i.e., when alienated from everyday use), generates the impression of a metaphysical opposition between interior (subjective) and exterior (objective) dimensions of existence. When he

²¹ Wittgenstein (2007: §542).

says ‘we have only rejected the grammar which tries to force itself on us here’ (i.e., in the semantics of sensation), in paragraph 304 of the *Investigations*, he suggests that the specific perplexity can be attributed to the tendency to misconstrue sensation-predicates as *objects*: yet, ‘if we construe the grammar of the expression of sensation on the model of “object and designation”,’ he observes, the object always ‘drops out of consideration as irrelevant’ (§293).

The subject-predicate (S-P) structure, traditionally, predicates a property of an object. In predicating sensation-concepts of a subject, however, the logical form creates the impression that a psychological property must correspond to the predicate in the same modality as that which corresponds to the subject-term (to which the concept is applied); and this carries the implication that the psychological property, in principle, has an ontological status independent of the body; but, ‘Where our language suggests a body and there is none: there, we should like to say, is a *spirit*’ (§36).²² Sensation-predication thus creates the hypostatic effect of an inner object (which corresponds to the psychological property) privately accessible only to the subject (and *identified* by interior observation) – and leads to a precise iteration of Descartes’s inaugural hypostasis of mind as thinking *thing* distinguished from extended things. If an ontology suggesting the existence of a colourless, shapeless residual object independent of its properties strikes us as unintelligible, however, what of the existence of a body divested of psychological properties (a zombie)?

Yet if the structure of our grammar here creates the impression that sensations are a species of uncanny ‘intangible objects’, however, it is only because we have momentarily forgotten the purpose of sensation-discourse: which is, simply, to *express*. In the discourse of sensation, language is used (at least in the first-person), not to predicate a sensation-concept of a subject, but rather to express how we feel. So although superficially identical, the grammar of first-person pain-statements differs to third-person predications of pain in that it relates to the act of conveying feeling. Who is the ‘subject of pain’? Wittgenstein inquires: ‘[it is] the person who gives it expression’ (§302). This helps to explain why no criterial evidence is relied upon to verify pains that *I* feel; the word “pain” is not used by me to confirm pains that *someone*

²² See Wittgenstein (1989: 263): ‘With the idea of a predicate, goes the idea of a *property* ... Suppose I say, “This sofa is green”, then the predicate is “is green”. If I then ask what it is that has the property green, you would imagine something like a colourless sofa.’

perceives: because I don't *observe* someone in pain when I 'predicate' the sensation of myself (§290). But when I *need* to use the word "pain," that is, when I *feel* pain, I do so to express how I am feeling in a spontaneous (if not completely involuntary) way. Regarding my pains, 'I *have* them' (§246). We don't therefore 'identify the sensation by criteria: but [rather] repeat an expression' (§290). In order to bring out the distinction,²³ the *Blue Book* distinguishes between first-person and third-person statements in a manner that directly relates the grammatical form to the expressive function of sensation discourse: 'The difference,' he says, 'between the propositions "I have pain" and "he has pain" is not that of "L.W. has pain" and "Smith has pain." Rather, it corresponds to the difference between moaning and saying that someone moans'.²⁴ In the discourse of sensation, language is deployed as a semantic conduit for the expression of mental states.

In the *Investigations*, the grammar of sensation is developed by recalling how we learned to refer to sensations, by considering, specifically, how we train children to verbalise their feelings. 'A child has hurt himself' Wittgenstein observes 'and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child', he concludes, 'new pain-behaviour' (§244). The expressive view therefore holds that, as we develop, we learn to mediate instinctive behavioural expressions of pain (such as crying, groaning etc.) *through* language. So the pain-sentence comes to *represent* a new and, importantly, more *controlled* means through which 'the child evinces his pain.' Semantic expressions of pain, as Kripke puts it, can be considered, therefore, to constitute 'more sophisticated, pain behaviour that adults teach the child as a substitute for the primitive, non-verbal expression of pain'.²⁵ A semantic expression of pain 'is not made on the basis of any special application of criteria any more than a cry is. In the most primitive case, it *escapes* from the speaker'.²⁶ Children are encouraged to recognise sensation-categories (hunger, toilet, fear, cold, warmth, as well as pain) independently *when they feel them* and to respond to them in the appropriate (socially conditioned) manner.

²³ See Malcolm (2001: 73) (he calls them 'declarations').

²⁴ Wittgenstein (1958: 68) See also 69: 'We feel that in the cases in which "I" is used as subject, we don't use it because we recognise a particular person by his bodily characteristics; and this creates the illusion that we use the word to refer to something bodiless, which however, has its seat in the body. In fact this seems to be the real ego, in one of which it was said, Cogito, ergo sum.'

²⁵ Kripke (1982: 134).

²⁶ Kripke (1980: 135) (cf. §302). See also Malcolm (1954: 541).

Pain-language is related, Wittgenstein concludes, to ‘the primitive [*ursprünglichen*], the natural, expressions of the sensation and used in their place’ (§244). Pain, therefore, is not a logical construction out of behaviour, as behaviourism maintains; rather, the ‘utterance’ as Malcolm finesses, ‘is itself an expression of sensation, just as flinchings, grimaces, and outcries are expressions of sensation’.²⁷ But rather than the pain-proposition expressing the pain in an identical way to the instinctive expressive behaviour, I read Wittgenstein as arguing that the sentence, “I am in pain,” is a grammatical reconstruction of the visceral expression of pain; the instinctive, uninhibited complaints of pain *remain* causally sublimated in the sentence (yet can still escape, as we are well aware, from an adult who suddenly feels intense pain). The exclamation *ow!* is not a sophisticated substitute for the expression of pain. It is, rather, again, *just the non-propositional expression of pain*. (And similar exclamations of sensations, both painful and pleasurable, can be adduced.) The chef whose freshly sliced finger lies among the cucumber on the chopping board, it may be objected, does not have time to think of sophisticated language to express his pain; he just expresses it. Pain, if intense enough, is still expressed, even by adults, according to the classic behavioural scripts: involuntary groans, howled expletives, rapid or laboured breathing, facial grimaces, squirming, etc. (We can make accurate sketches.) What the grammar of sensation refers to, however, is the acquired capacity to mediate this visceral catharsis of pain *through language*. Yet this doesn’t imply that the semantic expression is equivalent to the physiological behaviour (“non-verbal communication”) nor is it reducible to the phenomenological experience of the sensation ‘itself’. Rather the grammar here constitutes an apparatus through which the visceral expression is mediated such that, it could be suggested, both the instinctive behaviour and the syntactical expression act as different ‘modes of presentation’ with the same reference: PAIN.²⁸

We simultaneously learn (something that seems immensely important to the species) to *suppress* the instinctual behavioural impulses (to scream, shout out, etc.)

²⁷ Malcolm (1977: 127). Elsewhere, he writes: ‘my sentences about my present sensations have the same logical status as my outcries and facial expressions’. Malcolm (1954: 542).

²⁸ In the *Zettel* the relationship between instinctual anatomical expression and the grammar of sensation is very clear: ‘Surely that this way of behaving is *prelinguistic*: that a language-game is based on it, that it is the prototype of a way of thinking and not the result of thinking. ... Our language-game is an extension of primitive behaviour.’

associated with extreme sensation; and it is as a result of this process, arguably, that the picture of the privacy of sensation develops so powerfully in us. Wittgenstein's suggestion is that the *ursprünglich* expressions of pain remain sublimated in the sophisticated grammatical articulations that eventually, but never completely, take their place. Rather, we *train* the child to *recognise* the signs, i.e., the feelings, associated with these sensations and act on them (rather than cry each time). From a very early stage in human development, linguistic competence gradually becomes entwined with (but, I would argue, never totally 'replaces') behaviour as socio-cultural conventions become efficiently internalised. At the same time, the more 'primitive' instinctual impulses become, concurrently (and highly significantly), ever tightly controlled. It is no accident that the region of the brain responsible for the affective or emotional dimension of pain (its suffering associated with the physical intensity) is also responsible for impulse management.²⁹

Human language, although built up from within, informed by, and *shaped* according to instinctive species-expression, is not taught by mapping a conceptual embryonic grammar onto its latent Chomskian syntactic iteration. Rather, as any parent will appreciate, we develop into it through a lengthy program of initiation through imitation, tedious repetition and no small amount of correction. Learning language, as Wittgenstein famously observes, is just the mastery of a technique (§199; p. 208). As children we are inculcated in the norms of successful social integration; and it is according to this social program that grammar comes to prescribe correct linguistic behaviour, like all instruction in conventions, by correcting deviation and rewarding competence. Yet if grammar is normative (intended to straighten out 'people as well as thoughts'),³⁰ we must also bear in mind (before I start to sound too behaviouristic) that we are instructed in the structures of life until such time as we 'master' these structures – which means, paradoxically perhaps, that we are moulded by the prevailing socio-cultural matrix until such time as our own individually distinctive voice is enabled to emerge (and we become capable of creative agency or, at the limit, of dissension). Wittgenstein emphasises that 'we don't use language according to strict rules – it hasn't

²⁹ That is, the prefrontal cortex. See Apkarian et al. (2001).

³⁰ Garver (1996: 151).

been taught to us by means of strict rules, either.’³¹ That is the reason why grammar in this interpretation cannot be wholly identified with the syntagmatic rules of syntax. Although the formal rules governing conventional linguistic value necessarily determine the value of the exceptional creative gesture, it is a mistake to completely reduce Wittgensteinian grammar to acquired practices of rule-following.

Grammar, in Wittgenstein’s idiom, clearly has a much wider significance than the description of syntactical rules; it is intended to capture how conventions adopted by communities ultimately codify normative communicative behaviour. Yet such conventions, significantly distinguished as constitutive and not regulative³², are never, as Cavell observes, reducible to *merely* arbitrary codes. To a certain extent, admittedly, the conventions of language *are* arbitrary (in that it is possible, *à la* Saussure, to imagine some entirely alternative system of encoding) (§§496-7); but the point is they are not relative to random alteration where ‘convenience suggests a change’ without interrupting what Cavell calls the very ‘texture of our lives’.³³ Indeed, the ‘array of conventions’ signified by Wittgenstein’s category of grammar, is provided for by what he calls the ‘form of life,’ that is, the shared culture of ‘conduct and feeling’ as codified by the natural history of the species.³⁴ If the codes were entirely arbitrary this would imply that ‘nothing in the object of the game’ determines their purpose. ‘We don’t make up the rules of these games’ Wittgenstein adds: ‘we have *inherited* [them]’;³⁵ conventions may be crystallised by consensus but Wittgenstein emphasises that this is ‘a consensus of *action*: a consensus of doing the same thing, reacting the same way’.³⁶ Expressions, when alienated from their situational – conventional, cultural, social, pragmatic – forms of life where they function according to community consensus, are semantically empty. For a word, ‘To know its meaning is to use it *in the same way* as other people do. “In the right way” means nothing’.³⁷ Ultimately, the significance of Wittgenstein’s category of grammar is its acknowledgement of the normative dimension of human nature (i.e., the recognition that human nature *is* culture).

³¹ Wittgenstein (1958: 25). Compare (1969: §475): ‘Any logic good enough for a primitive means of communication needs no apology from us. Language did not emerge from some kind of ratiocination.’

³² Garver (1996: 149).

³³ Cavell (1979: 110).

³⁴ *Ibid.*: 111 cf. §435.

³⁵ Wittgenstein (1989: 143).

³⁶ *Ibid.*: (183-184).

³⁷ *Ibid.*: 183).

Explicitly identifying grammatical investigation with the ordinary language method, Cavell admires Wittgenstein's *modus operandi* for similarly directing 'a word back from its metaphysical capture by the appeal to its everyday use'.³⁸ Seeking to disclose, by description, how meaning is embedded in efficacious social activity, in pragmatic structures prescribed by 'human customs and institutions' (§337), Wittgenstein's method aims to describe the conventional 'criteria on the basis of which the word is applied in all the ... contexts into which it fits and will be found to fit'.³⁹ Like OLP, therefore, grammatical analysis identifies 'the use of words in the language'⁴⁰ – with what it makes sense to say – but with the important modification that Wittgenstein now prioritises (at the risk of cliché) the dimension of nurture over nature.⁴¹ One learns the concept "pain" he reminds us, when one learns language (§384). Repatriated into its natural habitat, "pain" is functionally efficient and *hence* semantically adequate. Fact: we feel pain, we *suffer*. Perhaps formerly expressed in a purely instinctive way, via anatomical behaviour, the visceral expression of pain is now mediated – and the expression (*not* the suffering) sublimated – through normative forms of communicative action; and these constitute the public criteria used to ascribe psychological (private) predicates to others.

Knowledge, Criteria and Conviction

There is an irresistible inclination to think that if I am in pain I *must* know it. This seems intuitively obvious: "I must know," in this instance means, of course, "I cannot *not* know." We also instinctively imagine that other people are epistemically restricted in this regard, having to reconstruct my mental attitudes from overt behavioural agency: I choose to reveal and they know as much as I (can) show them. Yet, against this backdrop, Wittgenstein insists, counter-intuitively, even bizarrely, that the proposition, "I know I am in pain," does not make sense. What he means, however, is that, because the epistemic operator has no function here, this locution is never *actually* used (and therefore is ungrammatical, hence meaningless). As it adds nothing to the expression "I am in pain," in ordinary discourse, *I know* is obviated (§246). Wittgenstein's criterion

³⁸ Cavell (1979: 385). See also §90.

³⁹ Ibid.: 77.

⁴⁰ Wittgenstein (1974: §60).

⁴¹ See Malcolm (1954: 543).

for judging whether a sentence is meaningful was to examine the result of contradicting it: ‘a proposition makes sense if and only if its negation [also] makes sense’.⁴² Now, clearly, because the proposition, “I don’t know she’s in pain,” is as meaningful as its affirmative counterpart, because it is *something*, some belief that possesses an intentional content open to confirmation or refutation by experience; “I know she’s in pain” makes perfect sense. Indeed, it seems trivially disjunctive: either I know she’s in pain or I don’t.

The method of semantic inversion however makes the absurdity of the first-person proposition patent: “I don’t know I’m in pain.” No disjunction is possible in this case (unless we assume counterfactually that there can be hallucinations of pain⁴³). Wittgenstein exposes the absurdity further by imagining someone claiming: “Oh, I know what the word ‘pain’ means; what I don’t know is whether *this*, that I have now, is pain” (§289; §408); because it makes no sense to say – that is, there are no ordinary circumstances in which I *would actually* use the sentence – “I do not know whether I am in pain or not” – it is meaningless. Admissions of ignorance are ruled out in first-person present-tense expressions of pain because it not possible to be mistaken about my sensations (or my expressions). But if disbelief is logically excluded, if doubt ‘has no place in [this] language-game’ (§288), then, by Wittgenstein’s conditions, the expression (sublimated in propositions such as “I am in pain”) does not qualify as a knowledge-claim. This grammatical analysis reveals that it is meaningful to say of other people that *I know* they’re in pain but not to say it of myself (§246; p. 222). “I know ...” is not the kind of epistemic function it makes sense to complete with a “that clause” taking my sensation as its semantic content; therefore it is vacuous, hence unusable.

Again, his advice regarding knowledge is to consider the grammar that supports its communicative function. In other words, when and how do we use the word “know”? What kinds of statements are made about ordinary successes or failures of knowing in public discourse? What do we regard as the content of the noun “knowledge”? If we are using “to know” as the verb is ‘normally used,’ he observes in

⁴² Garver (1996: 148-149).

⁴³ See Putnam (1991: 154) who, surely correctly, writes: ‘one can have a pink elephant hallucination but one cannot have a pain hallucination ... simply because any situation a person cannot discriminate from a situation in which he himself has a pain *counts* as a situation in which he has a pain’.

the *Investigations*, then, for instance, it is evident that others frequently *know* when I am in pain (contradicting the intuition that their epistemic capacity is restricted by precisely this fact) (§246). Elsewhere, he writes, ‘I would like to reserve the expression “I know” for the cases in which it is used in normal linguistic exchange’.⁴⁴ And as it functions in ordinary communicative contexts, to know is to know *something*: I find out about something, I pass from ignorance to knowledge, I learn.⁴⁵ But a declaration of knowledge functions as such if and only if one is capable – in principle – of reflecting upon or ultimately defending *how* one knows (or could have learned) the content of the proposition against credible challenges that may undermine it (this is, perhaps, the key intuition of Gettier’s critique of the justified true belief definition of knowledge). “I know ...” may signify “I do not doubt ...”, Wittgenstein remarks, ‘but that does not mean that the words “I doubt ...” are senseless, that doubt is logically excluded’ in this environment (p. 221). A knowledge-claim (or hypothesis), therefore, is relative to the possibility of its falsification by counterexample; such propositions are negated without incoherence. Knowing something implies informative contents that inherently (and, indeed, epistemically) *include* the potential for error. Fallibility and genuine (i.e., non-Cartesian) doubt are necessarily tolerated for the sake of maintaining our concept of what it means to know something.

Now, if on the basis of certain physiological signs I begin to suspect that a certain woman is suffering pain, and if I confirm this, then I *know* she’s in pain. Although this realisation will more than likely mean that I don’t harbour any reasonable doubts – does it mean that *doubt*, as such, is excluded? What ‘do we *call* “getting to know”’ in these circumstances?⁴⁶ To answer this, we advert to what Wittgenstein controversially⁴⁷ identifies as the *criteria* that enable the correct ascription of psychological states to others on the basis of corroborating behavioural data (this is simply given: the criteria for “pain,” “anger,” “fear,” are already *in the language*’⁴⁸). If challenged, “But how do you know?” we would gesture exasperatedly in the sufferer’s direction, indicating the obvious, characteristic expressive signs, and exclaim: “look, she is clearly in pain. And *she says* she is!” Above all, therefore, the principal criterion

⁴⁴ Wittgenstein (1969: §260).

⁴⁵ Cavell (1979: 16).

⁴⁶ Wittgenstein (1958: 24).

⁴⁷ See especially Wright (1982 & 1984).

⁴⁸ Malcolm (1995: 143).

for ascribing pain to another person is her statement “I am in pain,” which, in ordinary circumstances, we have no reason to disbelieve (because, in these circumstances, we accept that her experience is incorrigible *to her*); that is, what makes her statement a criterion – as opposed to a *symptom* (a statistically correlated probability)⁴⁹ – of her sensation is the fact that it is an expression of feeling mediated spontaneously, yet also volitionally (i.e., it is *not* suppressed), *through* her words. Her statement represents a first-person, present-tense vocalisation of (her) pain, incorrigible for *her*; although mediated linguistically via the statement, it is a category mistake to consider this a knowledge-claim, a declaration of certainty or a proposition descriptive of her behaviour. What *I* witness in this case, on the other hand, is a first-person (incorrigible) expression of pain by the subject of the experience (the person actually *suffering*); I *realise* she’s in pain *by virtue of* her expression (which is incorrigible to her).⁵⁰ The crucial passage in the *Investigations* concludes with the challenge: ‘Just try – in a real case – to doubt someone else’s fear or pain’ (§303).

Skeptical doubt arises exclusively in philosophical contexts where counterfactual fictionalisations (or sci-phi scenarios) are conjured from the armchair with the sole intent of raising ingenious doubts where there would be no natural inclination to doubt. (In this context Eli Hirsch instructively disambiguates the conviction that we *ought* to doubt in the critical mode from the possession of warrant for legitimate doubt.)⁵¹ Can I be wrong about someone else being in pain? It appears, intuitively, that I can; but the crucial point is that this fallibility (in itself) does not exclude *knowing* that someone else is in pain. We must inquire, however, what reason there could be, in quotidian circumstances, to doubt it. Is it accurate to claim that I know when someone is in pain precisely *because* I can be *mistaken* about it? Again, is the impossibility of error regarding my own pains equivalent to certainty?⁵² Wittgenstein diagnoses a tendency in this context to equivocate between the schemas of sensation and knowledge, an attempt (in the philosophical milieu) to construe private

⁴⁹ See the distinction between symptom and criterion in Wittgenstein (1958: 24-25).

⁵⁰ This point, it should be noted, does not beg the question: I’m not presupposing she’s in pain by assuming that what I witness is an expression of pain; rather, to repeat, I see she’s in pain *by virtue of* her expression. By incorrigible I mean (contra Armstrong 1963) *immunity to error through misidentification* (IEM) (Evans: 1982). See also Wittgenstein (1958: 67).

⁵¹ Thus it is ‘alienated’ doubt or ‘pseudo-doubt’ which cannot possibly occur in the absence of the real belief in an external reality’ (Hirsch 2011: 21).

⁵² ‘Ironically’ Priest remarks ‘it is the possibility of [genuine] doubt rather than the fact of absolute certainty that allows the use of “to know”’. Priest (1992: 60).

mental state as the epistemic paradigm and consequently to discover a form of public knowledge as *immune to doubt as feeling a sensation* (i.e., to raise the gut-feeling of *being sure* to epistemological status). It is for this reason that he regards Moore's known-with-certainty truisms ('(I know) my body exists' (I know that) 'Many humans other than myself exist', 'Here is one hand' etc.)⁵³ as anomalous uses of "I know ...". Moore treats the kind of "knowledge" he has in mind as immune to doubt, as incorrigible, that is, as a sensation. He doesn't *feel* doubt.⁵⁴ 'And this is because he wants to give himself the experience [the feeling] of knowing.'⁵⁵ This is why Moore's a priori propositions exhibit, as a result, Malcolm observes, 'a surprising amount of the logic of first-person declarations of sensation, feeling, or mood.'⁵⁶

In *On Certainty* Wittgenstein argues that Moore misconstrues a 'hinge' statement (an a priori belief) as an indexical proposition known with certainty; but Wittgenstein responds, we cannot claim to *know* these fundamental 'hinges' at all for, as Wright that later demonstrated, such 'unearned certainties' actually constitute the conditional grounds of all a posteriori epistemic achievement. The Moorean propositions may be regarded as certain but it is misconceived to surmise that the beliefs expressed by these propositions enjoy internal evidential support and hence are *known*. Hinges, according to Duncan Pritchard, 'are not evidentially grounded (since nothing is more certain than a hinge proposition)',⁵⁷ yet we may be entitled to regard them nevertheless as legitimate a priori beliefs despite their recalcitrance to external epistemic justification. *Life* involves accepting unwarranted presuppositions that remain unsupported by evidentiary conditions: 'rational agency is not just an optional aspect of our lives', Wright reads Wittgenstein's last writings as arguing, 'we are entitled – save when there is specific evidence to the contrary – to make the presuppositions that need to be made in living out our conception of the kind of world we inhabit and the kinds of cognitive powers we possess'.⁵⁸

In §377 Wittgenstein states unequivocally that the criteria of another person's sensations are, for me, just 'what he says and does.' And according to these criteria,

⁵³ See Moore (1966 & 1998). See also Malcolm (1964).

⁵⁴ Wittgenstein (1969: §178).

⁵⁵ Malcolm (2001: 73).

⁵⁶ Malcolm (1977: 190).

⁵⁷ Pritchard (2005: 196-97).

⁵⁸ Wright (2002: 41).

therefore, if we still doubt that she's in pain, having seen and heard (and understood) her expressions, then there must remain viable reason for disbelief. This is not to deny that there could be viable reason; but it is to claim that whatever reason there is will *not* be cause for *skeptical* doubt. The problem here is that the skeptic misconstrues predications of sensation as hypotheses that are falsified by exactly one counterexample. However, to reason from the empirical fact that *sometimes* we cannot know *what* an individual is thinking to the proposition that we can *never* know what she's thinking and thereby conclude that we can never know *that* she's thinking – is a fallacy. If no relevant grounds for disbelief are apparent in the circumstances then her expression means, *ceteris paribus*, that she's in pain, even, contra Malcolm, in the absence of any other (behavioural) criteria. This, to paraphrase Wittgenstein, is simply what *we call being in pain*. Beyond the limits of convention lies the expression.

April's expression of pain provides us with criterial conviction (a kind of contextual 'unearned' certainty – not strictly opposed to belief – stronger than inductive probability yet weaker than entailment) to justifiably identify her expression as an expression of pain.⁵⁹ By virtue of this, her pain becomes the content of a belief that 'I have no grounds for doubting ... but, on the contrary all sorts of confirmation'.⁶⁰ Doubts may still be *logically possible* but we need justifiable reasons for specific acts of dubiety if criterially supported judgments (made on the basis of a priori belief) are to be threatened (cf. §84). What Baker has dubbed 'C-support' is therefore assumed by Wittgenstein's commentators to yield conclusive confirmation (but, significantly, *not* logical entailment)⁶¹ that she's in pain (just in case she is); as a result of which the onus, as Baker argues, is transferred to the skeptic who seemingly without justifiable reason may still insist that we can *never know* (i.e., possess complete and indubitable evidence

⁵⁹ Baker (1974: 162): criterial justification, as employed here, is simply the concept of citable reasons (X, Y) for something's (Z) being the case; even though the reasons in question remain context-sensitive, circumstantial and generally, but not irreducibly, conventional, they are normative. Malcolm and Hacker's account of criteria begs the question because it assumes that skepticism is wrong and that the concept of criteria refutes it (i.e., the consequent of their argument is accepted prior to the antecedent). But if a criterion were such that it established with certainty that Y , then X would be identical to Y and then, by Leibniz's law, $X = Y$ i.e., the criterion would be that which it is taken to be a criterion of.

⁶⁰ Wittgenstein (1969: §288). It is, in fact, disbelief that requires the justification in the case of pain-behaviour as a criterion of pain. Baker (1974:163).

⁶¹ Baker defends the contentious claim that C-support yields semantic (grounds for) certainty despite admitting that a proposition p (about public X) C-related to another proposition q (about private Y) can be 'undermined since it doesn't amount to [logical] entailment' (i.e., q can be false and p true) (1974: 162, 176).

for her experience, the way she, transparently, can). Thus one principal motivation of the appeal to criteria is that of recasting ‘certainty’ as a form of conviction beyond actual (relevant) disbelief (in situations, in other words, where there is no evident reason – or natural inclination – to doubt) and not what is beyond *all imaginable*, logical or methodological doubt. Criteria are effective only in quotidian contexts where no credible grounds for the suspension of belief are apparent, and where certainty cannot be characterised as co-extensive with the *logical impossibility of doubt* – and relative to which the unconstrained patterns of doubt characteristic of skepticism look necessarily acontextual even pathological, conspiracy-theorist, paranoid. Certainty, as emerges in Wittgenstein’s last writings, is not a kind of ultra-knowledge but something categorically different (something like the necessary natural inclination to suspend disbelief).

Yet a certain fact may come to light that alters everything. Imagine someone who moments before was ‘writhing’ in obvious pain, abruptly stopping, getting up, dusting off her clothes and fetching her coffee: from behind the crowd a voice yells “Cut!” Now, contrary to former evidence, I possess justification for disbelief; I think, nevertheless: that was a convincing performance phenomenally indistinguishable from the expression of (real) pain.⁶² Thus as supplementary relevant data come to light, I may be epistemically compelled to reverse my initial judgment (but can I be said to doubt *this* initial judgment?). ‘C-justification’, to use Baker’s terminology, is clearly defeasible⁶³: it remains possible, that is, that, *even* in the presence of all salient criteria, the person may not *actually be* suffering (the) pain (that appears to be expressed). The outside may contradict the inside.

Criteria are subject to defeat, Baker concedes, by evidence delivered from an expanded frame of reference (as the counterexample shows).⁶⁴ According to Cavell, because of the defeasibility constraint, Wittgensteinian criteria fail, ultimately, to provide the certainty that the prevailing interpretation (Malcolm, Albritton, et al.) tends,

⁶² As McDowell (2009) observes, the antic scenario plays a role in the problem of other minds ‘analogous to the role of the concept of illusion in the traditional approach to the epistemology of the “external world”’ (76).

⁶³ ‘If *p* is C-related to *q*, then it is possible that *p* is true and *q* false because of the defeasibility of the C-relation’ Baker (1974: 167).

⁶⁴ *Ibid.*: 162.

despite acknowledging their limitations, to attribute to them.⁶⁵ Rather, the appeal to public criteria in order to mitigate the epistemological problems of predicating psychological (private) concepts to others, merely ‘reveals’ what Cavell refers to as ‘the truth of skepticism,’ i.e., that methodological dubiety is precipitated by recognising the limitations of the epistemological project per se. Such admission of epistemic threshold represents a genuine attempt to respond adequately to the reality of contingency, to the way, in other words, that a restricted cognitive perspective leads to a profound sense of disconnection from others. Relative to this acceptance of cognitive limitation, problems beset the entire ambition to transcend the conditions of situated, perspectival knowledge in order to access the mind of another – i.e., the desire to gatecrash someone’s private subjectivity, to invade their interior, secret lives, which are, according to Cavell, ‘*exactly the problems the skeptic sees*’.⁶⁶ Skepticism is thereby reparsed as a conscientious idea born of authentic insight that mobilises its cognitive armoury in defence of the interdictions of mutual distance. Engendered, one could say, less from the desire to challenge uncritical dogmatism than from a respect of privacy, it is, basically, an acceptance of what Cavell refers to, in existential mode, as our shared human ‘finitude’.⁶⁷ This constitutes the *truth* of the skeptical attitude, embracing the fallible, finite conditions of *human* knowledge. Thus we must remain, according to Cavell, sufficiently ‘open’, in Wittgensteinian mode, to the skeptic *in ourselves*⁶⁸ if we are to provide an adequate counterchallenge to its cognitive ‘threat’.

Even Malcolm is obliged to recognise that it is possible to imagine a situation in which *all* criteria are satisfied and yet the person manifesting the relevant behaviour does *not actually* experience pain. ‘If we come upon a man exhibiting violent pain-

⁶⁵ At the beginning of *The Claim of Reason*, Cavell criticises Malcolm’s identification of pain-behaviour with “criteria for pain” (and, via the criteria, for *pain* itself). If ‘the criterion of being in pain is satisfied’ (this, however, is not a concept Wittgenstein employs) Malcolm argues ‘then he *must* be in pain’ (Malcolm 1954: 544). “Pain,” we could say, for Malcolm, implies *pain*. In Cavell’s reading, Malcolm interprets Wittgenstein as construing the ‘outward criterion’ as a means of confirming for certain the presence of the inner phenomenon. Necessarily: ‘The satisfaction of the criterion of y’ he quotes ‘establishes the existence of y beyond question’ (Malcolm, 1954: 544). Malcolm’s exegesis interests Cavell who reads the *Investigations*, like Kripke, as an extended engagement with (as opposed to a solution of) what he terms, revealingly, the ‘*threat* of skepticism.’ Cavell does not refer to Baker’s 1974 (i.e., five years prior to the *Claim of Reason*); he also ignores Albritton’s recanting postscript.

⁶⁶ Cavell (1996: 64). See also Wright (2002). There is something important to be learned from skepticism (that Moore, for instance, didn’t appreciate) i.e., that the limits of justification they express derive from a genuine insight: namely, that ‘cognitive achievement must be reckoned to take place within such limits.’ Otherwise complete epistemic ‘paralysis’ would result (37).

⁶⁷ Cavell (1966: 172).

⁶⁸ Cavell (1979: 47).

behaviour, couldn't something show that he is not in pain?' To which he replies: 'Of course.'⁶⁹ For, the person in question, as considered, may have been acting, malingering, rehearsing a part ... etc. Can I be certain, *even* given the presence of all relevant behavioural criteria in the *right circumstances* that this person *actually feels* the pain that he expresses? Evidently not. To illustrate, we could say that it would be absurd to insist that, because all the relevant behavioural criteria are (more than) satisfied, that John Hurt, for instance, was actually suffering pain in the famous scene in Ridley Scott's *Alien*. But, as a result of such examples, 'how can we ever know whether another person is actually *suffering* pain?'⁷⁰ The outward criterion fails to establish, by observation of physiological behaviour that the inner (psychological) sensation is *present* ... that it *actually exists*: thus, by virtue of these counterexamples, as well as the admission of defeasibility, it seems undeniable that I cannot be certain that the agent of pain-behaviour actually *feels* pain.

Yet this uncertainty was precisely what the appeal to criteria was intended to eliminate. It follows, for Cavell, that there are no necessary and sufficient public (behavioural) conditions for *knowing with certainty* that the mental state exists, that another person *really* is experiencing, phenomenologically, a private sensation (or is deliberately deploying the classic behavioural signs in an antic way to fake it). Relative to Cavell's analysis, McDowell's disjunctive insistence that, despite their paradoxical defeasible status, criteria merely *appear* to be satisfied in the antic scenario – that they 'are not really satisfied' – may seem naïve.⁷¹ For Cavell's position amounts to the conditional argument that if 'the knowledge is not really available' in the antic scenario then neither can it be said to be 'available' in the *real* incidence.⁷² Failure to support certainty, for Cavell, makes criteria very 'disappointing'.⁷³

Do we never succeed in genuinely understanding the psychological lives of others? For even if someone, for example, fulfils "all the criteria" marked out by "all parts of the grammar of pain" then, to be sure, it is exceedingly likely that he is in pain

⁶⁹ Malcolm (1954: 545); see also Cavell (1979: 41).

⁷⁰ Cavell (1979: 44).

⁷¹ But supported by §304 where Wittgenstein has his imaginary antagonist say: "But you will surely admit that there is a difference between pain-behaviour accompanied by pain and pain-behaviour without any pain?" He answers: 'Admit it? What greater difference could there be?'

⁷² McDowell (2009: 76). McDowell proposes 'an anti-skeptical thesis which has all the main features of neo-Mooreanism' (Pritchard 2006: 290).

⁷³ Ibid.: 79 (and *passim*).

i.e., that he's not feigning, etc.' the point is, however, that this likelihood connotes at best merely abductive probability: *we cannot be certain*.⁷⁴ Indeed, even if we have to depend on behavioural (or, indeed, neurological) data for criteria to verify the existence of mental states then, as Noë has observed, this is equivalent to admitting that knowledge of other minds is beyond our epistemic capacity.⁷⁵ And, if so, it seems we cannot credibly claim to *know* (because we can't provide the theoretical justification in refutation of the skeptical counterexample that would supply the sufficient and necessary conditions for epistemic confidence). Criteria perhaps provide 'good evidence',⁷⁶ but acceptance of their mere adequacy (whether necessary or possible) acquiescently presupposes that we are obliged to rely on inconclusive behaviour for pseudo-access (reconstruction from behavioural cues) to the psychological attitudes of others; so although criteria may be sufficient to facilitate a *skeptical* solution (to the effect that our concept of the consciousness of others can be regarded as a kind of working stratagem for predicting agency) they fail to deliver the 'grail' of knowledge (i.e., certainty).⁷⁷

'C-justification' thus remains vulnerable to the skeptical challenge because, ironically, its entire motivation can be exposed as determined by a tacit acceptance that certainty still conditions the only kind of knowledge that counts if the ingrained impulse to doubt is to be satisfied. All discussion of criteria is thus hampered by a kind of subliminal disenchantment engendered by this admission of failure to achieve the elusive objective of certainty. Cavell notes the repeated use of qualified phrases like 'near certainty' and 'almost certain' in the official Malcolm-Albritton accounts.⁷⁸ Conceding the defeasibility of the criterion (that criteria are context-sensitive, circumstantial and conventional) amounts to retaining the concept of mind (as an expedient theoretical construct) only at the expense of acquiescing to the radically contingent relationship between behaviour and intentionality.⁷⁹ The epistemic 'gap' between the presence of public criteria and their satisfaction by private psychological content, Cavell comments, ultimately cannot be bridged by appeal to criteria: the quasi-

⁷⁴ Cavell (1979: 44, 79).

⁷⁵ Noë (2010: 28).

⁷⁶ Cavell (1979: 79).

⁷⁷ The phrase 'grail' of knowledge comes from Kołakowski (1988: 71).

⁷⁸ Cavell (1979: 39).

⁷⁹ Noë (2010: 30).

certainty provided by C-justification is preserved only at the cost of never knowing for sure ‘that the criterion is satisfied, that what it is *of* is *there*’.⁸⁰ Thus the connection between criterion and what it is a criterion *of* seems undermined by the defeasibility constraint. ‘The wince itself is one thing’, Cavell concludes, ‘the pain itself [remains] something else; the one can’t *be* the other’.⁸¹

But does criterial defeasibility provide evidence of the ‘truth’ of skepticism? I don’t believe so – but *not* because criteria fail to provide the anti-skeptical knowledge that Malcolm, Baker, and Hacker assume they do. What it does demonstrate, perhaps, is the irrelevance of skeptical dubiety to quotidian epistemic conditions (and conventions) that pertain to contexts where there is no natural inclination to doubt. (The relevant certainty here is that which, necessarily, ignores the myriad counterfactual alternatives, possible or impossible, that could conceivably be the case.)

Admittedly, we say *X* is a criterion of *Y* thus confirming that *X* is different (roughly: in ontological status) to *Y*. Necessarily, however, (i.e., by Leibniz’s law) no criterion can *be* that which it is a criterion of any more than a representation can be identical to what it represents. Yet *X* *still* remains a criterion of *Y* even if, in specific circumstances, *Y* is discovered not to be the case. And this is the key to the concept: criteria are always criteria *of* something. The pain-criterion, that is, represents its content (the concept of pain) and that is how it comes to *mean* pain to us: “Pain” represents PAIN which, in turn, represents *pain* (just as the simulated concept CAT in the skeptical scenario [BIV] still represents *cats*). Where we are referred *through* the representation (sense) to focus on *what* it represents (referent), the behaviourist mistakenly fixates on the representation itself, confusing it with what it represents (misidentifying sense as reference); the skeptic repeats the same error but, while focusing on the representation, *denies* that it represents anything (in both cases the criterion *qua* representation is treated as opaque and / or self-referential). By concentrating on the ‘subsidiary’ *representation* rather than what it ‘focally’ represents, both behaviourism and skepticism misconstrue the logic of representation (it is treated with paranoid suspicion by the skeptic and too much credibility by the behaviourist).⁸²

⁸⁰ Cavell (1979: 41).

⁸¹ Ibid.: 79.

⁸² Such a switch in perspective leads to the ‘aesthetic fallacy’, that is, the inference that the referent is reducible to its representation. Cashell (2009). These are two aspects of the structure of tacit knowing

Another way of putting this is that both reduce the epitome of the human *being*, as McDowell remarks, to the cipher of the human *body*.⁸³

By virtue of the way Cavell and (it could be argued) Wittgenstein develop the concept, criterion is synonymous with *representation*; so that, for instance, in the antic scenario, Cavell says that ‘*what* he is feigning must be precisely *pain*, what he is rehearsing must be the part of a man *in pain*, the hoax depends on his simulating *pain*, etc.’⁸⁴ Pretending to be in pain necessarily engages conventional pain-behaviour to convincingly imitate pain: that is to say, a person’s performance, to be convincing in this regard, must try to fulfil what we (the community) employ to establish, in the typical case, that someone is suffering pain; they may then be said to ‘satisfy’ the criteria *of* pain: they know that we know how someone is likely to behave when suffering pain and they reason, “if I behave like *this and this*, then it will appear to observers as if I’m really suffering pain.” It is only because it has the property of representing content that it is possible to ‘*retain the concept*’ of pain and flexibly apply it to heterogeneous cases. In other words, it is only because it represents (pain) that the criterion (of pain) can determine the correct application of the concept (of pain) to other, future, instances in the appropriate circumstances. If new information becomes a criterion for her *not* being in pain, then the criterion itself is clearly *not* invalidated by circumstantial change. Criteria provide reasons for reliable judgement: that’s all (but that’s enough i.e., it’s sufficient for making cognitive decisions that are beyond reasonable doubt).

Even in the antic scenario, the agonist satisfies ‘the criteria we use for applying the concept of pain to others. [But] It is because of *that* satisfaction that we know that he is feigning pain (i.e., that it is pain he is feigning), and that he knows what to do to feign pain’.⁸⁵ While many regard criteria as providing, at most, context-sensitive, circumstantial epistemic support (and this is accurate as far as it goes) analysis of the antic scenario demonstrates (contra the skeptical assumptions) that criteria do *not change* with context. For when new evidence is brought to light through context-expansion (think of a legal inquiry or on-going police investigation) such evidence

which determines the foundation of all knowledge achievements according to Michael Polanyi (1969: 182; 1967).

⁸³ McDowell (2009: 78).

⁸⁴ Cavell (1979: 45).

⁸⁵ Ibid.

invites the application of *different* criteria (having *different* content), criteria of something else that justify the judgment that something else is the case; we are (as above) compelled to revisit and alter (if needs be) our initial judgment on the basis that *new criteria* suggest that something *other* is, precisely, now the case. If criteria cannot confirm the necessary *existence* of something they can nevertheless ‘determine the [accurate] application’ of the concept in the appropriate circumstances.⁸⁶ Thus, for the sake of the provision of knowledge, criteria are not error-immune, that is, they do not provide ultra-knowledge exempt from the logical impossibility of doubt. So despite valid counterexamples, an expression of pain, regardless of the lack of supporting evidence, *still remains* a criterion – that is *evidence enough*, *ceteris paribus* – of pain. Although John Hurt may deploy representations that presuppose criteria for the *real* expression of *real* pain for their efficacy, different criteria, quite simply, apply in the antic scenario; criteria that are satisfied by a relevantly different state of affairs being the case (they are the criteria for *playacting*).

Criteria are *not relativised* to context. (And thus McDowell’s disjunctive interpretation turns out, in fact, to be the more accurate one). In imagining the antic scenarios, just as in cases of hiding ‘private’ sensations or beliefs, it is, Wittgenstein remarks, ‘important that I have to imagine an artful concealment’ (§391). We would be profoundly shocked if it were revealed that the actor was actually suffering *that* pain.

It is only because sensations are represented that we can represent sensation. The signs of pain, like all signs, can be imitated. Thus it is otiose to complain, pace Cavell, that the signs are *not* the pain itself; it makes more sense to observe that because the signs of pain are *about* pain, they refer to the concept PAIN, a reading supported by the cryptic aphorism, ‘An “inner process” stands in need of outward criteria’ (§580). Sensation, that is to say, must be represented in order to be communicated. How is a sensation represented? When pain is expressed in (physical, observable) signs, for instance, the body is the vehicle, the epitome of its representation. Predications of sensation, contra Baker (and Chihara & Fodor), are therefore *neither* statements about behaviour *nor* about states of mind but rather *both*, ‘not side-by-side, however, but

⁸⁶ Ibid.

about the one *via* the other' (p. 179).⁸⁷ At this junction, the body becomes a *living* sign through which (*durch das andere*) consciousness, interiority, sentience – is *represented*. And it is here that the infamous Wittgensteinian aphorism comes into its own: 'only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious' (§281). The living human is the epitome 'to which third-person attributions of consciousness, sensations, feelings, are related'.⁸⁸ As my verbal expression "I am in pain" draws attention to *myself* (§405) so I 'look to him for information about his thoughts and intentions,' Malcolm confirms, 'He is our primary source of information about himself'.⁸⁹

The epistemology of other minds is recast as the consequence of an entrenched conviction that the only knowledge that counts as adequate in this instance is certainty (i.e., omniscience); but, as William Poteat has observed, we don't even know *ourselves* in this way.⁹⁰ According to Cavell, skepticism does not succeed in developing its insights, however, because what it seeks is more than what any philosophical genre is capable of delivering; for the only thing likely to inoculate against the viral doubt of skepticism is a kind of *über*-Cartesian certainty, the kind of omniscient point of view that would immediately invalidate the very possibility of doubt (both reasonable *and* unreasonable). But this, as demonstrated by Cavell, is simply not achievable: the gap will never be bridged if approached from this vantage point; all it will achieve is to reinforce at a metaphysical stratum the existential condition of alienation recognised at the ordinary stratum. Thus skepticism ultimately ends up in its own dogmatic *cul de sac* with an (albeit aporetic) endorsement of the regnant disjunctions of post-Cartesian metaphysics – behaviour *or* consciousness, body *or* mind, object *or* subject. But our relation to (and relationships with) others are not primarily epistemic in nature: they are *not* relationships of 'knowing, where knowing construes itself as being certain. So it is also true that we do not *fail* to know' in this context.⁹¹ The moral of skepticism (and the reason why Cavell argues that it is important to take its impulse seriously) is that our

⁸⁷ This way of thinking about criteria strikes me as mistaken: it is wrong to construe criterial support as a semantic relation between a proposition about behaviour and a proposition about a mental state. There is no 'suppression of a middle term' in linking behaviour to mental state.

⁸⁸ Malcolm (1954: 548).

⁸⁹ Malcolm (1977: 131).

⁹⁰ Poteat (1994: 215).

⁹¹ Cavell (1979: 45).

relationship with the world and other people is, at best, derivatively cognitive (the essence of such relationships cannot be reduced to evidential justification for the folk-psychological belief in putative entities such as “mind-independent objects” or “the minds of others”). Skepticism misconstrues the intersubjective relationship as epistemic in structure (explicable only in relation to the failure to provide justification to support ungrounded inveterate beliefs). Intended to ‘find the other’, therefore, skepticism ultimately ‘closes’ the other ‘out.’⁹² In refusing to treat the body ‘as *expressive* of mind’ the skeptic, in Cavell’s arresting metaphor, ‘scoops mind out of it’.⁹³ Even if it is possible (at a stretch) to imagine a committed, systematic skeptic who *always* doubts whether another person suffers pain – even to the point of *surrealism* (to use Cavell’s own bromide),⁹⁴ such doubts, Cavell admits, often ‘seem to make good sense only on the basis of ideas of behaviour and of sentience that are invented and sustained by skepticism itself’.⁹⁵ It is important that, in imagining the hard-core skeptic, we think of someone cold and cruel, incapable of empathy or conditioned to suppress the normal response to suffering, someone like *A Clockwork Orange*’s Alex, pathologically indifferent to the pain of others. Though strictly speaking not illogical, the ‘abnormal reaction’ as Malcolm correctly insists ‘*must* be the exception and not the rule’.⁹⁶ Regarding the pain of others, it is de facto important to ‘shut our eyes’ in the face of de jure doubt (p.224). It is humanely, ethically important that, contra Cavell, we are clear that it is *doubts* and *not criteria* that come to an end.⁹⁷

The Ethics of Skepticism

Although it may be to dwell on its most troubling aspect, it is necessary at this late point in the discussion to inquire what putting the consciousness of others into question implies from a moral standpoint. This concern, frequently elided in discussion, becomes apparent with the realisation that ‘other minds’ skepticism, unlike its external world variety, pertains in a non-trivial way to *other people*. Recent research in this area

⁹² Cavell (1979: 84).

⁹³ Ibid.: 67 (my emphasis).

⁹⁴ In an earlier essay, Cavell had used the descriptor ‘surrealism’ to characterise the more bizarre arguments of philosophical discourse. Such statements of the surreal were presented as paradigmatic of the consequences of banishing words from their natural habitat. Cavell (1964: 89-90).

⁹⁵ Cavell (1979: 47). The language-game of (Cartesian) doubt, as Wittgenstein observes, ‘presupposes certainty’ Wittgenstein (1969: §115).

⁹⁶ Malcolm (1954: 547).

⁹⁷ Cavell (1979: 412); Wittgenstein (p. 181).

has significantly begun to recognise that the problem is not disinterestedly epistemological⁹⁸ in character, in relation to which, I want to contextualise Cavell's exegesis of Wittgenstein.

Although characterised (in my view) by an overreaction to the 'threat' of skepticism (and fascination with its alleged 'truth'), Cavell nevertheless develops an insightful attempt to redefine the problem from an ethical perspective while conceding its irrefutability according to the principles of classical epistemology; he manages, that is, to recast the epistemology of other minds in a way that shifts the focus of the skeptical challenge by compelling us (and even by virtue of the antecedent concession) to recognise the implicit (and non-benign) moral implications of withholding the concept of consciousness from others. Although it has been observed that the potentially harmful ('pernicious') effects of skepticism imply the normative injunction that 'We should not take such an attitude,'⁹⁹ skepticism regarding other minds may seem intuitive because irrefutable (the Cavellian truth is that it may be impossible to justify our tendency to ascribe consciousness to other people in epistemological terms). Yet, from an ethical standpoint, this still does not morally justify adopting the skeptical attitude.¹⁰⁰ From the stronger intuitive realisation that we have reason not to suspend belief in the minds of others, the general inference may be drawn that epistemic commitments have significant ethical consequences (which would warrant a conclusion that the ethical is supervenient on the epistemological) but this argument is not required (at least not entirely) to defend the normative justification for the existence of other minds proposed here. Indeed, what I propose implies that the epistemic justification sought for other minds becomes the problem of defending why it may appear intuitive to suspend belief. What reason or purpose would require me to deny body, face or speech as expressive of psychological attitudes – to what end? I would argue (more extensively if space permitted) that there is no reason to motivate the denial of subjective agency to another person that does not supervene on an antecedent ethical presupposition to the contrary. Yet the 'attitude' associated with the skeptical suspension is equivalent to the refusal to concede that the concept of consciousness applies to anyone else (and not just to the admission that it is possible that another

⁹⁸ Overgaard (2007), Hirsch (2011), Pritchard (2005, 2008), Williamson (2011).

⁹⁹ Williamson (2011: 8).

¹⁰⁰ See also Cavell, (2005: 210).

person does not have phenomenological experience of psychological states). In other words, it is defensible to construe other minds skepticism as an active decision (to regard the behaviour of others as *mere* behaviour, that is, to insist that human behaviour is an opaque presentation or stylised mime) rather than an innocuous armchair indulgence.

In practice, skepticism (that is, if the concept of “practical skepticism” is not an oxymoron) connotes an evaluative, axiological attitude that discriminates on the basis of some transcendental (and obvious) property that it is stipulated others lack, namely, the capacity for subjective life (and all that this implies of agency, freedom of thought, emotion, intentionality). On the basis of impeccably (and tortuously) argued inferences that appear to endorse the irrefutability of skepticism, it may seem reasonable to decide that others, because it cannot be established conceptually that they share our cognitive capacities, are not deserving of respect. Indeed, that professional philosophers have failed to disprove the skeptical hypotheses (or failed to establish that they are irrational) may be taken, erroneously, to provide reasonable corollary evidence of their truth. And as a consequence, we may come to believe, for instance, that our attitude to others ultimately doesn’t matter – that the effort to understand another’s point of view is academic or naïve, or that, because we cannot always be certain if a person actually suffers pain, that compassion (or care) is sentimental.

Skepticism is always presented as a position of epistemic superiority; yet I would claim that denying consciousness to others, when ‘everything speaks for and nothing against it,’¹⁰¹ actually represents a kind of cognitive perversion. It is therefore, I would argue, not reasonable to reject the necessary truth that others are subjects of conscious experience; it may not be an exaggeration, indeed, to suggest that the denial of mind (especially when the culture values it so inordinately) constitutes a kind of intellectual violence. Is this not perhaps the paradigm of irrationality? (Although, in practice, it should suggest that something is seriously wrong with the reasoning.)

One of Cavell’s most original contributions to the debate is his discussion of the other (typically neglected) alternative to first-person propositions of sensation, namely, second-person epistemic statements of the form: “I know *you* are *X* ” the correct analysis of which is as expressions of *sympathy*: Cavell treats such propositions as

¹⁰¹ Wittgenstein (1969: §117).

equivalent to “I know what you’re going through,” or “I’ve done all I can” or even, perhaps, “I’m here for you.” The expression of pain, when our vulnerability is also crucially exposed, precipitates a reciprocal response that Cavell captures with the category of ‘acknowledgment.’ The latter is elsewhere thematised as a response to the ‘call’ of the other: an interpellation which stakes a claim by invoking in me an imperative for ‘comforting, succoring, healing’.¹⁰² This, if anything, constitutes the only significant form of ‘knowing what pain is.’ To *know* pain in this way, in other words, is to acknowledge the person who *expresses* it, that is, to respond (emotionally) to her—to reply to her appeal.

“I know *you* are in pain” cannot be reduced to an epistemic statement because what it communicates is recognition (of your pain): ‘an expression of *sympathy*’, a state equal and opposite to your state, implying an emotional reaction determined by beholding your presence. This analysis allows us to differentiate between the concept of epistemic access to the psychological (or intentional) states of others (even if possible, always vulnerable to the ‘threat’ of skepticism) and emotional responses of sympathy evoked by another person’s presence (and, by extension, any being capable of expressing pain, therefore feeling pain, hence suffering). We respond to another person by *acknowledging* their presence in a way that a priori accepts their otherness (transcendence) without necessarily seeking to comprehend it.

It is not enough that I *know* (am certain) that you suffer – I must do or reveal something (whatever can be done). In a word I must *acknowledge* it, otherwise I don’t know what “(your or his) being in pain” means. Is.¹⁰³

In ‘its requirement that I *do* something or reveal something’, acknowledgement, Cavell says, actually ‘goes beyond knowledge’ and toward *agency*.¹⁰⁴ We may respond to expressions of pain with reactions of sympathy (or, indeed, as Cavell points out, with indifference) but we cannot but respond actively in some way that, paradoxically, transcends the classical epistemological categories. When someone expresses something, tells or reveals something to me, I am obliged to act in a way that ipso facto expresses my attitude (if, for instance, I *refuse* to act, this also expresses a very clear

¹⁰² Cavell (1979: 81).

¹⁰³ Cavell (1996: 68)

¹⁰⁴ Ibid.: 63.

attitude). Accordingly, my empathetic reaction (positive or negative) vividly reveals that I cannot but regard him or her as a conscious being; and it takes a good deal of concentrated (we could even say *perverse*) effort to doubt it.

So the category of acknowledgement importantly challenges the skeptic's official cognitive disinterestedness and, at the same time, helps to disclose the ethical structure of the interpersonal relationship. Cavell notes the fundamental distinction between the 'failure to know' and the 'failure to acknowledge' for, evidently, the latter, unlike the former, is liable to be evaluated as a lack of empathy: a kind of 'indifference ... callousness, an exhaustion, a coldness'.¹⁰⁵ Moral vacuity, he reminds us, can never be considered ethically neutral and, as a community informed by normative means of expression, we should not take it to be. Suffering matters: and 'your suffering makes a *claim* upon me'.¹⁰⁶ Even if – obviously – I may not immediately feel your pain as a tangible episode in my body, I may, nevertheless, be compelled to acknowledge it through my involuntary emotional reaction to you. For when *you* say "I'm in pain" this elicits a very strong emotional response – even though I may not, by contingent physiological facticity, physically *feel* the sensory intensity of your pain: but the accompanying distress expressed, affectively communicated and thus literally sharable by me, may be more important in this context than the fact of being unable to feel your sensation (this is not only meaningless, incidentally, but also strictly irrelevant to your suffering). It should be pointed out however that neuroscientific research has experimentally established that the affective dimension of pain, that to which I respond empathetically and which constitutes what, following Wittgenstein, we may call the 'physiognomy' of pain, is felt in an entirely non-metaphorical, vicarious way by observers of others in pain.¹⁰⁷ Not only this, but observing someone in pain (even their facial expressions) has been found to engage neural mechanisms in observers that duplicate the sensory *and* affective neurological dimensions of the phenomenology of pain. Thus the aphorism that concludes "Knowing and Acknowledging" anticipates the most recent findings of the neurological study of nociceptive empathy: 'I know your

¹⁰⁵ Ibid.: 69.

¹⁰⁶ Ibid.: 68.

¹⁰⁷ See Osborne & Derbyshire (2010) for bibliography of contemporary research in this area.

pain the way you do' (implying that, reciprocally, through acknowledging it, you know my pain the way I do).¹⁰⁸

Discussion of criteria tends to omit a crucial dimension of the discourse of pain: namely, sensitive response to the pain of others. Vicarious acknowledgement of another person's pain, like our empathetic response to the phenomenon in general, to use Wittgenstein's word, is '*ursprünglich*' – that is, again, visceral, impulsive, instinctual, above all, *emotional*. Yet pain is a purely physical phenomenon; a fact of physiology associated with the neurological structure of the sensitive beings we are (§281). We react instinctively adversely to its alarming nature, to its semiotics of harm. Yet it is also an irreducibly mental phenomenon toward which we develop codified cognitive (therapeutic) attitudes: instead of expressing pain through inarticulate ejaculations, we acquire normative grammatical modes of representing it (that, as argued, actually serve to suppress the more impulsive aversive reactions).¹⁰⁹ My reaction to another person in pain, therefore, may be as instinctual, visceral, and immune to error as my own experience of pain: 'I can be certain of someone else's sensations as of any fact'. "But if you are *certain*, isn't it that you are shutting your eyes in the face of doubt?" – They are shut' (§224). Eyes shut, perhaps, but emotionally open. In paragraphs §§ 540-545 of the *Zettel*, Wittgenstein confirms that responses of concern for others in pain are as *ursprünglich* as my pain in the same circumstances: 'being sure that someone is in pain, doubting whether he is, and so on, are so many natural, instinctive, kinds of relationship towards other human beings, and our language is merely an auxiliary to, and further extension of, this behaviour'.¹¹⁰ These remarks clarify that the expressive view of the discourse of pain is what primarily (but not exclusively) motivates the ethical approach to the problem of other minds suggesting the identification of new normative criteria that may be taken to refute the skeptical hypotheses in practice if not in theory (but this is *precisely* where it matters).¹¹¹ It is my instinctive *attitude* (toward others), paradoxically, that confirms that they, in fact, suffer: this attitude (of empathy and the

¹⁰⁸ See especially Avenanti et al. (2005); Botvinick et al. (2005); Fan et al. (2008); Jackson et al. (2005; 2006a; 2006b). Osborne & Derbyshire (2010) have found that the *physical sensation* of pain (and not just the affective dimension) is evoked in observers of others in pain.

¹⁰⁹ It is important to remind ourselves that acted pain tends to be more exaggerated than expressions of pain in ordinary real-life circumstances.

¹¹⁰ Wittgenstein (2007: §545). See also Loggia et al. (2008).

¹¹¹ Note that it does not follow from this that the ethical position defended here is itself expressivist in orientation.

capacity to identify with), to paraphrase Wittgenstein, ‘is a form of conviction that someone else is in pain’ (§287). It is this attitude (the ethos of compassion) that, above all, provides reason for the ethical commitment to the consciousness of others.¹¹²

It is becoming increasingly clear that the “theory of mind” still current in contemporary cognitive science needs to be abandoned if the issue of the consciousness of others is to be realistically addressed. Jettisoning the entrenched picture of the mind as an ‘unobservable domain ... whose effects are evident in what people say and do’¹¹³ is required for the acknowledgment that our ‘commitment’ to the ‘alive consciousness’ of others cannot be justified because it is simply not the kind of commitment amenable to justification according to conceptual criteria of epistemological analysis. Our commitments in this regard, as Wittgenstein argues in *On Certainty*, are, contrary to the whole approach associated with the post-Cartesian epistemological paradigm, unjustifiable in essence; nevertheless we are entitled to (believe in) them despite their lack of supporting ‘evidence.’ Of course this implies, according to Pritchard, that ‘the propositions that we hold to be most certain are *of their nature*, lacking in evidential support’.¹¹⁴ Even if warranted justification cannot be earned for our commitment to the consciousness of other people (or the external world), and this distresses us, ‘the fact remains that I really believe it’.¹¹⁵ Because the skeptical suspension of belief in the consciousness of others has non-trivial moral consequences, however, I have argued that justification *should* not be demanded for this commitment. Ultimately, therefore, Wittgenstein’s dissolution of the problem of other minds is normative in three intimately related senses: his emphasis on the grammar of sensation (that is, on the norms of linguistic behaviour) and on the criteria (i.e., the norms) for the ascription of psychological concepts to others also suggests that we do not doubt that other people are subjects of consciousness because in an important sense we *ought* not to.¹¹⁶

¹¹² Loggia et al. (2008).

¹¹³ Noë (2010: 29).

¹¹⁴ Pritchard (2005: 198).

¹¹⁵ Hirsch (2011: 19).

¹¹⁶ See also *ibid.*: 195. The link made here, incidentally, is also tacitly developed in Wright (2002): Not accepting unwarranted (yet transcendental) presuppositions, he says, ‘is to avoid having a life’; for all ‘rational thought and agency involves ineliminable elements of cognitive risk’ (41). Hirsch and Wright both refer to intellectual or epistemic ‘responsibility’ in this context.

Kieran Cashell

Limerick Institute of Technology

kieran_cashell@hotmail.com

References

- Albritton, R. (1966) 'On Wittgenstein's Use of the Term "Criterion"' IN G. Pitcher (ed.), *Wittgenstein: The Philosophical Investigations*, New York: Doubleday, 231-250.
- Apkarian, A. V. et al. (2001) 'Prefrontal Cortical Hyperactivity in Patients with Sympathetically Mediated Pain', *Journal of Neuroscience Letters*, 311, 193-197.
- Armstrong, D. (1963) 'Is Introspective Knowledge Incorrigible?', *The Philosophical Review*, 72, 417-432.
- Avenanti, A. et al. (2005) 'Transcranial Magnetic Stimulation Highlights the Sensorimotor side of Empathy for Pain', *Natural Neuroscience*, 8, 995-960.
- Ayer, A. J. (1953) 'One's Knowledge of Other Minds', *Theoria*, 19, 1-20.
- _____. (1954) *Philosophical Essays*, London: Macmillan.
- Baker, G. (1974) 'Criteria: A New Foundation for Semantics', *Ratio*, 16, 156-189.
- Banville, J. (1982) *The Newton Letter: An Interlude*, London: Minvera.
- Blackburn, S. (2008) 'Criteria', *Oxford Dictionary of Philosophy*, Oxford: OUP.
- Botvinick, M. et al. (2005) 'Viewing Facial Expressions of Pain Engages Cortical Areas Involved in the Direct Experience of Pain', *Neuroimage*, 25, 312-319.
- Cavell, S. (1966) 'The Availability of Wittgenstein's Later Philosophy' IN G. Pitcher (ed.), *Wittgenstein: The Philosophical Investigations*, New York: Doubleday, 151-185.
- _____. (1979) *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy*, Oxford: Oxford University Press.
- _____. (1996) 'Knowing and Acknowledging' IN S. Mulhall (ed.), *The Cavell Reader*, Oxford: Blackwell, 46-71.
- _____. (1964) 'Must We Mean What We Say?' IN V. C. Chappell (ed.), *Ordinary Language: Essays in Philosophical Method*, New Jersey: Prentice Hall, 75-112.
- _____. (2005) *Philosophy the Day After Tomorrow*, Cambridge MA & London: Belknap Press of Harvard University Press.
- Cashell, K. (2009) 'Reality, Representation, and the Aesthetic Fallacy: Critical Realism and the Philosophy of C. S. Peirce', *Journal of Critical Realism*, 8 (2), 135-171.
- Chihara, C. S. & J. Fodor (1991) 'Operationalism and Ordinary Language: A Critique of Wittgenstein' IN D. Rosenthal (ed.), *The Nature of Mind*, Oxford: OUP, 137-150.
- Critchley, S. (2007) *Infinitely Demanding: Ethics of Commitment, Politics of Resistance*, London: Verso.
- Evans, G. (1982) *The Varieties of Reference*, Oxford: OUP.

- Fan, Y. et al. (2008) 'Temporal Dynamic of Neural Mechanisms Involved in empathy for Pain: An Event-Related Brain Potential Study', *Neuropsychologia*, 46, 160-173.
- Garver, N. (1996) 'Philosophy as Grammar' IN H. Sluga & D. G. Stern (eds), *The Cambridge Companion to Wittgenstein*, Cambridge: University Press, 139-170.
- Hacker, P. M. S. (1972) *Insight and Illusion*, Oxford: OUP.
- _____. (1997) *Wittgenstein on Human Nature*, London: Phoenix.
- Hampshire, S. (1952) 'The Analogy of Feeling', *Mind*, 61, 1-12.
- Hirsch, E. (2011) 'Is it Possible for me to Doubt that my Children Exist?' (unpublished).
- Jackson, P.L. et al. (2006a) 'Empathy Examined through the Neural Mechanisms Involved in Imagining How I Feel Versus How you Feel Pain', *Neuropsychologia*, 44, 752-761.
- _____. (2005) 'How do we perceive the Pain of Others? Window into the Neural Processes of Empathy', *Neuroimage*, 24, 771-779.
- _____. (2006b) 'To What Extent do we Share the Pain of Others? Insight from the Neural Basis of Empathy', *Pain*, 125, 5-9.
- Kirk, G.S. and J.E. Raven (1964) *The Presocratic Philosophers: A Critical History with a Selection of Texts*. London: Cambridge University Press.
- Kołakowski, L. (1988) *Metaphysical Horror*, London: Penguin.
- Kripke, S. (1982) *Wittgenstein on Rules and Private Language*, Malden MA & Oxford: Blackwell.
- Loggia, M. L. (2008) 'Empathy Hurts: Compassion for another Increases both Sensory and Affective Components of Pain', *Pain*, 136, 168-176.
- Malcolm, N. (1958) 'Knowledge of Other Minds', *The Journal of Philosophy*, 55 (6), 969-978.
- _____. (1964) 'Moore and Ordinary Language' IN V. C. Chappell (ed.), *Ordinary Language: Essays in Philosophical Method*, New Jersey: Prentice Hall, 5-23.
- _____. (1977) 'Moore and Wittgenstein on the Sense of "I Know"', *Thought and Knowledge*, Ithaca & London: Cornell University Press, 170-198.
- _____. (1977) 'The Privacy of Experience', *Thought and Knowledge*, Ithaca & London: Cornell University Press, 104-132.
- _____. (1954) 'Wittgenstein's *Philosophical Investigations*', *The Philosophical Review*, 63 (4), 530-559.
- _____. (1995) *Wittgensteinian Themes: Essays 1978-1989*, Ithaca & London: Cornell University Press.
- Malcolm, N. & G. H. von Wright (2001) *Ludwig Wittgenstein: A Memoir*, Oxford: Clarendon Press.
- McDowell, J. (2009) 'Criteria, Defeasibility, and Knowledge' IN A. Byrne & H. Logue (eds), *Disjunctivism: Contemporary Readings*, Cambridge MA: MIT, 76-90.
- Mill, J. S. (1865) *An Examination of Sir William Hamilton's Philosophy*, Boston: Spenser.
- Moore, G. E. (1998) 'A Defense of Common Sense' IN L. P. Pojman (ed.), *Classics of Philosophy*, New York, Oxford: OUP, 1140-1146.
- _____. (1966) 'Proof of an External World' IN J. N. Findlay (ed.), *Studies in Philosophy* London: Oxford University Press, 65-92.
- Nielsen, H. A. (1958) 'Wittgenstein on Language', *Philosophical Studies*, 8, 115-121.

- Noë, A. (2010) *Out of Our Heads: Why You Are Not Your Brain and Other Lessons from the Biology of Consciousness*, New York: Hill & Wang.
- Osborne, J. & Derbyshire S.W.G. (2010) 'Pain Sensation Invoked by Observing Injury in Others', *Pain*, 148 (2), 268-274.
- Overgaard, S. (2007) *Wittgenstein and Other Minds: Rethinking Subjectivity and Intersubjectivity with Wittgenstein, Levinas, and Husserl*, New York: Routledge.
- Plath, S. (1963) *The Bell Jar*, London: Faber.
- Poteat, W. (1994) *Recovering the Ground: Critical Exercises in Recollection*, New York: State University Press.
- Priest, S. (1991) *Theories of the Mind*, London: Penguin.
- Pritchard, D. (2006) 'McDowellian Neo-Mooreanism' IN F. Macpherson & A. Haddock (eds), *Disjunctivism: Perception, Action, Knowledge*, Oxford: OUP.
- _____. (2005) 'Wittgenstein's *On Certainty* and Contemporary Anti-Skepticism', IN D. Moyal-Sharrock & W. H. Brenner (eds), *Readings of Wittgenstein's On Certainty*, Basingstoke & New York: Palgrave, 189-224.
- Polanyi, M. (1969) 'Sense-Giving and Sense-Reading' IN M. Grene (ed.) *Knowing and Being*, Chicago: University of Chicago Press, 181-207.
- _____. (1967) *The Tacit Dimension*, Garden City, NY: Doubleday.
- Poteat, W. (1994) *Recovering the Ground: Critical Exercises in Recollection*, Albany: State University of New York Press.
- Price, H. H. (1938) 'Our Evidence for the Existence of Other Minds', *Philosophy*, 13, 425-456.
- Putnam, H. (1968) 'Brains and Behaviour' IN D. Rosenthal (ed.), *The Nature of Mind*, Oxford: OUP, 151-159.
- Russell, B. (1948) *Human Knowledge: Its Scope and Limits*, London: Simon & Schuster.
- Tanesini, A. (2001) 'In Search of Community: Mouffe, Wittgenstein and Cavell', *Radical Philosophy*, 110, 12-19.
- Williamson, T. (2011) 'Philosophical Expertise and the Burden of Proof' (forthcoming *Metaphilosophy*): www.philosophy.ox.ac.uk/__data/assets/pdf_file/0013/19300/metaphilosophy.pdf
- Wittgenstein, L. (1958) *The Blue and Brown Books*, New York: Harper and Row.
- _____. (1989) *Lectures on the Foundations of Mathematics, Cambridge 1939*, C. Diamond (ed.), Chicago: University of Chicago Press.
- _____. (1969) *On Certainty* (trans. D. Paul & G. E. M. Anscombe), Oxford: Blackwell.
- _____. (1974) *Philosophical Grammar* (trans. A. J. P. Kenny), Oxford: Blackwell.
- _____. (1953) *Philosophical Investigations* (trans. G. E. M. Anscombe), Oxford: Blackwell.
- _____. (1922) *Tractatus Logico-Philosophicus* (trans. C. K. Ogden), London: Routledge.
- _____. (2007) *Zettel* G.E. M. Anscombe & G. H. Von Wright (eds), Berkeley: University of California Press.
- Wright, C. (1982) 'Anti-realist Semantics: The Role of *Criteria*', *Royal Institute of Philosophy Lecture Series*, 13, 225-248.
- _____. (1984) 'Second Thoughts About *Criteria*', *Synthese*, 58, 383-45.
- _____. (2002) 'Wittgensteinian Certainties': www.st-andrews.ac.uk/~arche/papers/Wittgensteinian%20Certainties.pdf 1-44.

MULTIPLE REALIZABILITY AND *NOVEL* CAUSAL POWERS

Ricardo Restrepo

Abstract

Framed within the dialectic of the causal exclusion argument (Kim 2005), this paper does two things. One, it clarifies some properties of multiple realizability based on its true origin (Turing 1950). And two, it challenges a form of argument Noordhof (1997), Clarke (1999), and Whittle (2007) employ to support the idea that the mental has causal powers not had by its physical realization base (*Novel*). The paper challenges *Novel* with ideas derived from multiple realizability, among others.

1. Introduction

Does the mental have novel causal powers its physical realization base does not have? Taking the lead from Noordhof (1997) and Clarke (1999), Whittle (2007) employs an argument which concludes that the mental does have such novel powers (*Novel*). The conclusion is exciting and has received relatively little critical attention. I will argue, however, that the grounds for *Novel* are unpersuasive. Building on Turing's remarks, an analysis of multiple realizability reveals a preferable alternative explanation for the cases Whittle, Noordhof, and Clarke use to support the novelty of the causal powers of the mental. *Novel* arises in the context of providing a solution to the problem of causal exclusion originally developed by Jaegwon Kim. It is a "problem" in that it aims to display an inconsistency in certain non-reductionist assumptions, and then argues by *reductio ad absurdum* that non-reductionism is not true. The challenge for the non-reductionist is to provide a solution that rescues the position's core commitments. The problem of causal exclusion says that since every effect has a sufficient physical cause, and there is no systematic overdetermination, mental properties must either be reduced to physical properties in order to be causal or be ontologically eliminated, on pain of inconsistency (Kim 2005). Whittle considers *Novel* to be an obstacle to her own proposed solution to the problem of causal exclusion. However, since *Novel* is ineffective, she may see this as an upside in the problem of causal exclusion dialectic, even though I am not a proponent of her solution, Noordhof's, or Clarke's.

2. Multiple Realizability

Multiple realizability is an important element in the contemporary conception of higher-level things (we will centrally deal with properties). Multiple realizability pertains to things which can be made real in more than one way, and its existence supports the view that properties in higher-level domains are distinct from the properties that realize them (Pereboom 2002). In this section I wish to develop some lessons from the origins of the conception of multiple realizability. In my view, these lessons invite us to take an appropriate stance toward *Novel* (to be articulated in the next section).

The conventional wisdom is that '[i]n a series of papers published throughout the 1960s, Hilary Putnam introduced multiple realizability into the philosophy of mind' (Bickle 2006).¹ Claims similar to this one have been made by Block (1980; 1990), Kim (1992), Shapiro (2000: 635) and Funkhouser (2007b). However, while Putnam was the first to use the multiple realizability of the mental to argue for non-reductionism, he did not introduce the concept into the philosophy mind. Turing made distinctive use of the concept in the philosophy of mind no less than ten years earlier, when he was thinking about the possibility of building artificial intelligence in his article in *Mind*. Turing wrote:

The storage was to be purely mechanical, using wheels and cards.

The fact that Babbage's Analytical Engine was to be entirely mechanical will help us rid ourselves of a superstition. Importance is often attached to the fact that modern digital computers are electrical, and the nervous system is also electrical. Since Babbage's machine was not electrical, and since all digital computers are in a sense equivalent, we see that this use of electricity cannot be of theoretical importance... If we wish to find such similarities we should look rather for mathematical analogies of function (Turing 1950: 446).

¹ A reviewer for this journal noted that the multiple realizability of mental properties is 'quite obvious and has been known all along'. The above quote, written by a prominent philosopher in the entry on *multiple realizability* in one of the top contemporary encyclopedias of philosophy, shows that this is not so, for it says it was only introduced in the 1960s (e.g. 1960; 1967). That there are many things which can share a property- that properties are multiply exemplified- is of course ancient information, about which Socrates inquired much. But this phenomenon, while conceptually close, cannot be assumed to be the phenomenon of interest to philosophers of mind when they speak of 'multiple realizability'. It must be distinguished from the claim that there is multiple realization- which, according to a live approximation, is sameness of type through differences in the (lower-level) conditions that give rise to instances of that type (Funkhouser 2007a). This is precisely what Turing points to in his above remarks; for claiming that computing properties are realizable by wheels and cards as well as by electrical discharges, is of course to claim that two drastically different lower conditions give rise to instances of the same (computational) type.

As can be seen, this passage contains, in addition, key elements of functionalist and computational theories of mind which are commonly thought to have been introduced into the philosophy of mind by Hilary Putnam at least ten years later.

Theoretical importance, Turing argued, should not be attributed to just any property of considered machines. Two machines might weigh 10 kilograms but this could scarcely suffice for them to realize the same relevant type of computing machine. It is only some properties of physical machines that realize their being computing machines of the relevant types.

Similarly, to realize mentality, what is needed is the realization of certain, but not all, properties of things that have mentality. The hypothesis of computational cognitive science is that realizing certain computational properties of us is sufficient for having certain mental properties in question (a complementary formulation is found in Chalmers (1996: 309; Restrepo 2009)).

In this spirit, in a 1952 BBC discussion with Richard Braithwaite, Geoffrey Jefferson, and Max Newman, Turing said: ‘we are not interested in the fact that the brain has the consistency of cold porridge. We don’t want to say ‘This machine is quite hard, so it isn’t a brain, so it can’t think’’ (495). In order to create artificial intelligence or any other mental phenomenon, we must look at certain properties of the brain and not others. This opens the possibility that there are many ways in which mental properties can be realized.

The brain has certain causal powers in virtue of the properties it has. Normally, a living brain has the power to cause characteristic effects of having the consistency of cold porridge. However, this property and these powers are not important for psychology. Rather, examples of properties of the brain that are interesting for psychology are those that cause intelligent behaviour in persons who display it, those that realize the perception of redness in people who have those perceptions, feeling exhilarated, having pain, etc. The multiple realizability assumption has the fruitful effect of focusing our attention on those properties we wish to understand, and potentially replicate, without irrelevant properties confounding our attention.

Contrary to Turing’s claim, however, if electricity makes a positive causal contribution to our mental life, then it is of theoretical importance for mentality. An alternative multiple realization of mentality will have to duplicate those properties of

electrical transmission which contribute to mentality; never mind that having the consistency of cold porridge is not a property of importance. That positive causal contributions of multiply realizable properties are important features multiple realizations must duplicate is a recent development in the inquiry into the nature of multiple realizability. Shapiro (2000: 647) argues that philosophers who hold the multiple realizability of mental face the challenge of clarifying what multiple realizability amounts to. For if realizing kinds do not differ in the causally relevant properties then we do not have a legitimate case of multiple realizability, and if they do differ, then they are different kinds. "Steel and aluminum are not different realizations of the waiter's corkscrew because, relative to the properties that make them suitable for removing corks, they are identical" (Shapiro 2000: 644). Further, corkscrews can have different colors, but this is not sufficient for their being multiply realizable. Pereboom (2002: 524-525) counters that the difference is that colors do not make positive causal contributions to the nature of corkscrews- what they do- while being made of steel or aluminum does. Pereboom (2002: 525) concludes that '[a]ccordingly, *making a causal contribution to the nature of the thing that has it* might be the notion of causal relevance that is pertinent to a condition on multiple realization. This alternative conception would license steel and aluminum but not distinct colors as multiple realizations of a corkscrew, and for silicon and neural systems to count as multiple realizations of psychological features".

Just like the color of corkscrews, having the consistency of cold porridge does not realize mental properties; while electricity in machines that compute through electrical transmission does. In addition, while a special purpose adding machine that weighs 10 kilograms and a word processing machine that also weighs 10 kilograms have a similarity that may be of interest for some purposes, for other purposes, they are appropriately grouped differently. This interest-relativity is not ontologically degenerate. All things have real differences and similarities, which determine the sort of effects they have. Sometimes we are interested in some properties and some effects, and sometimes in others. It is to be expected that different grouping practices, *on the basis of real shared properties and causal powers*, are employed to serve our interests. This selection of properties in virtue of our interest and the fact that they fit into a causal

pattern is what Macdonald and Macdonald (2006) capture under their understanding of causal relevance.

Searle (1980; 1992) believes that such interest-relativity makes multiply realizable properties ontologically degenerate in comparison to physical properties. However, there are paradigmatic physical properties that are multiply realizable as well. Being a solid is a multiply realizable physical property. Being a solid can be realized by being made of Carbon and it can be realized by being made of Iron, for instance. For some purposes, these properties are grouped together and at other times they are grouped differently for different purposes.

Further, some possible realizations of a higher-level property do not always realize that property: at least in some cases, adequate background conditions need to be in place. Wheels and cards are one of the multiple realizations of digital computers; silicon-based electronic devices are another. Like silicon-based electronic devices, wheels and cards have the characteristic causal powers which enable them to realize the property of being a digital computer. However, in order for a collection of silicon-based electronic devices, and wheels and cards to activate the causal powers which enable them to compute a relevant function, they have to be appropriately put together into a computing machine in question. Just any organization will not do, of course; a specific one is required. Depending on where they are located within the machine, different causal powers of different wheels and cards, and electronic circuits will be activated.

In a particular computation, a few wheels and cards within the machine might move in such a way that we can narrowly attribute to them being the realization of the computation of the function in question. Shoemaker (1984) would call them “core realizations”. However, without the rest of the machine holding them appropriately positioned and timed relative to each other to generate certain effects, those parts would not realize the computation. These background conditions are composed of causal antecedents and part-whole relations of the object in question (in this case, a machine of a particular sort). Background conditions are quite important for core realizations, for without them core realizations would not be successful realizers of higher-level properties. The background conditions of the sort illustrated here together with the core realizations are what is sufficient for the complete realization of a higher-level property,

and channel the causal flow in such a way that the appropriate causal powers of the realization are activated. Shoemaker (1984) calls them ‘total realizations’.

We can see that core realizations by themselves will not be sufficient for higher-level properties they could realize; they need to be under appropriate conditions. When those conditions are in place, however, the realization is complete. It is total realizations that are sufficient for the things they could realize. When core realizations are not under appropriate background conditions, they cannot be reliably expected to realize the property in question and have its characteristic effect.

3. An Alternative Explanation to *Novel*

Whittle’s (2007) *Novel* runs as follows:

... suppose that C-fibres firing realizes pain in me, but sleepiness in my dog, whereas D-fibres firing realizes pain in my dog. One evening, I accidentally touch a hot hob and the pain/C-fibres firing causes me to cry out. Concurrently, my dog is also experiencing C-fibres firing, and yawning as a result. But if he had been in pain, he would have howled, so pain, unlike C-fibres firing, does cause him to cry out. This case is problematic not because C-fibres firing bestows causal powers not associated with pain, i.e. in this case, the causal power to make dogs sleepy; but because pain seems to have a causal power that C-fibres firing doesn’t, namely, the power to make my dog howl (Whittle 2007: 25-26).

We can see that Whittle is committed to the existence of multiple realizability. Whittle takes it that having pain could be realized by having C-fibres fire and by having D-fibres fire. Additionally, Whittle asks us to consider this as a case in which one instance of having C-fibres fire (the dog’s) does not realize having pain and does not have the characteristic effect of having pain (howling). Whittle concludes that having pain has a causal power that having C-fibres fire lacks. This, she thinks, is just a special case of mental properties having causal powers their physical realization bases lack. Consequently, the mental has novel causal powers its physical realization base does not have.

The problem with Whittle’s thought is that there exists a more obvious and simple alternative explanation for her scenario. All that is needed to explain the supposed existence of her scenario is the fact that in her chosen scenario the background conditions (including causal antecedents and part-whole relations) of having C-fibres fire in her and her dog are such that only when they fire in her are the causal powers

characteristic of having pain activated. The total realization was different in her and in her dog. If one changes these background conditions appropriately, having C-fibres fire in the dog would realize having pain and cause the dog to howl.

It might be true in a narrow stipulative sense that if the dog's C-fibres fire, pain would not be realized and the dog would not exhibit the characteristic behavior of having pain. But the alternative explanation is just that in the scenario she chose, the stipulated background conditions do not activate the causal powers relevant to having pain in the dog and that were those background conditions in the dog adjusted, having C-fibres fire in the dog would realize having pain and cause it to howl. For an argument with the conclusion that the mental has causal powers that its physical realization base does not have to work, we should not be able to straightforwardly explain the scenario by these means.

It is an uncontroversial truth that in different circumstances different causal powers of properties are activated, and that the activation of different causal powers of having C-fibres fire in different circumstances is no exception. One should expect quite different effects of having C-fibres fire between a case in which they are hooked up to realize pain and a case in which they are hooked up to realize sleepiness, as Whittle's scenario demands. But provided one adequately adjusts the peculiar way in which Whittle has stipulated her dog's C-fibres are hooked up, her dog's having pain would be realized by its having C-fibres fire and it would howl when they do.

In sum, the dog's having C-fibres fire does not realize his having pain and cause characteristic effects of pain. Whittle explains this fact by saying that having C-fibres fire does not have causal powers that having pain does have. A simpler explanation is just that Whittle has stipulated background conditions for the firing of C-fibers which prevent the causal powers characteristic of having pain from being activated. Were those background conditions to be adjusted, the firing of C-fibers would make pain be instantiated in the dog and cause it to howl.

4. The Empirical Grounds

Whittle thinks this kind of response does not work. She says:

Whilst it is no doubt true to say that, in the right circumstances, if the total realizer had been instantiated by my dog, then he would have howled, this does not establish that, in

the actual world, the total realizer has the causal power to make my dog howl. It may, for instance, be nomologically impossible for the properties of C-fibres firing to realize the pain role in dogs. So, granted such a scenario, claiming that this total realize has the power to make my dog howl in this world, smacks of desperation (Whittle 2007: 26-27).

It is, however, puzzling why Whittle thinks her proposed nomologically impossible scenario should be taken for granted. Normally, humans and dogs have the same realization base for having pain (for humans see Rosenzweig *et al.* 2002; for dogs see Stafford 2007). So given that we can suppose that having C-fibres fire in Whittle is a possible realization of her having pain, we must be able to suppose that having C-fibres fire in her dog is a possible realization of having pain in her dog. The scientific understanding of the realization of having pain supports the idea that if having C-fibres fire is a nomologically possible realization of having pain in a human, then it is a nomologically possible realization of having pain in a dog. Consequently, given that having C-fibres fire in a human could realize having pain and cause pain behaviour, as Whittle supposes, having C-fibres fire in a dog could realize having pain in the dog and cause it to howl.

The only way in which one could grant Whittle what she asks us to grant is by applying in an *ad hoc* manner our understanding of multiple realizability. We would have to say that even though in the actual world the normal realization of having pain in humans and in dogs is the same, without further argument, in a human, but not in a dog, having C-fibres fire could realize having pain.

5. Methodological Dualism Galore

Whittle thinks the kind of alternative explanation proposed here for her chosen scenario has a fatal consequence:

This would be to abandon the idea that there is a mental property of pain. What we would have *instead* is the property of pain-in-humans and the property of pain-in-dogs...perhaps pain-in-humans has a very different quality than pain-in-dogs and thus warrants the ascription of different mental properties (Whittle 2007: footnote 12, emphasis added).

Compare Whittle's statement to someone who said that to suppose that having C-fibres fire happens only under certain circumstances has the unacceptable consequence that there is no such thing as having C-fibres firing itself, but C-fibres-fire-in-humans and C-

fibres-firing-in-dogs, and so on. The first thing to notice is that it is *true* that C-fibres fire only under certain conditions, just like everything else that has a cause and is a part of other things. So it should not be unacceptably surprising that pain likewise occurs under the conditions of human, of dogs, and so on.

It seems that by Whittle's standards we should be prepared to use hyphens for expressing just about any other idea about things in the world, since everything has causes and can be seen to be a part of other things (except, perhaps, the universe as a whole). However, I do not see what the point of that kind of use of hyphens would be, what confusion it would enlighten, what new fact would become clear, whether this would be a real replacement, or what value would be added.

Secondly, the fact that having C-fibres fire only happens under certain circumstances does not imply the elimination of having C-fibres fire *per se*, and its replacement by something else. Causes don't eliminate their effects, and wholes don't eliminate the existence of their parts. Similarly, having pain should not be eliminated and replaced just because it has causes and is a part of other things which form the total realization.

Thirdly, even if having pain has at least two varieties, in humans and in dogs, for example, this would not eliminate pain. There would still be a mental property, having pain, of at least two varieties. Consider the fact that there are atoms of tens of varieties, star cores of many varieties, and geometrical shapes of infinite varieties, for example. How would it entail from this multiplicity that there is no property of being an atom, or of being a star core, or of having a geometrical shape, or that no object has any of these properties (with their characteristic causal powers)? That this multiplicity with unity is commonplace is not only commonsense; it's also scientifically endorsed. For example, while different atomic numbers determine differences in atomic varieties, it is still true that atoms exist and that they are united in that they are composed of a nucleus with (a) proton(s) and (an) outer layer(s) of electrons. Further, there might well be a qualitative difference between dogs having pain and humans having pain, but it would still be a difference in their *having pain*, just as while there is a qualitative difference between Hydrogen and Helium, they are still atoms. If the answer to the above question is that there is no such entailment, and that this multiplicity with unity is *bona fide* and commonplace, then the unpersuasive and *ad hoc* move is in fact being made Whittle by

just assuming that somehow the mental-to- physical case is different from the rest of the physical world in this respect.

6. Noordhof's novel mental causal powers

Before Whittle, Noordhof (1997: 242) put forth the *Novel* argument like this:

Suppose that a causal role property C can be realized by three grounds: G1, G2 and G3. Let these grounds be causes of three other properties that variably realize a role-characterizing property S. Call these other realizing properties R1, R2, and R3. Then a common feature of variable realization is that although G1 - R1, G2 - R2 and G3 - R3 (where '-' means 'is causally related to') it is not the case that G1 - R2 nor is it the case that G2 - R3. The grounds of a causal role property are causally connected with some realizing properties of a role-characterizing property and not others. So although C is causally connected to S however S is realized, the grounds are not so connected. My thought is that it is in this sense that causal role properties introduce new causal relations. G1 may determine the instantiation of C because G1 is a member of the family of properties upon which C supervenes, but C introduces relations in which G1 couldn't possibly stand.

Perhaps a multiply realizable property has different causal powers than *one* possible realization base. Nevertheless the following theses seem to be unaffected: (i) that all the causal powers of the C coincide with (at least some of) the causal powers of the complete set of multiple realizers, and (ii) that in any particular case where C causes S, all the causal powers of C will be completely coincident with (at least some of) the causal powers C's ground, in virtue of which S is caused. To claim that there are causal relations between higher-level properties not coincident with causal relations between their realizers requires additional justification.

Suppose that there is some grounding property G1 causally related to R1, but not to R2, as Noordhof suggests. *On Noordhof's supposition, when C is realized by G1 it will not be able to cause R2.* The causal powers of C, when realized by G1, are completely limited by the causal powers of G1; not some other possible realization of C. The same is true of any other multiple realization of C: it will constrain the causal powers of C anytime it realizes C. It is hard to see a novelty worth asserting here.

7. Clarke's novel mental causal powers

Clarke puts his version of *Novel* like this:

... M must carry with it causal powers that no one of its base properties has. Suppose that M is realized in a human being by P1, in a lizard by P2, in an octopus by P3, and so on. Consider the causal powers of P1. Presumably, an exemplifying of P1 in a human will cause an exemplifying of P1*. What, however, about P1 in a lizard, in an octopus, etc.? Recall that P1 is a micro-structural property of a human being; it is, for example, a property that a creature has in virtue of the presence within that creature's brain of a certain pattern of neural activity in the structure that is characteristic of a human brain. It would seem to be impossible (at least nomologically) for P1 to be instantiated in a lizard or an octopus. Indeed, it appears that P1 could only be instantiated in a human being. If so, then P1 cannot be instantiated in a lizard, in an octopus, etc., and it cannot carry with it the power to cause in a lizard an exemplifying of P2*, the power to cause in an octopus an exemplifying of P3*, etc. P1, then, does not carry with it some of the causal powers that M carries with it. Or, to put it the other way around, M carries with it some causal powers that P1, one of the subvenient physical properties that realizes it, does not carry. And likewise for the relation of M's causal powers to those of any other subvenient physical property that realizes it (Clarke 1999: 304-305).

Suppose the realization bases of interest are total realizations and are necessarily different, as is most plausible when considering humans and octopi. Consider Babbage's machines and electronic computers. They provide different total realizations, but they can compute the same relevant functions, and in this important sense they are equivalent. Two adding machines with these variable realizations will have the property of processing the addition of 8 and 3. Call one realization in Babbage's wheels and cards machine *C1* and the other realization in the electronic machine *C2*. These two variably realizing properties, causally deliver the same effect: say an 11 paper print. Consequently, it is not necessarily true that two different total realizations do not have the same effect. Similarly, humans and octopi may totally realize being in pain differently, but that realization may have equivalent effects (pain behavior).

True, the pain behavior of humans and octopi have different total realizations and the total realization of pain in humans does not cause the total realization of pain in octopi, nor vice versa. But this case completely correlates with the fact that the pain of humans does not cause the pain behavior of octopi. So it does not support the idea that pain causes things its realizations don't.

Suppose a human has M in virtue of having P1, as Clarke supposes, and an octopus has M in virtue of having P3. Clarke's analysis is that this human's having M by having P1 will cause the human to have P1*, but will not cause P3*, whereas the octopus' having M by having P3 will cause it to have P3*.

However, just as the human's having M will not cause P3*, the octopus' having M will not cause P1*. The differences seen in the causal profile of M are perfectly

correlated with the differences in the causal profile of its multiple total realizations. When M is realized by P1, it will cause P1* and not P3*, and when M is realized by P3, it will cause P3* and not P1*.

We can account for the differential effects of M by considering that it is realized by P1 in one case and by P3 in another, as a consequence of the fact that Clarke supposes: that only one of them could be realized in the human and the other in the octopus and that whenever they are realized in their corresponding organism they have certain characteristic effects and not others. What causal powers does M have that cannot be fully accounted for by its realization base? When M is had by a human it causes P1* in the human and not in octopi, and when M is had by octopi it causes P3* in the octopi and not in the human. For *Novel* to work, it requires that M cause something that cannot be fully accounted for in this way.

As responded to Noordhof, on the supposition that M is multiply realizable, the following theses seem to be unaffected: (i) the causal powers of M will completely coincide with, and never outstrip, the causal powers of its realizers (or a subset thereof), and (ii) in any particular thing that has M, the causal powers of M will not be anything over and above the causal powers of its actual realization base. A case worth arguing for needs to provide grounds to deny these theses.

8. Reductionism

Whittle, Noordhof, and Clarke use the *Novel* argument on the way to providing a non-reductionist solution the problem of causal exclusion. Further, some of the remarks I have made have been in line with some of what Kim has argued (though Kim has not addressed *Novel* directly). So it may seem that my negativity towards *Novel* has the ultimate goal of supporting reductionism. This, however, is far from the case because I think Kim's argument is logically invalid for the reasons Restrepo (2012a; 2012b) expresses. While this is not the place to do full justice to the causal exclusion argument, since this paper is centrally about certain aspects of multiple realizability and the *Novel* argument, it is framed in the causal exclusion debate; so I must at least gesture at a key line of reasoning for not endorsing Kim's reductionist conclusion.

A key premise in Kim's argument is *Closure*: Each physical event has a sufficient physical cause (insofar as causes of events are sufficient) (Kim 2005). Of

course, given physicalism, every cause is physical. But being physical can be a fundamental physical property (a physical₁ property) or a property that is completely resultant from, but not reductively identical with, fundamental properties (a physical₂ property). Jackson illustrates this nicely:

We need... an extended sense (of the physical) because the patterns that economics, architecture, politics and very arguably psychology, pick out and theorise in terms of, include many that do not figure in the physical sciences. The reason is no mystery: it is that aggregation creates new properties ... because aggregations fall under patterns, kinds etc. that the items they are aggregations of do not fall under... Physicalists must allow that the world contains aggregations that have properties that are not physical₁ properties for the same reason, when all is said and done, that someone who holds, rightly, that a triangle is an aggregation of straight lines must allow that the triangle is not itself a straight line (Jackson 2006: 234, parentheses added).

The view I favor is that mental properties are physical₂ properties, which leaves the conclusion *that mental properties are physical causes* without its putative reductionist element. This is not inconsistent with Kim's (2005: 43) explicit conception of the physical as it figures in *Closure*, which he borrows from Papineau (2002). This conception is that the physical is to be understood as what is inorganically identifiable, or "identifiable non-mentally-and-non-biologically". That is, the physical is what can be referred to "independently of this specifically mental conceptual apparatus" (Papineau 2002: 41). This means that those concepts that do not use terms like *seeing* and *believing* are sufficient to refer to anything that exists in nature and participates causally in the world. Certainly, one can refer to triangles, like mental physical₂ properties, without explicitly using the term *triangle*, or *mental*. This could be done with terms like *lines put together in a certain way*, or *quarks and electrons put together in a certain way*. This, however, does not eliminate the causal relevance of physical₂ properties and a logically valid reductionist conclusion is blocked. The details of this "logical reply" are provided by Restrepo (2012a).

9. Conclusion

Whittle (2007) mentions *Novel* in developing a solution to the problem of causal exclusion. She, however, finds her own solution problematic because she thinks that the mental has powers not had by its physical realization base. The problem of causal

exclusion attempts to show that the mental does not have irreducible mental causal powers. To worry that one's response to this argument is problematic because of the supposed fact that the mental has *additional* causal powers is surprising in its own right. Nevertheless, it may be of theoretical comfort that *Novel* does not really pose an obstacle to any solution to the problem of causal exclusion. This is something that reflection on Turing's comments on building computing and thinking machines sheds some light on.

Ricardo Restrepo

Instituto de Altos Estudios Nacionales, Ecuador

ricardo.restrepo@iaen.edu.ec

References

- Bickle, J. (2006) 'Multiple Realizability' IN *Stanford Encyclopedia of Philosophy*, E. Zalta (Ed.), <http://plato.stanford.edu/entries/multiple-realizability/> . Accessed August 18, 2011.
- Block, N. (1980) 'Intro: What is Functionalism?' IN *Readings in Philosophy of Psychology* vol. 1, Cambridge: Harvard University Press.
- _____. (1990) 'Can the Mind Change the World?' IN *Meaning and Method*, G. Boolos (Ed.), Cambridge: Cambridge University Press.
- Chalmers, D. (1996) 'Does a Rock Implement Every Finite-State Automaton?', *Synthese*, 108, 309-333.
- Clarke, R. (1999) 'Nonreductive Physicalism and the Causal Powers of the Mental', *Erkenntnis*, 51, 259-322.
- Funkhouser, E. (2007a) 'A Liberal Conception of Multiple Realizability', *Philosophical Studies*, 132, 467-494.
- _____. (2007b) 'Multiple Realizability', *Philosophy Compass*, 2, 303-315.
- Jackson, F. (2006) 'On Ensuring that *Physicalism* is not a Dual Attribute Theory in *Sheep's Clothing*', *Philosophical Studies*, 131, 227-249.
- Kim, J. (1992) 'Multiple Realizability and the Metaphysics of Reduction', *Philosophy and Phenomenological Research*, 52, 1-26.
- _____. (2005) *Physicalism, or Something Near Enough*, Princeton: Princeton University Press.
- Macdonald, C. and Macdonald, G. (2006) 'The Metaphysics of Mental Causation', *The Journal of Philosophy*, 103, 539-576.
- Noordhof, P. (1997) 'Making the Change: the Functionalist's Way', *British Journal of the Philosophy of Science*, 48, 233-250.
- Papineau, D. (2002) *Thinking about Consciousness*, Oxford: Oxford University Press.

- Pereboom, D. (2002) 'Robust Non-Reductive Materialism', *The Journal of Philosophy*, 99, 499-531.
- Putnam, H. 1960 (1975) 'Minds and Machines' IN *Mind, Language, and Reality*, Cambridge: Cambridge University Press.
- _____. (1967) 'Psychological Predicates' IN W.H. Capitan and D.D. Merrill (eds.), *Art, Mind, and Religion*, Pittsburgh: Pittsburgh University Press.
- Restrepo, R. (2009) 'Russell's Structuralism and the Supposed Death of Computational Cognitive Science', *Minds and Machines*, 19, 181-197.
- _____. (2012a) 'Two Myths of Psychophysical Reductionism'. *Open Journal of Philosophy*, 2, in print. A version can be found at <http://philpapers.org/archive/RESTMO-3>. Accessed Feb. 4, 2012.
- _____. (2012b) 'Thinking about Physicalism'. *Open Journal of Philosophy*, 2, in print. A version can be found at <http://philpapers.org/archive/RESTAP>. Accessed Feb. 4, 2012.
- Rosenzweig, M., Breedlove, M. and Leiman, A. (2002) *Biological Psychology*, Sunderland: Sinauer Associates Inc.
- Searle, J. (1980) 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3, 417-57.
- _____. (1992) *The Rediscovery of the Mind*, Cambridge: MIT Press.
- Shapiro, L. (2000) 'Multiple Realizations', *The Journal of Philosophy*, 97, 635-54.
- Shoemaker, S. (1984) *Identity, Cause and Mind*, Cambridge: Cambridge University Press.
- Stafford, K. (2007) 'Pain in Dogs: Its Recognition and Alleviation', IN *The Welfare of Dogs*, Dordrecht: Springer.
- Turing, A. 1950 (2004) 'Computing Machinery and Intelligence', IN J. Copeland (Ed.), *The Essential Turing*, Oxford: Oxford University Press.
- Turing, A., Braithwaite, R., Jefferson, G., and Newman, M. 1952 (2004) 'Can Automatic Calculating Machines Be Said to Think?' IN J. Copeland (Ed.), *The Essential Turing*, Oxford: Oxford University Press.
- Whittle, A. (2007) 'The Co-Instantiation Thesis', *Australasian Journal of Philosophy*, 85, 61-79, page numbers used from URL = http://personalpages.manchester.ac.uk/staff/ann.whittle/documents/coinstantiation_thesis.pdf.

PHENOMENAL CHARACTER AS THE MODE OF PRESENTATION OF ENVIRONMENTAL PROPERTIES

Roberto Horácio de Sá Pereira

Abstract

The purpose of this paper is to defend and further develop an account of the phenomenal character of perceptual experience. Rather than identify the phenomenal character with the intrinsic properties represented by perceptual experience (phenomenal externalism), my aim is to support the alternative claim that the phenomenal character of a perceptual experience is to be identified with the mode of presentation of environmental properties.

0. Introduction

As Block says, the concept of consciousness is a mongrel one: consciousness is articulated in different ways. Among these, the one that certainly raises more problems is consciousness in the phenomenal sense. Our perceptual experiences differ from our other intentional states, because they have essentially phenomenological features. When tasting a wine, we become aware of its tannins, aroma, ruby color, or other traits. The properties of the wine phenomenally appear to us in a specific way. However, if perceptual experiences have essentially phenomenological features, like other intentional states, they are also characterized by their intentional status, that is, their particular representational content. As propositional attitudes, perceptual experiences place certain satisfaction conditions on the world that, once satisfied, make the experiences in question veridical.

It is important to note that, although natural, this assumption is in no way consensual in the contemporary philosophy of perception. Numerous authors reject the supposed representational status of perceptual experience. According to them, instead of placing accuracy conditions on the world, perceptual experiences put us in direct contact with the objects perceived (relationalism). Thus, normal perceptual experiences (veridical or illusory) and hallucinatory experiences have nothing in common (disjunctivism). However, to undertake here a defense of the representational status of perceptual experience would lead me far afield, requiring, in fact, a new paper. Thus, the fundamental axis on which the present work hinges may be expressed in a

conditional form: if we endorse the idea that at least so-called perceptual experiences have not only an essentially phenomenal character but also representational content, it is natural to suppose that there is a connection between these two crucial aspects of the experience.

Such a connection places few constraints on any satisfactory view of perceptual experience that here takes the form of three desiderata. However, it is important to emphasize that, in contrast to the status of representational content of perceptual experiences, all three desiderata are highly controversial. Indeed, the truth is that none of the important names in contemporary philosophy of perception is willing to acknowledge the legitimacy of the three desiderata together. Despite these circumstances, I find all of the desiderata intuitive, and I reiterate that only a conception that satisfies all three as a whole may be considered satisfactory.

The first is the thesis that perceptual states are individuated, in part on the basis of their representational content, and these, in turn, on the basis of patterns of relations that species (to which the perceiver belongs) has with the different objects, properties, and kinds of natural environments. I call this first desideratum *representational externalism*.

The second desideratum establishes the most tenuous connection between the phenomenal character of perceptual experience and its representational content: the phenomenal character of perceptual experience *determines* its representational content in the sense that there could be no difference in representational content if there was no difference in phenomenal character. The fundamental idea here is that the phenomenal character is a crucial element in recognizing the representational content of perceptual experience. In the absence of a better name, I call this second desideratum *representationalism*.

In order to avoid misunderstanding, it is important to distinguish representationalism, the label I use here, from the thesis expressed by the same label according to which the representational content determines the phenomenal character of experience in the sense that there could be no difference in phenomenal character if there was no difference in representational content. In short, we can have perceptual experiences with different phenomenal characters representing the same content, but not

the other way around; perceptual experiences with the same phenomenal character cannot represent different content.

For this reason, it worth noticing that, as I conceive it here, representationalism is agnostic about the ways the relationship between phenomenal character and representational content are usually understood. This relationship is neutral on the assumption that the phenomenal character of experience is founded on its representational content. In other words, it is agnostic about the assumption that we can provide a reductionist account of the phenomenal character in non-phenomenal terms. For the same reason, representationalism is also agnostic about the identification of phenomenal character with the properties of the natural environment represented by the experience, a thesis that, from now on, I refer to as *phenomenal externalism*.

The third and last desideratum is the thesis of the local supervenience of phenomenal character of perceptual experience on the relevant physical properties of the biological substrate. In opposition to doxastic states, the phenomenal character of perceptual experience seems to depend crucially on the physical properties of the biological substrate from which individuals are made.

Taken together, the desiderata seem inconsistent. The problem takes the form of a classic trilemma in which the satisfaction of each pair excludes the possibility of satisfying the remaining third. In principle, the trilemma would be insoluble, as only the possibility of its *dissolution* by the rejection of one or more desiderata would remain. The most important contemporary positions on this topic can be characterized according to the different attempts to dissolve the trilemma.

Thus, despite their differences, by accepting both representationalism and representational externalism, Harman (1990), Dretske (1995), and Tye (1995) are forced, inter alia, to reject the local supervenience of phenomenal character on the biological substrate and hence to embrace what I previously named phenomenal externalism. That is, they must embrace the thesis that identifies the phenomenal character of perceptual experience with the physical environmental properties represented by experience.

However, what the satisfaction of the desideratum of representational externalism effectively excludes is what we may refer to here as *representational internalism*, that is, the assumption that the content of perceptual experience, unlike the

doxastic content of propositional attitudes, can be individuated and constituted independently of any reference to the perceiver's external environment. What satisfaction of the desideratum of representationalism excludes is what Block (2003, p. 165) calls *phenomenalism*, that is, the assumption that the phenomenal character of perceptual experience *outruns* its representational content, or the assumption that neither does representational content determine phenomenal character nor does that phenomenal character determine representational content. Now, even if we reject both representational internalism and phenomenalism, we do not need to give up what I have called phenomenal internalism.

The second proposal to dissolve the trilemma is formulated by Block (2003). Block endorses representational externalism. However, in response to the phenomenal externalism of Dretske, Tye, and others, he opposes phenomenal internalism, that is, the thesis that the phenomenal character locally supervenes on the biological substrate of the organism, or “depends on the details of the physiology or physico-chemical realization of the computational structure of the brain” (2003, p. 166). Thus, he must reject what I have here called representationalism and embrace what he calls phenomenalism: that neither the representational content determines the phenomenal character of experience, nor the phenomenal character of experience determines the representational content.

Now, the assumption of phenomenal internalism (the local supervenience of the phenomenal character of experience on the biological substrate) rules out phenomenal externalism (the assumption that phenomenal character consists of the properties represented by perceptual experience). Further, the assumption of representational externalism excludes representational internalism, that is, the assumption that the content of perceptual experiences can be individuated independently of any possible relations an individual may have with different objects, properties, and kinds of natural environments.

Thau (2002) suggests the third and last form of dissolution of the trilemma. Like Block, Thau endorses phenomenal internalism. However, Thau supports representationalism. This means he must reject representational externalism and embrace what I have called here representational internalism; both the phenomenal character and the representational content of perceptual experience would be

individuated independently of any reference to an individual's external environment. However, what the assumption of representationalism excludes is phenomenalism, and what the assumption of phenomenal internalism excludes is phenomenal externalism.

In this paper, I intend to show that contradiction between the three desiderata is only apparent, or, in other words, that there is a solution to the trilemma. The simultaneous satisfaction of the three desiderata depends on the assumption of two central theses. The first concerns the representational content of perceptual experience. If it is true that content is individuated, *in part*, by patterns of relations that the individual has with the objects, kinds, and properties of his natural environment (representational externalism), it is also undeniable that such content is individuated, *in part*, on the basis of the mode of presentation of those entities. Thus, the content of perceptual experience cannot be Russellian, that is, purely referential. In addition to objects, properties, and relations, such content is also constituted by the way that these entities are given to perceptual experience.

The second fundamental thesis concerns the phenomenal character of perceptual experience. As representational content is also individuated, in part, by the individual's discriminatory and recognitional abilities, the natural assumption is to understand that the phenomenal character of perceptual experience is the conscious element by which objects, properties, and kinds of natural environments are *given to* perceptual experience. Thus, based on a critical examination of Dretske's design problem, I claim in this article that the phenomenal character of perceptual experience is nothing but a way of processing information about the instantiation of some property peculiar to individuals, groups, or entire species. It is because of the phenomenal characteristics that emerge from the causal interaction of the individual with the properties of his natural environment that his neural states are recruited by natural selection to indicate that such properties are instantiated.

In addition to this general introduction, this work comprises five further sections. In each, I want to make plausible each of the three desiderata, showing, at the same time, its compatibility with other remaining desiderata. However, it is important to note that I do not intend here to refute any of the three opposing theses (namely, phenomenal externalism, representational internalism, and phenomenalism). Such an enterprise would extend beyond the limits of a single article. I critically appreciate these

competing views only to the extent that they help motivate and situate the view I develop and defend. My aim is to show that the alternative proposal defended here is more plausible than any of its competitors.

1. Phenomenal Internalism: Representationalism without Phenomenal Externalism

Assume initially (a) that the phenomenal character determines the representational content of experience (representationalism) and (b) that perceptual states are individuated, *in part*, by their representational content and these, in turn, individuated, in part, on the basis of the patterns of relations that the species to which the individual belongs have with the different objects, properties, and kinds of species in a given natural environment (external representationalism). The initial question that arises is whether, assuming (a) and (b), we are committed to phenomenal externalism, according to which the phenomenal character of perceptual experience is identical to the physical properties represented by experience.

Inverted spectrum scenarios, originally suggested by Shoemaker (1994), raise doubts about phenomenal externalism. Suppose that John and Peter are individuals who grew up and live in the same environment, belong to the same species, and belong to the same speech community. Consequently, they discriminate and represent the colors red and green in exactly the same way. Both agree that ripe tomatoes are red and unripe ones are green. This supports the assumption that their perceptual states are individuated, in part, by their interactions with the red and green colors of the natural environment they inhabit. Thus, their visual experiences represent the color of unripe tomatoes as green and the color of the ripe tomatoes as red. Yet the way the color of the unripe tomato phenomenally appears to John (the inverted) is the same as the way the color of ripe tomatoes phenomenally appears to Peter. In sum, respective visual experiences of the same color possess different phenomenal characteristics.

The natural assumption here would be that the phenomenal differences between Peter's and John's perceptual experiences is due to a physical distinction between them. Relative to some relevant physical aspect of their brains, or perhaps relative to certain relevant physical aspects of their respective visual apparatuses, Peter is different from John. Thus, the assumption that the phenomenal character of perceptual experience locally supervenes on physical properties of the brain or of sensorial apparatuses

(phenomenal internalism) is the only one consistent with the philosophical intuition expressed in these scenarios of inversion. If two individuals are functional replicas in the sense that they likewise represent colors, although their visual experiences of colors differ phenomenally, the moral to be drawn is that they must differ in relation to some relevant physical aspect.

However, the simple scenario of inversion is not *per se* a decisive argument against phenomenal externalism. Under the assumption that phenomenal character is one and the same as the physical properties represented by experience (phenomenal externalism), the inverted would be wrong; that is, he would be misrepresenting the colors, albeit systematically. If John's visual experiences of unripe tomatoes represent them as red, and if his visual experiences of the ripe ones represent them as green, then John misrepresents the colors in the sense that his experiences do not correctly track the colors of his environment.

This problem arises only in phenomenal externalism when we add the (quite plausible) assumption that scenarios of inversion are relatively common. Under this assumption, we have no means of non-arbitrarily telling apart *normal* individuals, whose experiences were correctly tracking colors, and the *abnormal* individuals, whose experiences were tracking them incorrectly. So, if by means of his visual experiences of ripe tomatoes (whose phenomenal character is, say, phenomenal redness) Peter correctly represent their color as red, Peter's visual experience of the same ripe tomatoes (whose phenomenal character is the phenomenal redness) also correctly represents their color as red.

As I anticipated in the introduction, I do not intend here to refute phenomenal externalism. That would lead me far afield. My aim is only to point out the greater plausibility of phenomenal internalism. When considering single individuals or subpopulations of a species, we can perhaps accept the characterization of perceptual states of the inverted as misrepresentations of colors. However, when considering entire species, it is far more plausible to assume that the perceptual experiences of these individuals are representing these colors correctly. Further, what explains the phenomenal difference between experiences of *normal* individuals and experiences of the inverted is their distinctive biological makeup.

2. Representational Externalism:

Phenomenal Internalism without Representational Internalism

Suppose now (c) that the phenomenal character of perceptual experiences locally supervenes on the biological substrate (phenomenal internalism) and (a) that the phenomenal character of perceptual experience determines its representational content (representationalism). The question that now arises is whether, assuming (c) and (a), we are committed to representational internalism. Thau's account is a case in point. According to him, inversion spectra scenarios establish only that the phenomenal features of perceptual experiences are not determined by the way "the outside world is represented" (2002, p. 31). Nonetheless, nothing prevents the supposition that these phenomenological characteristics are determined by the way in which the subject represents external properties *for himself*. Thus, unripe tomatoes not only appear phenomenally red to the inverted; they also represent as red for him. Thus, the phenomenal character of the perceptual supervenes locally on the physical properties of the biological substrate (phenomenal internalism), and determines the representational content of the experience (representationalism), but only because such content is individuated independently of the external environment (representational internalism).

It is inevitable to think that one of the motivations of representational internalism is its adherence (albeit unconsciously) to the traditional act-object model. In light of this model, the experience of an external object is understood as the perception of an internal image resembling the object, thus mediating our cognitive access to it. We would become directly aware of this datum only indirectly through the external properties of objects. We would perceive a red bulgy tomato by means of our immediate perception of an internal datum that is red and bulgy.

Thus, the color of the unripe and ripe tomato not only appears phenomenally inverted to normal and inverted individuals, they would also be represented in different ways, since the internal data would be materially different. Normal and inverted people only agree in relation to the color of ripe and unripe tomatoes in the doxastic domain of beliefs and thoughts, that is, when they acquire the concepts of "red" and of "green" by means of the acquisition of language.

However, if we accept that the content of doxastic states are individuated, in part, based on patterns of relations the species to which the individual belongs has with

the different objects, properties, and kind of natural environment (doxastic externalism), how could we assume that the content of perceptual states is individuated independently of the external environment?

The inverted Earth scenario was originally conceived by Block as an argument against phenomenal externalism. However, it also allows us to illustrate the difficulties of representational internalism. Like Putnam's famous Twin-Earth idea, the inverted Earth would be similar to Earth, except for two crucial aspects. (I) All objects that are red on Earth are green there, and vice-versa; all objects that are green on Earth are red there. (II) People on the inverted Earth employ the predicate "red" to refer to the color of green objects, and the predicate "green" to refer to the color of red objects. Suppose now that Peter is transported to the inverted Earth through inverting his lenses. He could not possibly notice any difference. If he contemplated a ripe tomato on the inverted Earth, the phenomenal character of his visual experience would be qualitatively identical (in all its relevant aspects) to the phenomenal character of his visual experience of a ripe tomato on Earth.

The crucial point is the following. If, according to Putnam's original thought experiment, Oscar would be wrongly thinking of water by contemplating the substance in the rivers, lakes, and oceans of the Twin-Earth, Peter would also be misrepresenting the color of ripe tomatoes as red and the color of unripe tomatoes as green on inverted Earth and for exactly the same reason. If Oscar's doxastic states are individuated in part by their representational content and these, in turn, based on patterns of relations that individuals of Oscar's community have with H₂O on Earth, Peter's perceptual states would also be individuated in part by their representational content. This content would be based on patterns of relations that individuals of the species to which Peter belongs have with the color red on Earth. If Oscar is mistaken when he thinks of water (H₂O) on the Twin-Earth, Peter also must be wrong if he represents a ripe tomato as red on the inverted Earth.

Interestingly, Block believes that once Peter has adapted to the inverted Earth, he will begin to correctly represent the ripe tomato as green and the unripe tomato as red, even though ripe tomatoes appear phenomenally red to him while unripe tomatoes appear phenomenally green. Block notwithstanding, it is much more reasonable to think that Peter's visual experiences and memories are causally tied to his original

environment. The causal ties are linked to recognitional abilities, which are prompted by the qualitative nature of experiences. Thus, when transported to the inverted Earth through inverting lenses, Peter must be misrepresenting the colors of ripe and unripe tomatoes when he contemplates them.

The fundamental question that now arises is whether the rejection of representational internalism also commits us to the rejection of phenomenal internalism, that is, the rejection of the assumption that the phenomenal character of perceptual experiences locally supervenes on the physical properties of their biological substrates. In other words, we wonder if representational externalism necessarily leads to phenomenal externalism.

Indeed, this is Dretske's position. In his famous 1995 work, he initially suggests that the main motivation of phenomenal internalism would be what I called above the act-object model (Dretske, 1995, pp. 127-128). In light of this model, we would become directly aware of the phenomenal character of our experience and only indirectly of external objects and properties that the same experience represents. Now, if representational internalism seems to assume an adherence, though implicit, to the act-object model, we cannot say the same of phenomenal internalism. To say that the phenomenal character of perceptual experience is "in the head" is not to endorse in any way the literal assumption that we would perceive such experience as a datum inside our consciousness, or, more precisely, as a supposed mental image. It only means we assume that phenomenal character supervenes locally on the physical properties of a biological substrate.

One year later, Dretske (1996) acknowledges that he can provide no argument that counters phenomenal internalism. All he has against it is that, in his opinion, it is incompatible with the main intuition—that the sensory qualities through which perceptual experiences are individuated "are not in the person where it is the experience" (Dretske, 1996, p. 144). The idea is that sensory qualities are not properties of experiences themselves, but rather "relational properties" (1996, p. 145). In representational externalism, we identify beliefs based on what they represent. Likewise, we would identify the sensory qualities of experience based on what they represent. Thus, the only reason Dretske presents for rejecting phenomenal internalism is its previous adherence to phenomenal externalism, that is, the assumption that the

phenomenal character of perceptual experience is one and the same as the physical properties represented by the experience.

The internal phenomenal stance is not only compatible with the external representational one, but it also provides a more plausible alternative than phenomenal externalism. The inverted Earth scenario allows us not only to counterfactually dissociate the phenomenal character of perceptual experience from the physical properties such experience represents. The scenario clearly also suggests that the phenomenal character locally supervenes on the physical properties of a given organism. If, when on the inverted Earth, Peter misperceives a ripe tomato as being red because his experience of red is caused by inverting lenses, then phenomenal redness supervenes locally on the physical properties of his perceptual apparatus.

3. Representationalism: Phenomenal Internalism without Phenomenism

Suppose, finally, (d) that perceptual states are individuated, in part, by their representational content and this, in turn, in reference to patterns of relations that the species to which the individual belongs have with objects, properties, and kinds of external environments, and (c) that the phenomenal character of perceptual experiences locally supervenes on the biological substrate (phenomenal internalism). The question that now arises is whether by assuming (d) and (c) we are committed to the rejection of what I have called representationalism and hence to the acceptance of phenomenalism, according to which the phenomenal character of perceptual experience does not determine the representational content.

As we saw, the scenarios of inversion raise doubts about phenomenal externalism, that is, the thesis that the phenomenal character of perceptual experience is one and the same as the physical properties represented by such an experience. It is reasonable to suppose that John (the inverted) represents unripe tomatoes as green (just like Peter) although they phenomenally appear to him just as ripe tomatoes phenomenally appear to Peter. However, if phenomenalism is correct, the reverse must also be true. It would thus be reasonable to assume that John's and Peter's visual experiences could represent ripe tomatoes as red and unripe tomato as green, even though they appear to them in the same phenomenal way.

The crucial difficulty of phenomenalism lies in the fact that the phenomenal character of perceptual experience is one of the key elements in the individuation of representational content. As we saw, representational content is individuated, *in part*, on the basis of different objects, properties, and kinds of perceivers' natural environments. Nevertheless, representational content is also individuated, *in part*, on the basis of the perceiver's discriminatory and recognitional abilities. John (the inverted) is only able to recognize the ripe tomato as red by means of his experiences of phenomenal green, just as he is only able to recognize unripe tomatoes as green through his experience of phenomenal redness. Thus, if John's visual experiences of ripe tomatoes and unripe tomatoes possess exactly the same phenomenal character, he could not possibly represent them as different.

It is worth emphasizing, once again, that I do not intend here to refute phenomenalism. My aim is only to show that representationalism (in conjunction with phenomenal internalism and representational externalism) provides us with a far more plausible picture of perceptual experience than the alternative represented by phenomenalism in conjunction with representational externalism.

4. Phenomenal Mode of Presentation

However, if phenomenal character is one of the key elements of individuation of perceptual content and of perceptual modes of presentation, the question that arises is how we should understand such modes of presentation and such content. Frege has never clearly defined what he means by "mode of presentation" of the reference or "the ways that objects are given" <Gegebenheitsweise der Gegenstände>. His most famous examples are properties uniquely instantiated, such as the "property of being the evening star" and "property of being the morning star," as the ways of presenting the referent of "Phosphorus" and "Hesperus," respectively.

For an initial approximation, following Chalmers (2004, 2010), "modes of presentation" can be cast out to extensions of linguistic expressions as certain identifying conditions. For example, the color red (understood as a physical property of reflectance of a certain spectrum of light) can be identified, roughly, *as the property that normally causes experiences of phenomenal redness in certain individuals under normal lighting conditions* (Chalmers, 2010). In a more precise way, perceptual modes

of presentation can be formally modeled on Chalmers' two-dimensional semantic framework as functions from centered worlds (in which a subject, a given time, and the perceiver's perceptual experiences are at the center) to extensions. When we consider as an argument of this function a possible world for Peter and imagine his experience of phenomenal redness at the center, the value of the function is the color red, understood as a physical property of light-reflectance. In contrast, when we consider as an argument of this function a possible world with John and his experience of phenomenal redness at the center, then we have as a value the color green, understood as a physical property of light-reflectance.

Such an approach raises a question about the status of perceptual modes of presentation of objects and properties, namely, whether they are descriptive or not. Descriptive modes of presentation (*de dicto*) are essentially characterized by the fact that the reference is determined indirectly, that is, by the subject's propositional knowledge that the referents uniquely satisfy one of those identifying conditions expressed by the respective mode of presentation. For example, if my reference to the color red is determined by my knowledge that the color red satisfies the identifying condition of being the color *that normally causes experiences of phenomenal redness in certain individuals under normal lighting conditions*, then the mode of presentation in question is descriptive or *de dicto*. In contrast, non-descriptive modes of presentation (*de re*) are essentially characterized by the fact that the reference is determined in a purely relational way, or by means of the existence of some relation between the subject and referent (Bach, 1987). Consequently, if my reference to the color red is determined by the fact that this color is what *normally causes experiences of phenomenal redness in certain individuals under normal lighting conditions*, then the mode of presentation in question is non-descriptive or *de re*.

Now, when the subject refers to the color red through his visual experience, he does not do so based on the propositional knowledge that *such color experiences typically cause experiences of phenomenal redness in individuals such as himself*. Such an assumption would be a form of hyper-intellectualism, and even Chalmers recognizes that the perceiver could not possibly be representing the color red in terms of the proposed description (2010, pp. 368-369). In short, Peter's visual experience of a ripe tomato represents the color of the tomato as red *because* that color is causing his visual

experience, but certainly not *as* the color that uniquely satisfies the condition of being the color that normally causes experiences of phenomenal redness in individuals like him under normal conditions. That said, perceptual modes of presentation are essentially non-descriptive.

Now, although the perceptual mode of presentation is *de re*, or non-descriptive, representational content cannot be reduced to Russellian content consisting only of one or more physical properties, like the property of light-reflection, as phenomenal externalism assumes. For one thing, representational content is individuated not only on the basis of the patterns of relations between the individual and the objects, kinds, and properties of the natural environment. It is also individuated based on the individual's recognitional and discriminatory abilities. Peter is only capable of recognizing the color red by means of his visual experiences of phenomenal redness, while John is only capable of recognizing the same color by means of his experiences of phenomenal greenness. Thus, although Peter's and John's visual experiences of ripe tomatoes refer to the same red color, their representational content differs slightly. While Peter's visual experience represents the color red *as* the color that appears phenomenally red to him, John's experience represents the same color *as* that which appears phenomenally green to him.

Three observations are crucial here. First, when I say, for example, that John represents red *as* the color that appears green to him, I do not presume that John has to possess the concept of green to refer to the color in question. The representational content of experience is non-conceptual content in the broadest sense of the term: the perceiver need not possess any concepts involved in the canonical specification of the representational content of his perceptual experiences. To represent the red of the ripe tomato as the color that appears green to him, John does not need to possess the concept of the color green; as he does not even need to possess concepts of causality, experience, or other such ideas.

The second observation is this. The proposal does not assume in any way the act-object model. When I say that John represents red *as* the color that appears green to him, I am not assuming that John first perceives the phenomenal greenness of his visual experience of red, say, as an internal datum to his consciousness, and only later perceives the color itself. Rather, the suggestion proposed here is fully compatible with

the thesis of transparency (Tye) and model of introspection known as displaced perception (Dretske). The proposal is entirely congruent with the thesis of transparency in that it assumes that, on behalf of his visual experience of phenomenal greenness, John does not perceive anything other than what his own visual experience represents, namely, the red color of ripe tomatoes. Moreover, the proposal is also congruent with the displaced-perception model of introspection, that is, with the assumption that John can only become introspectively aware of the phenomenal greenness of his visual experience of ripe tomatoes as the result (output) of a reliable process whose input is the perception of the color of ripe tomatoes itself, considered an external property. In short, to say that John's visual experiences represent the red of ripe tomatoes *as* the color that appears green to him is only to say that the phenomenal greenness is the peculiar way in which John himself perceiving the color red; that is, it is the conscious element by means of which John recognizes that color.

The third observation is this. Even though the proposal assumes that the phenomenal character of perceptual experience is a crucial element in individuation of the perceiver's representational content, it does not need to further assume that the phenomenal character of perceptual experience is nothing more than some ability or know-how to discriminate and recognize instances of properties in the perceptual field. To be sure, phenomenal character enables the perceiver to discriminate and recognize instances of properties. Still, phenomenal character is not manifest in those abilities. Rather, it manifests as a physical property of the brain state that emerges from the brain through causal interaction with instances of environmental properties.

My proposal depends on a critical re-examination of Dretske's design problem (Dretske, 1988, pp. 96-98, 101-103). We begin by reviewing the naturalization of the representational content of experience proposed by Dretske. All events and facts in the world convey information. This comprises the set of all possibilities that a given event or fact excludes. For instance, if it is raining, this fact excludes the possibility that it is snowing or that the sun is shining. We can then characterize the information conveyed by an event or fact as a narrowing of the set of possibilities consistent with the occurrence of the event or fact in question. This is only possible, however, in situations in which events and facts A depend nomically or statistically on other events or facts B in such a way as to make the following counterfactual true: an instance of A would not

have occurred if B had not been instantiated. When a fact or event A covaries nomically or statistically with a fact or event B, the occurrence of A carries information about the occurrence of B, or, more specifically, instances of the carriers of information are generated from the occurrences of B. Thus, for example, if the diameter of the rings of a tree covaries nomically with the age of the tree, then a measure of the diameter of the rings of a certain tree carries information about the age of the tree in question.

However, what particularly interests us are physical events in the brains of sentient creatures and laws or neurobiological statistical covariances prescribing which events should occur under which conditions. The first step in this direction is to recognize the existence of statistical correlations between certain neural states N and certain properties F of proximal stimulation triggering physiological activity in a sensory organ. Thus, a state N could not occur in an individual S, unless S bears some relation to F. Consequently, occurrences of N provide information about the instantiation of F.

Nevertheless, here emerges the traditional problem of disjunction. The measurement of the diameter of the tree rings in question not only covary nomically with age of the tree. It also covaries with the rainfall of the region, with soil nutrients, with the intensity of sunlight, etc. Likewise, the same neurological state N, which covaries nomically with some property of the proximal stimulation, also covaries with many other distal properties of physical events that trigger immediate physiological activity in a sensory organ. In other words, the available information in proximal stimulation significantly underdetermine the distal causes of that stimulation, hence the objects and properties represented in perception and the representational content of perceptual experience. For example, the same firings of retinal sensors are compatible with numerous possible causes. Consequently, any given pattern of information carried by proximal stimulation underdetermines the types of environmental entities perceived by humans and other animals.

Although Dretske does not take the problem of underdetermination into account, one of his favorite examples does illustrate the problem. Suppose that two speedometers are connected to the axle of two vehicles with tires of different diameters. In principle, those speedometers only register the speed of rotation (RPM) of their respective axes (proximal stimulation). In the vehicle with the larger tire, the rotation speed is lower.

This means that, in order to represent the speed of locomotion of their vehicles (distal property) they must be properly “calibrated” (Dretske, 1995); that is, they must acquire the function of indicating the vehicle’s locomotion speed. However, once calibrated, the way each vehicle processes the information or represents the same speed of locomotion has to be different.

Thus, besides the nomological or statistical covariation, the representation of distal properties requires the satisfaction of an additional teleological condition. This additional condition is what Dretske calls *indicator function*. Initially, a neural state N carries information about the instantiation of properties on the basis of the proximal stimulation that triggers the neuronal activity that results in N. However, as we saw, this initial information is compatible with many distal causes. In order to *represent* a particular distal property D, besides covarying nomically with D (via the nomological covariance with F), N has to acquire the function of indicating D specifically (Dretske, 1988, pp. 53-59; 1995, pp. 48-50).

Let us say that Peter is a humanoid inhabitant of the primitive African savannah and that ripe tomatoes there are red and unripe ones are green. Suppose that, because of their red color, proximal stimulation coming from ripe tomatoes under normal environmental conditions triggers a neurophysiological activity in normal individuals like Peter that results in the neural state N. Following the same reasoning, because of its green color, proximal stimulation coming from unripe tomatoes under normal environmental conditions triggers neurophysiological activity in normal individuals such as Peter resulting in the neural state N'. Initially, N covaries statistically with the properties P of proximal stimuli, which are compatible with several distal causes, among them the color red. By the same token, N' covaries statistically with a certain property P' of the proximal stimulus P', which is compatible with several distal causes, among them the color green.

Let us now suppose that ripe tomatoes are edible while unripe ones are indigestible. This makes it indispensable to the survival of the species to which Peter belongs that its members eat the ripe tomatoes (action M) and avoid the unripe ones (action M'). Now, as ripe tomatoes are red and the unripe ones are green, Peter’s neural states N and N' are recruited by natural selection to act as structuring causes of actions M and M', respectively. The fundamental point for Dretske is as follows. The neural

states N and N' are recruited as *structuring* causes of Peter's actions M and M' *because of what they indicate*, namely, the instantiation of the color red and the color green, respectively. Thereafter, neuronal states N and N' not only supply the information that the colors red and green, respectively, are instantiated. They also acquire the function of conveying such information.

As we saw, Dretske identifies the phenomenal character of visual experiences with the colors themselves represented by the experience (phenomenal externalism). Here, however, the problem of inversion returns. Suppose that John is another primitive humanoid inhabitant of the primitive African savanna who possesses an inverted spectrum relative to Peter. While the presence of a ripe tomato, by means of some proximal stimulation, triggers neurophysiological activity in Peter from which results a neural state N, in John, the same ripe tomato triggers another neurophysiological activity from which the neural state N' results.

Therefore, while, in Peter, the neuronal state N conveys the information that the color red is instantiated by ripe tomatoes in his visual field, in John, it is the neural state N' that carries the same information. While, in Peter, the neural state N' conveys the information that the color green is instantiated in his visual field, in John, the neural state N carries this information. Now, to the extent that John (the inverted) is as adapted to his natural environment as Peter, the natural assumption is that, in John, the neural state N', and not the neural state N, is recruited as the structuring cause of John's action M, which is eating ripe tomatoes. Conversely, in John, the neural state N, and not the neural state N', is recruited as the structuring cause of John's action M', which is to avoid eating the fruit.

Thus, Dretske's solution to the design problem requires some repairs. First, the inversion of spectra clearly indicates there are different solutions to the problem of natural design and, most importantly, whatever form it takes, the solution crucially depends on the physical constitution of organisms. If Peter and John are physically distinct individuals, then the distinct neural states in Peter and John will be recruited as structuring causes for the same types of action.

However, the most important addendum is as follows. As the actions in question are conscious, *it is because of the phenomenal characteristics that emerge from the causal interaction between the brain and the distal properties of his environment that*

his neural states are recruited by natural selection to indicate that such properties are being instantiated. It is because of the phenomenal redness of Peter's visual experiences of ripe tomatoes that his neural state N is recruited to represent the instantiation of the color red. Furthermore, it is due to the phenomenal greenness of John's visual experiences of the same ripe tomatoes that his neuronal state N' is recruited to represent the instantiation of the same red. Thus, phenomenal character is a physical property of a brain state that emerges from the brain's causal interaction with instances of environmental properties by virtue of which that brain state is recruited to indicate that such properties are being instantiated. Therefore, in opposition to what Dretske (1995, pp. 82-84) states, if we are to understand the phenomenal character, it is not enough to know which properties the experience in question has the function of indicating. It is also essential to know how the information about their instantiation is being internally processed.

That said, the most plausible suggestion is one that identifies the phenomenal character of perceptual experience with the conscious element in a way of processing information about the instantiation of distal physical properties in their perceptual field peculiar to individuals, groups, or a species. In other words, phenomenal character is the element of conscious experience by which the individual recognizes the instantiations of properties in his perceptual field. The phenomenal character of Peter's experience of the color of ripe tomatoes is the conscious element by means of which he discriminates and recognizes the color red.

5. Solving the Trilemma

As indicated in the introduction, any satisfactory conception of perceptual experience must meet at least three desiderata, namely, the desideratum of representational externalism, the desideratum of representationalism, and the desideratum of phenomenal internalism. The first is the thesis that perceptual states are individuated, in part, on the basis of their representational content and this, in turn, is individuated on the basis of the patterns of relations that the species to which the individual belongs has with different objects, properties, and kinds of natural environments. The second is the thesis that the phenomenal character determines the representational content. Finally,

the third is the thesis of the local supervenience of the phenomenal character on the biological substrate.

The solution proposed here is the only one that can solve the trilemma without abandoning any of the three mentioned desiderata. Furthermore, when compared to available alternatives, it presents itself as the most plausible conception of perceptual experience. First, as the phenomenal character of perceptual experience is understood as a way of processing information about the instantiation of a particular property peculiar to individuals, groups, and species, then the desideratum of phenomenal internalism is trivially satisfied: the phenomenal character of experience locally supervenes on the physical properties of the biological substrate. If John and Peter are individuals who differ in some relevant physical property of their brain or of their visual system, then the way in which the same property of light-reflectance appears to Peter is phenomenally different from the way it appears to John.

Nevertheless, even if the phenomenal character of perceptual experience is nothing but the manner by which the information about the instantiation of properties is internally processed, perceptual states are individuated, in part, by means of their representational content, and this, in turn, by means of the environmental properties those states represent. Therefore, this proposal satisfies the desideratum of representational externalism.

Finally, the proposal also satisfies the desideratum of representationalism. As phenomenal character is one of the crucial elements in the individuation of representational content itself, there can be no difference in representational content without a difference in phenomenal character.

Roberto Horácio de Sá Pereira

Universidade Federal do Rio de Janeiro

robertohsp@gmail.com

References

- Bach, K. (1987) *Thought and Reference*. Oxford: Oxford University Press (revised with postscript 1994).
- Block, N. (2003) 'Mental paint'. IN Hahn, M. and Ramberg, B. (EDS.). *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge, MA: MIT press.
- Chalmers, D. (2004) 'The representational character of experience'. IN Leiter, B. (ED). *The Future for Philosophy*. Oxford: Oxford University Press. Re-impresso em: 2010, *The Character of Consciousness*. Oxford: Oxford University Press.
- Dretske, F. (1981) *Knowledge and the Flow of Information*. Cambridge: MIT Press.
- _____. (1988) *Explaining Behavior*. Cambridge: MIT Press.
- _____. (1995) *Naturalizing the Mind*. Cambridge: MIT Press.
- _____. (1996) 'Phenomenal externalism, or if meanings ain't in the head, where are qualia?'. *Philosophical Issues* 7:143-158.
- Harman, G. (1990) 'The intrinsic quality of experience'. *Philosophical Perspectives*, 4, 31-52.
- Shoemaker, S. (1994) 'Phenomenal character'. *Nous* 28: 21-38.
- Thau, M. (2002) *Consciousness and Cognition*. New York: Oxford University Press.
- Tye, M. (1995) *Ten Problems of Consciousness*. Cambridge: MIT Press.
- _____. (2000) *Consciousness, Color, and Content*. Cambridge: MIT Press.

**NIETZSCHE OU MACINTYRE:
DUAS ALTERNATIVAS À MORALIDADE MODERNA?**

Helder Buenos Aires de Carvalho

José Elielton de Sousa

Resumo

O artigo pretende mostrar que, ao confrontar as alternativas éticas apontadas por Nietzsche e MacIntyre, face às suas análises críticas da condição moral moderna feitas em *Genealogia da moral* e *After Virtue*, respectivamente, podemos encontrar pressupostos teóricos comuns que tornam essas alternativas em diversos pontos convergentes entre si, não simplesmente excludentes, contrariando até mesmo a leitura proposta por MacIntyre desse confronto.

Abstract

This paper aims to show that by confronting the ethical alternatives proposed by Nietzsche and MacIntyre as resulting of their critical analyses of the modern moral condition respectively done in *Genealogy of the Morals* and *After Virtue*, we can find shared theoretical presumptions that point to those alternatives as convergent in many cases and not simply as excluding each other, even in a opposed direction to the MacIntyre's point of view concerning that issue.

0. Introdução

Nietzsche foi um dos primeiros filósofos a tematizar criticamente a condição moral moderna, identificando os possíveis equívocos deste projeto e apontando suas consequências para a moralidade. Em *Genealogia da Moral* (1998), uma de suas obras mais importantes, este filósofo elabora uma das críticas mais radicais à moralidade moderna, da qual somos herdeiros, e aponta a necessidade de uma radical transvaloração de todos os valores.

O filósofo escocês radicado nos Estados Unidos, Alasdair MacIntyre, por sua vez, retoma a discussão iniciada por Nietzsche acerca da condição moral moderna como pano de fundo para seu diagnóstico sobre a moralidade atual. Em *After Virtue [Depois da Virtude]* (2007), sua obra mais polêmica e influente, este filósofo traça uma narrativa de decadência do mundo moderno e das consequências disso para a contemporaneidade, apontando a ética aristotélica das virtudes como saída para a situação caótica em que a moralidade se encontra.

Genealogia da Moral e *After Virtue* são obras que apresentam, assim, um diagnóstico da situação em que se encontra a moral moderna, ao tempo em que

postulam a necessidade de uma medicação urgente e eficaz para os males identificados no diagnóstico. Ambos oferecem uma análise minuciosa da condição moral moderna que tem como pano de fundo a decadência, isto é, o fracasso do projeto iluminista em justificar de forma racional e independente a moralidade, retratando assim uma situação de desmoronamento dos antigos fundamentos de nosso quadro moral, na qual se faz necessário construir propostas éticas capazes de reerguê-los ou substituí-los.

Nesse processo, evidencia-se um movimento constante de aproximação e afastamento entre esses dois autores, tanto no que diz respeito ao diagnóstico feito como nas alternativas apontadas. No entanto, se ambos abordam problemáticas afins munidos de pressupostos comuns, o fazem em registros diferentes. Ao tempo em que partilham a crítica à moralidade moderna, acusando-a de decadente, de fracassada, adotam perspectivas de análise totalmente diferente; aceitam alguns pressupostos morais comuns, mas assumem posturas díspares.

Assim, a questão que se impõe, então, é saber se as alternativas propostas à fracassada moralidade moderna em *Genealogia da Moral* e *After Virtue* são diametralmente opostas entre si *ou* possuem alguma proximidade. O que pretendemos mostrar é que, ao confrontar as alternativas éticas apontadas por Nietzsche e MacIntyre, face às suas análises críticas da condição moral moderna nessas duas obras mencionadas, podemos encontrar pressupostos teóricos comuns que tornam essas alternativas, sob diversos aspectos, convergentes entre si.

1. A crítica à predominância do caráter excessivamente deontológico dos conceitos morais modernos

Nietzsche e MacIntyre compartilham um pressuposto fundamental a respeito da condição moral moderna: a crítica à predominância do caráter excessivamente deontológico dos conceitos morais modernos. Esses autores focalizam o deslocamento efetuado pelos pensadores modernos no que diz respeito à tarefa da moralidade: a pergunta sobre *que tipo de pessoa devo me tornar* foi substituída pela pergunta sobre *quais normas devo obedecer*. As normas tornam-se, assim, o conceito principal da vida moral moderna, uma vez que são elas que determinam, por meio de leis ditadas pela razão, os critérios e os procedimentos que permitem reconhecer uma ação moralmente boa e/ou justa.

Apesar de divergirem quanto às causas e origens desse processo de deontologização excessiva dos conceitos morais – Nietzsche o interpreta como sintomatologia do instinto gregário típico da moral de rebanho, MacIntyre o interpreta como consequência do abandono da teleologia efetuado pelo projeto das luzes –, ambos os autores reconhecem na moralidade kantiana o esforço máximo para tornar os juízos morais objetivos por meio de uma justificação racional dos mesmos¹.

A filosofia moral kantiana se baseia na busca por um princípio supremo da moralidade, um princípio com autoridade racional para ordenar e conduzir as paixões e a si mesmo. Esse princípio, uma vez que se baseia nas leis ditadas pela razão, teria validade universal e seria observado por todas as criaturas racionais. Nessa perspectiva, uma ação é moral quando é regida pelo imperativo categórico, isto é, quando observa a lei ditada pela razão no que ela tem de pura, sem a presença do desejo ou do costume. Esse imperativo é desprovido de qualquer conteúdo moral, mas sua fórmula encerra em si o princípio de todos os deveres². Assim, “o projeto da descoberta de uma justificativa racional para a moralidade é, por conseguinte, o mero projeto de descobrir um exame racional que discrimine as máximas que são a expressão genuína da lei moral” (MacIntyre, AV, 44 [86]).

Para Nietzsche, Kant, assim como todos os demais filósofos que se ocuparam da moral, se equivoca já no ponto de partida: ele preocupou-se tão somente com a fundamentação da moral. Segundo ele, tais tentativas de fundamentação moral não passam de um esforço, ingênuo e carente de autocrítica, de justificação filosófica da moral vigente, uma vez que “o que os filósofos denominavam de ‘fundamentação da moral’, exigindo-a de si, era apenas, vista à luz adequada, uma forma erudita de ingênua fé na moral dominante, um novo modo de *expressá-la*, e portanto um fato no interior de uma determinada moralidade” (BM, 186). Ao considerar a moral como “dada”, como algo natural, a tentativa filosófica de fundamentar a mesma expressa somente uma determinada espécie de moral, a saber, a moral vigente, aquela de uma determinada cultura, não tendo, assim, validade universal como querem os teóricos morais iluministas, especialmente Kant.

¹ Quanto à teoria moral kantiana, nos limitamos aqui somente às leituras que estes autores fazem da mesma, sem entrar no debate acerca da correção dessas leituras.

² Ver KANT, I. *Fundamentação da Metafísica dos Costumes*. Tradução de Paulo Quintana. 2ª edição. São Paulo: Abril Cultural, 1980.

Nesse sentido, na perspectiva nietzschiana, querer que um juízo moral seja válido universalmente, tal como pretende Kant, é, no mínimo, egotismo exacerbado (Cf. GC, 335). Fixar o próprio ponto de vista como o único válido, universalizá-lo, absolutizá-lo, é não apenas desprezar a existência de outros ângulos de visão, mas principalmente desconsiderar o fato de que um juízo moral tem uma história. Para Nietzsche, o fato de que se “sinta algo como *certo*, pode ser devido a que você nunca tenha meditado sobre si e tenha cegamente acolhido o que desde a infância lhe foi designado como certo” (GC, 335. Grifo nosso). Assim, a ideia de “consciência do dever” tem uma longa história e uma variedade de formas atrás de si.

Na *Genealogia da moral*, essa consciência do dever está relacionada ao amplo processo de domesticação do animal homem, que tem por objetivo torná-lo “até certo ponto necessário, uniforme, igual entre iguais, constante e, portanto, confiável” (GM, II, 2). Ela faria parte da longa história da origem da responsabilidade, isto é, da capacidade de fazer promessas, de se comprometer, de responder por si. Associando o conceito de dever ao de culpa, Nietzsche afirma que esses têm sua origem “no conceito material de ‘dívida’. A internalização da consciência do dever, da obrigação, só foi possível graças ao doloroso e cruel processo de criação da memória no animal homem, pois “apenas o que não cessa de *causar dor* fica na memória” (GM, II, 3).

Na visão nietzschiana, o moralmente correto é, assim, fruto da longa história da crueldade. Não apenas isso, Nietzsche vai mais longe: “e não poderíamos acrescentar que no fundo esse mundo jamais perdeu inteiramente um certo odor de sangue e tortura? (Nem mesmo no velho Kant: o imperativo categórico cheira a crueldade...)” (GM, II, 6). Para Nietzsche, portanto, a sacralidade do dever, a obediência às leis tem uma história, uma história largamente banhada de sangue.

MacIntyre, por sua vez, tal como Nietzsche, critica o caráter não-histórico do imperativo categórico kantiano e sua pretensão de universalidade e objetividade, mas o faz de uma perspectiva diferente. MacIntyre vê Kant como expressão máxima do projeto de justificar de forma racional e independente a moralidade, iniciado em fins do século XVII e no século XVIII, no qual a diferenciação da moral do teológico, do jurídico e do estético tornou-se doutrina aceita por todos. Na leitura macintyreana, moralidade, para tal projeto, é o nome da esfera especial em que as *normas* de conduta

que não são teológicas, estéticas e nem jurídicas conquistam espaço cultural próprio, tornando-se, assim, predominantes na apreciação moral (Cf. AV, 38-39 [77-78]).

Nesse contexto, o projeto moral kantiano consistiria, segundo MacIntyre, em descobrir um exame racional que discrimine as máximas que são a expressão genuína da lei moral, máximas essas que possuem um caráter categórico incondicional ditado pela razão. Para MacIntyre, Kant compartilha com os demais teóricos morais iluministas a concordância acerca do caráter da moralidade e do que seria a justificativa racional da mesma. A filosofia moral kantiana expressa, segundo ele, uma concepção de natureza humana, por um lado, e, por outro, uma explicação e uma justificação das normas morais como sendo as regras que se poderia esperar que um possuidor de tal natureza humana aceitasse (Cf. AV, 52 [98]). De acordo com MacIntyre, Kant considera o caráter universal e categórico de certas normas da moralidade a característica mais importante da natureza humana, pois pertence à essência da razão estabelecer princípios que sejam universais, categóricos e internamente compatíveis.

Segundo MacIntyre, porém, qualquer projeto dessa forma estava fadado ao fracasso, pois se assenta numa mutilação do esquema moral aristotélico-medieval ao eliminar-lhe a concepção teleológica da qual é portadora (AV, 52 [99]). Para ele, a concepção de natureza humana compartilhada pelos iluministas terminou por se contrapor aos preceitos morais herdados da tradição (preceitos esses que seus teóricos pretendiam justificar racionalmente), pois o elemento mediador entre eles – o *télos* do agir humano – desapareceu como um dado racional. A falta de percepção desse fato impediu-os de reconhecer o caráter impossível de sua tarefa de fundamentar racionalmente e de forma independente a moralidade, pois passaram a manejar um vocabulário moral descontextualizado do esquema conceitual que lhes provia o sentido. Sem o *télos*, as normas morais entram em choque com a natureza humana em sua facticidade, não havendo mediação, tornando-se disponíveis para quaisquer usos.

MacIntyre analisa esse fracasso do empreendimento moral kantiano se detendo em duas teses, consideradas por ele, tanto enganosamente simples como fundamentais à filosofia moral de Kant. Para MacIntyre, a pretensa universalidade do exame racional kantiano não se sustenta, pois “Kant não duvidou nem por um instante que as máximas que aprendeu com os próprios pais virtuosos fossem as que deveriam ser justificadas por um exame racional” (AV, 44 [86]). Tal exame serviria, portanto, somente para

discriminar as máximas que não corresponderiam à lei moral e justificar racionalmente as que Kant partilhava e reconhecia como boas. O conteúdo da moralidade kantiana seria, assim, o de uma determinada moralidade, a saber, da moralidade cristã herdada, não possuindo a validade universal pretendida.

Quanto à segunda tese, MacIntyre considera que Kant, ao adotar um conceito de razão prática que não emprega qualquer critério externo a ela própria, não apelando assim aos dados da experiência, abre espaço para que outras máximas não-morais e imorais sejam válidas tanto quanto as que ele deseja fundamentar e sustentar. Segundo MacIntyre, Kant com certeza recusaria essas máximas, pois ele acreditava que seu exame de possibilidade coerente de universalização dessas máximas tinha um conteúdo moral determinado e consistente que as excluía, uma vez que sua formulação do imperativo categórico equivaleria a uma formulação completamente diferente delas (Cf. AV, 45-46 [88-89]). No entanto, para MacIntyre, Kant não oferece bons motivos para defender tal posição, permitindo assim que tais máximas fossem universalizadas sem incoerência nenhuma e os preceitos morais do próprio Kant situados no mesmo nível delas. O teste de universalização proposto se revelou estritamente formal, incapaz de filtrar máximas imorais e não-morais.

Para MacIntyre, portanto, a tentativa para fundamentar o que Kant crê serem as máximas da moralidade no que ele acreditava ser a razão fracassou (AV, 46-47 [90]). Desde então, “o projeto de oferecer uma justificativa racional da moralidade fracassara decisivamente; e, daquele ponto em diante, a moralidade de nossa cultura predecessora – e, por conseguinte, da nossa própria cultura – carecia de fundamentos lógicos ou justificativas públicas e compartilhadas” (AV, 49-50 [96]).

Nietzsche e MacIntyre reconhecem assim, cada um a seu modo, os limites do projeto moral moderno no que diz respeito às suas pretensões de universalidade e objetividade. Ambos os autores apontam o caráter subjetivo dos juízos morais modernos, mostrando a gênese histórica e contextual da razão iluminista que se pretendia autônoma e objetiva e, por conseguinte, trans-histórica. Esse fato fornece, assim, o ponto de partida para as alternativas éticas elaboradas por esses autores.

2. A ideia de florescimento humano

A crítica ao caráter deontológico dos juízos morais modernos feita por Nietzsche e MacIntyre impôs-lhes a necessidade de iniciar suas investigações acerca da moralidade de forma bem diferente daquela dos pensadores iluministas. Ambos constroem suas teorias éticas com base, em última instância, em alguma concepção do bem humano, passando a focalizar o centro irradiador do valor moral na natureza do próprio agente. Em Nietzsche, essa concepção de bem último está vinculada à noção de *grande saúde*, o caminho que possibilita ao homem atingir o supremo brilho e potência de que ele é capaz. Em MacIntyre, por sua vez, essa concepção de bem último está relacionada à noção de teleologia, à busca pela realização do bem propriamente humano. Ambas as teorias podem ser consideradas, assim, modelos éticos de caráter eudaimonístico, na medida em que consideram a noção de florescimento humano como fundamental para o empreendimento ético – essa ideia de florescimento permite pensarmos os universos éticos de Nietzsche e MacIntyre nas antípodas do deontologismo moderno.

2.1. A grande saúde *nietzschiana*

O problema mais geral da filosofia de Nietzsche, a crítica dos valores morais, remete imediatamente ao problema do valor da vida em relação às possibilidades de aperfeiçoamento, de florescimento do tipo homem. O florescimento humano não é um conceito elaborado de forma explícita por Nietzsche. Ele aparece em sua obra diluído nas diversas metáforas acerca da condição humana, podendo ser encontrado de forma mais direta em sua teoria da vontade de poder³.

Segundo Nietzsche, há uma vontade de poder operante em todo acontecer, uma atividade constante de subjugamento e assenhoreamento na qual o “sentido” e a “finalidade” anteriores são necessariamente obscurecidos ou obliterados (GM, II, 12). Todo acontecimento é uma luta pelo poder, pela expansão, na qual se sobressai o mais forte naquele momento; é uma sucessão ininterrupta de um processo de subjugamento, que sempre aparece em forma de vontade e via de maior poder e se impõe à custa de

³ Considerado central para compreensão da obra nietzschiana pela maioria dos comentadores, esse conceito de vontade de poder é, contudo, objeto das mais diversas interpretações. Marton afirma, em nota, que, além de existirem pressupostos suficientes para justificarem-nas, uma dificuldade técnica contribui para tanto: o fato de tal conceito estar presente, sobretudo, nos fragmentos póstumos, redigidos entre o verão de 1883 e os primeiros dias de janeiro de 1889, só recentemente publicados na íntegra (Cf. 2000, p.236).

inúmeros poderes menores. Apesar de não se limitar a isso, a vontade de poder pode ser compreendida como uma força central na qual repousam todas as atividades humanas. Desse modo, ela é, primeiramente e antes de tudo, o conceito-chave de uma hipótese psicológica por meio da qual Nietzsche vincula as questões acerca da cultura às relativas à natureza, tematizando, assim, as ações do homem tanto como indivíduo social quanto como espécie biológica⁴. O conceito de vontade de poder vincula, portanto, as questões relativas ao indivíduo às suas possibilidades de expansão, de florescimento.

A concepção de florescimento em Nietzsche lhe é muito peculiar. Ele não o compreende como um progresso rumo a um *télos*, pois para ele “o ‘desenvolvimento’ de uma coisa, um uso, um órgão, é tudo menos o seu *progressus* em direção a uma meta, menos ainda um *progressus* lógico e rápido, obtido com um dispêndio mínimo de forças” (GM, II, 12). De acordo com Nietzsche, o verdadeiro progresso consiste numa “sucessão de processos de subjugamento que ocorrem, mais ou menos profundos, mais ou menos interdependentes, juntamente com as resistências que cada vez encontram, as metamorfoses tentadas com o fim de defesa e reação, e também os resultados de ações contrárias bem-sucedidas”⁵ (GM, II, 12). Ele o entende, portanto, como um constante vir-a-ser, uma ação contínua de destruição e recriação com vistas a um poder maior, isto é, como um processo contínuo de reinterpretação, de redirecionamento para novos fins.

Para Nietzsche, o fenômeno moral, assim como todas as outras manifestações da vontade de poder, também passa por esse processo de expansão, de florescimento. Ela pode significar tanto uma grande saúde, como no caso da moral aristocrática, quanto uma grave doença, como no caso da moral escrava, dominante na modernidade. Desse

⁴ Apesar de reconhecer outras interpretações possíveis para o conceito nietzschiano de vontade de poder, adotamos aqui a interpretação feita por Walter Kaufmann em sua obra *Nietzsche, philosopher, psychologist, antichrist*, para quem esse conceito permite a Nietzsche insistir na relação agora renovada entre natureza e cultura. Para Kaufmann, a concepção nietzschiana do conceito de vontade de poder não é metafísica, nem no sentido heideggeriano nem no sentido positivista, sendo antes de mais nada uma hipótese psicológica (Cf. 1974, p. 175-177).

⁵ Em *O Anticristo*, Nietzsche afirmaria que segundo o que ele entende como progresso, a humanidade certamente não representa uma evolução em direção a algo melhor, mais forte ou mais elevado. Este “progresso” é apenas uma ideia moderna, ou seja, uma ideia falsa. O processo da evolução não significa necessariamente elevação, melhoramento, fortalecimento. Nietzsche concorda que é bem verdade que ela tem sucesso em casos isolados e individuais em várias partes da terra e sob as mais variadas culturas, e nesses casos certamente se manifesta um tipo superior; um tipo que, comparado ao resto da humanidade, parece uma espécie de super-homem. Contudo, segundo ele, tais golpes de sorte sempre foram possíveis e, talvez, sempre serão. Até mesmo raças inteiras, tribos e nações podem ocasionalmente representar tais ditos acidentes (Cf. AC, IV).

modo, na medida em que o filósofo admite a possibilidade de um tipo de homem superior (mais especificamente, um super-homem), a ideia de florescimento em Nietzsche pode ser compreendida, dentre outras formas, como a busca por uma *grande saúde*⁶, pois ela é propriamente o pressuposto desse tipo superior (Cf. EH, IX, 2).

O conceito de grande saúde torna-se operante na obra nietzschiana a partir de 1886, quando ele lança a segunda e definitiva edição de *A gaia ciência* (1882) acrescida de um prefácio, um quinto capítulo, no qual o filósofo enuncia pela primeira vez tal conceito, e um apêndice com mais poemas. Esse conceito pertence, portanto, ao chamado período de transvaloração de todos os valores, com o qual partilha o tom e a densidade:

Nós, os novos, sem nome, de difícil compreensão, nós, rebentos prematuros de um futuro ainda não provado, nós precisamos, para um novo fim, também de um novo meio, ou seja, de uma nova saúde, mais forte[,] alerta alegre[,] firme[,] audaz[,] que todas as saúdes até agora. Aquele cuja alma anseia haver experimentado o inteiro compasso dos valores e desejos até hoje existentes e haver navegado as praias todas desse “Mediterrâneo” ideal, aquele que quer, mediante as aventuras da vivência mais sua, saber como sente um descobridor e conquistador do ideal, e também um artista, um santo, um legislador, um sábio, um erudito, um devoto, um adivinho, um divino excêntrico de outrora: para isso necessita mais e antes de tudo uma coisa, *a grande saúde* – aquela que não apenas se tem, mas constantemente se adquire e é preciso adquirir, pois sempre de novo se abandona e é preciso abandonar (GC, 382).

Em suas exposições sobre a grande saúde, Nietzsche a apresenta como condição de possibilidade de um novo ideal, um ideal prodigioso, tentador, pleno de perigos, para o qual o mais elevado, aquilo em que o povo encontra naturalmente sua medida de valor, já não significaria senão perigo, declínio, rebaixamento. O filósofo aspira um tipo que seja capaz de, em um tempo vindouro, fazer diferentes ensaios, experimentar outras formas de valorar, criar novas “tábuas de valores”, que não mais expressem um profundo mal-estar com os processos efetivos. Espera a vinda de homens dotados de grande saúde que possam livrar a efetividade da “maldição” deposta sobre ela (Cf. GM, II, 24). A grande saúde implica, portanto, algo bem diferente desse animal doente que é o homem moderno; ela implica mesmo um tipo superior, “o homem redentor, o homem

⁶ Dentre as diversas formas que podemos utilizar para abordar tal questão na filosofia nietzschiana, escolhemos o conceito de *grande saúde*, tanto por ser a metáfora que melhor se ajusta aos fins e à temática que propomos no presente trabalho, quanto por ser um conceito menos problemático – ele não tem, por exemplo, o peso metafísico que tem o conceito de “super-homem” – e pouco utilizado pelos especialistas em Nietzsche.

do grande amor e do grande desprezo, o espírito criador cuja força impulsora afastará sempre de toda transcendência e toda insignificância” (GM, II, 24).

Essa ideia de um tipo superior, em Nietzsche, está vinculada ao seu conceito de *Übermensch* (super-homem), um conceito que, com exceção do aforismo 143 de *A gaia ciência*, em que faz uma aparição fugaz, aparece somente em *Assim falou Zaratustra* (1885), mas que é (juntamente com os conceitos de vontade de poder e eterno retorno) fundamental para a compreensão do conjunto da obra nietzschiana⁷. Eis, em síntese, como Nietzsche o apresenta: “*Eu vos ensino o super-homem*”⁸. O homem é algo que deve ser superado. Que fizestes para superá-lo?” (ZA, Prólogo, 3). Com o conceito de super-homem, Nietzsche começa a delinear os traços de um ideal inverso ao predominante na modernidade e com isso passamos do aspecto crítico de sua filosofia ao aspecto positivo, afirmativo, a parte que diz “Sim”. Assim, longe de representar uma nova espécie biológica⁹, o super-homem é propriamente o homem de uma nova cultura, uma cultura em tudo superior à *décadence* moderna. Ele traduz uma nova forma de pensar, sentir e avaliar, expressa por meio de um *pathos* afirmativo por excelência.

Assim sendo, o super-homem é um criador, e é no reconhecimento e exercício desse poder único de criação que reside a sua saúde. A afirmação dessa saúde se traduz por meio da aceitação do caráter temporal e finito da condição humana, da precariedade e das benesses dessa condição finita:

⁷ Segundo Kaufmann, Nietzsche não foi o primeiro a usar o termo *Übermensch*, ele remonta à antiguidade clássica, especialmente os escritos de Luciano, retórico e filósofo, que se tornou conhecido pelos seus diálogos satíricos e histórias fantásticas. Nietzsche, como filólogo clássico, estudou Luciano, sobre quem se referiu com frequência no seu *Philologica*. Ainda de acordo com Kaufmann, o termo já tinha sido usado também por Heinrich Müller (*Geistliche Erquickungsstunden*, 1664), por J. G. Herder, por Jean Paul - e por Goethe, num poema (*Zueignung*) e no Fausto (parte I, linha 490), onde um espírito zomba de um assustado Fausto que o tinha evocado, e o chama de *Übermenschen*. Kaufmann chama a atenção, porém, para o fato de que o sentido que Nietzsche mais tarde deu a esse termo é inteiramente distinto dos predecessores acima citados (Cf. 1974, p. 307-308).

⁸ Reiteremos aqui as observações feitas por Paulo César de Souza, em uma nota de *Ecce Homo*, acerca da tradução para o português do termo *Übermensch*. Segundo ele, apesar das restrições que Rubens Rodrigues Torres Filhos faz à tradução do mesmo por “super-homem” (ver comentários no volume dos “Pensadores”, p. 236, 313, 383), propondo que se use, ao invés disso, o termo “além-do-homem”, esse deixaria a desejar formalmente – o que se torna claro quando no texto é aproximada ao adjetivo *übermenschlich* (sobre-humano). De acordo com ele, não soa bem em português dizer “sobre-homem” ou “supra-homem”, restando-nos somente o contentamento – provisório, talvez – com “super-homem” (Cf. EH, Notas, p. 116).

⁹ Contra a ideia de associar o super-homem nietzschiano a uma espécie de fenômeno biológico, Marton afirma que “não se trata de um tipo biológico superior ou de uma nova espécie engendrada pela seleção natural, mas de quem organiza o caos de suas paixões e integra numa totalidade cada traço de seu caráter” (1994, p. 19).

O que há de grande, no homem, é ser ponte, e não meta: o que pode amar-se, no homem, é ser *transição* e um *ocaso*. Amo os que não sabem viver senão no ocaso, porque estão a caminho do outro lado [...] Amo aquele que prodigaliza a sua alma, não quer que lhe agradeçam e nada devolve: pois é sempre dadivoso e não quer conservar-se [...] Amo aquele que atira palavras de ouro precedendo seus atos e, ainda assim, cumpre sempre mais do que promete: pois quer o seu ocaso. Amo aquele que justifica os seres futuros e redime os passados: porque quer perecer dos presentes. [...] Amo aquele cuja alma é tão transbordante, que se esquece de si mesmo e que todas as coisas estão nele: assim, todas as coisas tornam-se o seu ocaso. Amo aquele cujo espírito e coração são livres: assim, nele, a cabeça é apenas uma víscera do coração, mas o coração o arrasta para o ocaso (ZA, Prólogo, 4).

Deste modo, o super-homem é, antes de tudo, aquele que vence a si mesmo; é alguém capaz de criar seu próprio futuro, de criar-se a si mesmo. Cabe ressaltar, contudo, que Nietzsche sabe das exigências desse tipo superior, sabe que a tarefa de autocriação que ele exige é uma coisa para poucos, daí seu caráter aristocrático. Para o filósofo, o que esse tipo superior, esse tipo saudável, pode e deve, jamais poderiam poder e dever os enfermos (Cf. GM, III, 14). O “*páthos* da distância” visa, portanto, manter esse tipo superior saudável, preservá-lo dos doentes, dos impossibilitados de responder por si mesmos.

Nesse sentido, a ideia de super-homem implica uma relação ética entre iguais: “companheiros, procura o criador, e não cadáveres; nem, tampouco, rebanhos e crentes. Participantes na criação, procura o criador, que escrevam novos valores em novas tábuas [...] Que tem ele a ver com rebanhos, pastores e cadáveres!” (ZA, Prólogo, 9). Para Nietzsche, falar de companheiros é falar de uma elite, de indivíduos soberanos, com vontade própria, duradoura e independente, capazes de seguirem-se a si mesmos, de inscrever novas tábuas de valores. Trata-se, portanto, de um aristocratismo ético e não de um solipsismo moral, do qual muitas vezes Nietzsche é acusado.

2.2. O télos *macintyriano*

A pretensão macintyriana de construir uma alternativa ética à condição moral moderna está vinculada à sua concepção de homem e à possibilidade que o mesmo tem de se aperfeiçoar, florescer enquanto tal. Apesar de ser desenvolvido de forma explícita somente em *Dependent Rational Animals* (1999)¹⁰, é o conceito de florescimento

¹⁰ Nessa obra, MacIntyre pretende continuar e também corrigir alguns aspectos de suas obras anteriores, notadamente a recusa da biologia metafísica de Aristóteles, que foi categoricamente rejeitada durante sua reabilitação da ética aristotélica das virtudes efetuada em *Depois da virtude*. Com a correção desses

humano que permite a MacIntyre, já em *Depois da Virtude*, esclarecer melhor suas reflexões anteriores acerca de sua concepção de homem e fundamentar sua proposta de retomar uma perspectiva ética teleológica¹¹.

Para MacIntyre, a capacidade de florescimento não é uma característica exclusivamente humana. O florescimento é um conceito aplicável também a membros de diferentes espécies de animais e plantas (DRA, 64). De acordo com MacIntyre, do que necessita um membro de uma espécie para florescer é desenvolver as faculdades características que possui enquanto membro dessa espécie (DRA, 64). Nesse sentido, o florescimento acontece em virtude da posse de determinadas características próprias de cada espécie de seres vivos, características essas que os definem enquanto tal. O florescimento do ser humano acontece, portanto, na medida em que este desenvolve suas faculdades racionais próprias, isto é, quando vive como teria que viver o ser humano. Desse modo, a ideia de florescimento humano pressupõe um conceito de bem, pois “florescer se traduz como *eu zen e bene vivere*” (DRA, 65).

Segundo MacIntyre, essa definição de bem envolve uma investigação conceitual e valorativa, uma vez que existem, pelo menos, três diferentes formas de atribuição do bem. Em uma primeira forma, o bem é compreendido como um meio para se atingir outro bem. Outra forma de atribuição do bem é compreendê-lo como qualidades intrínsecas de determinadas práticas socialmente estabelecidas, as quais são usadas como critérios de avaliação dos agentes nelas envolvidos. O juízo sobre o que é melhor para a vida de um indivíduo ou comunidade – a melhor maneira de ordenar seus bens –, não apenas enquanto agente que participa de uma ou outra atividade em uma ou outra comunidade, mas também enquanto ser humano, ilustra a terceira forma de atribuição do bem. Este tipo de juízo é um juízo sobre o florescimento humano, sua finalidade última, seu *télos* (DRA, 65-67).

Para MacIntyre, no entanto, distinguir e aplicar o bem são coisas que o indivíduo aprende somente no interior de uma *prática* social. Usando esse termo de uma forma não-convencional, MacIntyre explica que prática é

“erros”, MacIntyre elabora um conceito amplo de florescimento, estendendo-o a outras espécies de animais não-humanos e plantas.

¹¹ Izquierdo comenta que este conceito de florescimento humano está presente de tal modo em *Depois da Virtude* que certamente se pode falar de uma *antropologia do florescimento* nessa obra (Cf. 2007, p. 69-70).

Qualquer forma coerente e complexa de atividade humana cooperativa, socialmente estabelecida, por meio da qual os bens internos a essa forma de atividade são realizados durante a tentativa de alcançar os padrões de excelência apropriados para tal forma de atividade, e parcialmente definidores, tendo como consequência a ampliação sistemática dos poderes humanos para alcançar tal excelência, e dos conceitos humanos dos fins e dos bens envolvidos (AV, 187 [316]).

A definição macintyriana de prática implica dois termos fundamentais: a noção de bens externos e bens internos a uma prática. Segundo MacIntyre, “é característica dos bens externos que, quando conquistados, sempre são de propriedade e posse de alguém. Além disso, são tais que quanto mais se tem, menos há para outras pessoas” (AV, 190 [320-321]). Os bens externos são conquistados por acidente das circunstâncias sociais, as quais permitem que haja modos alternativos de alcançá-los. Eles têm, portanto, uma relação direta com dinheiro, fama, poder, relação essa que pressupõe uma concorrência em que deve haver tanto vencedores como derrotados.

Os bens internos, por sua vez, são assim chamados por dois motivos: primeiramente só podemos especificá-los dentro de uma determinada prática e por meio de exemplos dessa prática; e, em segundo lugar, porque só podem ser identificados e reconhecidos pela experiência de participar da prática em questão (AV, 188-189 [317-318]). Eles implicam padrões de excelência e obediência a normas, pois julgar esse tipo de bem requer um tipo de competência que só se adquire praticando uma determinada atividade ou aprendendo sistematicamente o que uma determinada prática exige. Os bens internos são, portanto, consequência da competição pela excelência, mas é característica deles que sua conquista seja boa para toda a comunidade que participa da prática (AV, 190-191 [321]).

Contudo, como há uma multiplicidade de práticas, muitas vezes formuladoras de bens não necessariamente convergentes, conflituosos, é necessário que se tenha uma concepção unitária do bem humano, levando em consideração sua vida como um todo, de tal forma que se possua um critério de hierarquização dos bens vinculados às práticas. MacIntyre concebe então a vida do indivíduo como uma narrativa, isto é, uma unidade narrativa, na qual nascimento, vida e morte fazem parte de uma mesma narrativa com começo, meio e fim (AV, 205 [345]).

O conceito macintyriano de identidade narrativa pressupõe dois aspectos importantíssimos: inteligibilidade e responsabilidade (AV, 217-119 [365-367]). Para MacIntyre, ser o sujeito de uma narrativa que vai do nascimento à morte é ser

responsável pelos atos e experiências que compõem uma vida narrável. O indivíduo deve ser capaz de explicar os atos e experiências que compõem sua narrativa de vida, de modo a torná-la um todo inteligível tanto para si mesmo quanto para os demais. A ideia de inteligibilidade é, assim, o elo conceitual entre a ideia de ação individual e a de narratividade.

Outro aspecto do conceito de identidade narrativa, que é correlato ao de inteligibilidade, é o de responsabilidade¹². Uma vez que a narrativa de qualquer vida faz parte de um conjunto interligado de outras narrativas, não sou somente responsável, sou alguém que pode sempre pedir explicação aos outros, questioná-los. A possibilidade de justificação, seja pessoal seja social, dos atos é constituinte essencial de uma narrativa. Para MacIntyre, portanto, sem a responsabilidade do *eu*, cujas sequências de eventos constituem as narrativas, as ações perderiam seu caráter inteligível. Assim, é possível dizer que a unidade da vida humana consiste em uma narrativa incorporada em uma única vida, e que o bem desta vida é obtido pela resposta de como posso viver esta unidade e levá-la à sua completude. Essa busca de unidade implica, portanto, um *télos* que vai sendo desenhado pelas respostas dadas às dificuldades que vão aparecendo.

Mas não se trata de uma busca individual, uma vez que nossas identidades são socialmente constituídas. MacIntyre acredita que não é possível procurar o bem somente *qua* indivíduo, pois o que é viver a vida boa concretamente varia de uma circunstância para outra e, mais que isso, todos tratamos nossas circunstâncias como portadoras de uma identidade social particular. Nesse sentido, a procura individual do próprio bem é, em geral e caracteristicamente, realizada dentro de um contexto definido pelas tradições das quais a vida do indivíduo faz parte, e isso é verdadeiro com relação aos bens internos às práticas e também aos bens de uma única vida (AV, 222 [373-374])¹³. Longe da conotação depreciativa que o termo tradição adquiriu modernamente, MacIntyre o entende como

Uma argumentação racional, desenvolvida ao longo do tempo, na qual certos acordos fundamentais são definidos e redefinidos em termos de dois tipos de conflitos: os

¹² Em *Three rival versions of moral enquiry*, MacIntyre define responsabilidade como uma exigência para tornar inteligível para si mesmo e para as demais pessoas que participam da mesma comunidade, o que eu estava fazendo ao me comportar assim em determinada ocasião particular e se, em algum tempo futuro, estarei preparado para voltar a valorar minhas ações à luz dos juízos propostos por outros (TRV, 197).

¹³ Para uma visão ampla e detalhada do conceito macintyreano de tradição, ver CARVALHO (1999).

conflitos com os críticos e inimigos externos à tradição, que rejeitam todos ou pelo menos partes essenciais dos acordos fundamentais, e os debates internos, interpretativos, através dos quais o significado e a razão dos acordos fundamentais são expressos e por cujo progresso uma tradição é constituída (JR, 23).

As tradições são, assim, um conjunto de práticas formadas e transmitidas através de gerações, as quais definem o contexto no qual a busca individual pelo bem-viver acontece. Desse modo, para MacIntyre, a busca individual do próprio bem é, em geral e caracteristicamente, realizada dentro de um contexto definido pelas tradições das quais a vida do indivíduo faz parte, e isso é verdadeiro com relação aos bens internos às práticas e também aos bens de uma única vida. Ele propõe, portanto, uma concepção narrativa do sujeito em busca do pleno florescimento que capacita à realização de um plano de vida, acompanhada de uma ênfase contínua na importância da tradição, constituída precisamente de um conjunto de práticas formadas e transmitidas através de gerações e que definem o contexto no qual se desenrola a busca individual do bem-viver

A noção macintyriana de florescimento humano, entendida como uma busca pela realização de seu *télos*, que é o bem humano propriamente, está relacionada à sua noção de natureza humana¹⁴. MacIntyre adota uma concepção de natureza humana segundo a qual os seres humanos têm uma natureza específica que determina, de forma apropriada, seus fins e metas. Trata-se de um esquema presente na ética grega clássica, especialmente na obra aristotélica, segundo o qual há, por um lado, a natureza humana tal como é, em estado bruto, movida por desejos e paixões ainda não instruídos, a natureza humana tal como poderia ser se realizasse sua essência ou seu verdadeiro fim, por outro lado, e um conjunto de preceitos éticos que possibilitam a passagem de um estado a outro da natureza humana (AV, 52 [99]).

Temos, então, um esquema tripartite no qual a natureza humana em estado natural é inicialmente discrepante e discordante dos preceitos da ética e precisa ser transformada pelo ensino e experiência da razão prática em natureza humana como ela poderia ser se realizasse o seu *télos* (AV, 52-53 [100]). MacIntyre concebe o homem como um ser que tem uma natureza própria, a qual lhe confere o fim a que se destina,

¹⁴ Sobre essa relação entre a ideia de florescimento e o conceito macintyriano de natureza humana, Izquierdo comenta que, embora MacIntyre não adote explicitamente essa relação, ela está presente de forma implícita em sua obra. Para Izquierdo, sem a noção de uma natureza essencial, em uma ética aristotélica tal como a que MacIntyre propõe, caem as noções de *bem* e do *bom e melhor* para os membros de uma classe específica que compartilham tal natureza, e, portanto, se derruba o esquema moral que relaciona o ser (a natureza) com o bem e com o fim (Cf. 2007, p. 116-117).

seu *télos*, a ser realizado ao longo de seu processo de florescimento. O *télos* da vida de um indivíduo é, portanto, seu florescimento (DRA, 12-13)¹⁵.

3. Educação moral e virtudes

Nietzsche e MacIntyre, ao fundamentarem suas propostas éticas na ideia de florescimento humano, deixam transparecer um aspecto comum importante acerca da condição humana: o homem não é compreendido como um ser moral por natureza, mas como alguém que precisa ser educado para a moralidade. Ambos os autores compartilham a ideia de um projeto educativo que visa ao florescimento de um tipo humano, no mínimo, contrário ao homem moderno. Assim, apesar das diferenças teóricas que existem entre eles, tanto a versão nietzschiana de uma busca por uma *grande saúde* como a versão macintyreana de uma busca pelo *télos* propriamente humano, pressupõem um *processo educacional* vinculado à condição moral humana.

O projeto educativo nietzschiano está relacionado à sua intenção de elevar o homem a um estado superior, por meio de sua contínua superação deste homem por ele mesmo, pois, para o filósofo, a educação autêntica é uma atividade constante de construção, de criação de si mesmo (Cf. Co. Ext. II, 6). A defesa de uma educação aristocrática e elitista, baseada não na superioridade de uma raça ou de uma classe social, mas na supremacia da vontade de poder e da inteligência, radica na concepção que Nietzsche tem do homem superior. Delineado explicitamente já nas *Considerações Extemporâneas* (1873-1876), esse projeto educativo sofre algumas transformações no decorrer de suas obras subsequentes, culminando, durante o projeto de transvaloração, na ideia de criação desse tipo superior¹⁶.

Já o projeto educativo macintyriano tem por objetivo a retomada de um conceito funcional de homem, compreendido como ser que tem uma natureza essencial e uma finalidade ou função essencial, um conceito que remete aos teóricos da tradição grega clássica, anteriores mesmo a Aristóteles (Cf. AV, 58 [109]). Delineado em *Depois da virtude*, esse projeto se fundamenta no já mencionado esquema tríplice presente na ética grega clássica, especialmente na obra aristotélica, que consiste em três elementos:

¹⁵ Em AV esse *télos* tem uma ancoragem nas práticas, ou seja, de natureza pragmaticamente estabelecida; já a partir de TRV, MacIntyre dá efetivamente um giro tomista em sua filosofia e ancora esse *télos* em uma perspectiva metafísica.

¹⁶ Sobre os diversos sentidos que o projeto educacional nietzschiano adquire no decorrer de suas obras, ver DIAS, R. M. *Nietzsche educador*. São Paulo: Scipione, 1993; FREZZATTI JUNIOR (2006).

a natureza humana tal como é, em estado bruto, movida por desejos e paixões ainda não instruídos; a natureza humana tal como poderia ser se o ser humano realizasse sua essência ou seu verdadeiro fim; e um conjunto de preceitos éticos que possibilitam a passagem de um estado a outro da natureza humana. Essa concepção de natureza humana pressupõe, assim, a passagem de seu estado natural (não-instruído) para um estado humano educado (instruído), durante a qual o homem vai descobrindo sua própria essência.

Assim, para ambos os filósofos o processo educacional deve submeter a natureza humana a uma rígida disciplina, por meio do cultivo de determinadas virtudes sem as quais tal processo educacional não poderia ser levado a cabo.

3.1. *As virtudes aristocráticas*

Dada a preocupação nietzschiana com o cultivo de tipos superiores, mais plenos de valores, mais merecedores de vida, durante o projeto de transvaloração, essa preocupação se volta para seu aparecimento não mais como um acaso feliz, como uma exceção – como aconteceu no passado –, mas como resultado da vontade (Cf. AC, III). O filósofo propõe que não dependamos do acaso para a produção de homens superiores, mas que façamos isso por meio da educação (*Erziehung*)¹⁷, entendida no sentido de formação. Assim, tal como vimos anteriormente, o que Nietzsche aponta é a necessidade de criação de um novo homem, um homem com uma vontade afirmativa, criadora de valores; um tipo saudável, aristocrático, oposto ao animal de rebanho, ao animal doente que é o homem moderno.

Entretanto, as condições de possibilidade de um tipo superior implicam a posse e o cultivo de determinados traços de caráter que são virtudes próprias desse tipo superior¹⁸, pois, mesmo em se tratando de tipos superiores, “é provável que também nós teremos ainda nossas virtudes, embora naturalmente não serão aquelas ingênuas, inteiriças virtudes pelas quais temos em alta estima, mas também um pouco à distância,

¹⁷ Frezzatti Junior ressalta que “a preocupação nietzschiana sempre foi com a *Erziehung* (educação) e não com a instrução escolar (*Schulung*) e a erudição livresca (*Gelehrsamkeit*)”, pois, para ele, “o objetivo da educação e da cultura nietzschianas é a produção contínua, embora esporádica, de gênios e culturas superiores” (2006, p.185-186).

¹⁸ Para Giacóia Junior, “um périplo pelas formas mais refinadas e grosseiras de sentimentos e juízos de valor moral permite [a Nietzsche] não apenas colher elementos para uma tipologia das morais existentes como também indicar virtudes que estiveram historicamente ligadas aos respectivos tipos de moral” (2005, p.65).

os nossos avós” (BM, 214). Contudo, Nietzsche se posiciona ambigualmente em relação à temática das virtudes ora criticando-as severamente, ora considerando-as importantíssimas para os propósitos de um tipo superior. A discussão mais comum sobre a natureza das virtudes em Nietzsche centra-se em torno de alguns dos tópicos que compõem a primeira parte de *Assim falou Zaratustra*. Porém, para uma visão mais ampla acerca da compreensão nietzschiana das mesmas, deve-se considerar também outras passagens de sua vasta obra em que esse tema é abordado, uma vez que o mesmo é mais insinuado do que tematizado diretamente¹⁹.

Nietzsche deixa transparecer que o valor da virtude depende de seu objetivo, de seu horizonte, de suas forças e impulsos, seus erros e, sobretudo, dos ideais de quem a possui, pois não existe valor em si na virtude; o valor de algo depende da perspectiva que se utiliza (Cf. GC, 120. Assim, quando Nietzsche se posiciona contrário à ideia de virtude, ele está se referindo especificamente às virtudes enaltecidas pela moral de rebanho, quais sejam, espírito comunitário, benevolência, diligência, moderação, modéstia, indulgência e compaixão, que são tidas como as virtudes propriamente humanas (Cf. BM, 199).

Assim sendo, a concepção nietzschiana de virtude é claramente aristocrática:

Deve-se defender a virtude contra os que a exaltam: estes são seus piores inimigos. Pois eles ensinam a virtude como um ideal *para todos*; tomam à virtude o atrativo, que lhe é característico, do raro, inimitável, excepcional e não mediano, – seu *fascínio aristocrático* [...] Reconheço a virtude onde ela 1. não exige ser reconhecida 2. não estabelece virtude como algo que está por toda parte, mas antes justamente outra coisa 3. *não sofre* com a ausência da virtude, mas antes, pelo contrário, considera isso como a relação de distância pela qual algo deve ser honrado na virtude: ela não se compartilha 4. não faz propaganda... 5. não dá licença a ninguém para bancar o juiz, pois é sempre uma virtude *para si* 6. faz justamente tudo o que de costume era *proibido*: virtude, tal como eu a compreendo, é o *virtutum* propriamente dito no interior de toda legislatura do rebanho 7. é virtude no sentido renascentista, *virtú*, virtude livre da estreiteza moral [*moralinfreie*]... (VP, 317).

A teoria das virtudes de Nietzsche é nitidamente diferente das concepções correntes acerca de tema. Ao acentuar o caráter aristocrático de sua concepção das virtudes, o filósofo descreve o cultivo das mesmas como um processo de auto-superação: “O homem é algo que deve ser superado; por isso, cumpre-te amar as tuas virtudes: pois

¹⁹ Hunt comenta que a teoria das virtudes de Nietzsche não é muito clara, pois ele não especifica que objetivos são superiores, que paixões contribuem para a posse dessas virtudes e, consequentemente, não determina os traços particulares que ele faz referência (Cf. 1991, p. 62).

delas perecerás” (ZA, I, 5). Dito isso, compete-nos, então, agora elencar as virtudes enaltecidas por Nietzsche no decorrer de sua obra. Cabe ressaltar, todavia, que essas virtudes não se encontram explicitadas de forma organizada, elas estão espalhadas nas diversas caracterizações acerca do tipo nobre, isto é, os aristocratas, os homens superiores, feitas pelo filósofo em sua obra.

A referência mais direta a uma tábua de virtudes nietzschiana encontra-se em *Aurora* quando, num movimento de inversão e ironia às chamadas virtudes cardeais, o filósofo nomeia suas “quatro boas” virtudes, a saber: “*Honesto* conosco mesmos e quem mais é nosso amigo; *valentes* contra o inimigo; *generosos* para com os vencidos; *cortesês* – sempre: assim nos querem as quatro virtudes cardeais” (A, 557). Nietzsche retoma essa tábua de virtudes diversas vezes em outros momentos de sua obra, mas com o mesmo significado. Em *Além do bem e do mal*, Nietzsche retoma-a acrescida de uma nova e importante virtude: a solidão.

Conservar suas trezentas fachadas; e também os óculos escuros: pois existem casos em que ninguém não deve olhar nos olhos, menos ainda no “fundo”. E escolher como companhia esse vício velhaco e jovial, a cortesia. E continuar senhores de nossas quatro virtudes: coragem, perspicácia, simpatia, solidão. Pois a solidão é conosco uma virtude, enquanto sublime pendor e ímpeto para o asseio, que percebe como no contato entre pessoas – “em sociedade” – as coisas se dão inevitavelmente sujas (BM, 284).

A referência nietzschiana a essas tábuas de virtudes deixa transparecer que o filósofo tem em vista uma forma de aristocracia marcada pela excelência do espírito, pois essas virtudes referem-se a um delicado estado anímico de tensão máxima, a saber, o domínio de *si mesmo*, característico de uma aristocracia espiritual. Nessa recôndita possibilidade de autêntica grandeza habita uma vontade própria, que se proíbe deixar-se arrastar pela vulgaridade, condenando-se assim à solidão. Porém, esse destino é abraçado por um tipo superior como preciosa virtude, pois este *páthos da distância* o mantém preservado para si mesmo e afastado do perigo de sucumbir à tirania anônima do modo comum de sentir e pensar. Esse *páthos da distância* ajuda-o a preservar sua individualidade por meio de uma rígida autodisciplina para consigo mesmo.

3.2. *As virtudes comunitaristas*

A pretensão macintyriana de retomar a tradição clássica, da qual Aristóteles é o principal representante, leva-o a elaborar uma teoria moral fortemente centrada na ideia de virtudes²⁰. Ao propor a recuperação da tradição das virtudes, MacIntyre empreende uma viagem histórica de busca das suas origens nas sociedades heróicas, passando por Atenas e seus poetas e teatrólogos, pela filosofia aristotélica e chegando ao mundo medieval²¹. MacIntyre objetiva oferecer elementos teóricos que permitam sistematizar um conceito de virtude no contexto contemporâneo, recuperando o modelo teleológico aristotélico, mas deixando de lado aqueles elementos que esse mesmo contexto não mais admite como sustentáveis, de tal modo que seja uma conceituação da virtude que respeite a historicidade inerente ao agir humano e a sua necessária dimensão comunitária²².

De acordo com MacIntyre, nas sociedades heróicas, o homem é o que ele faz, ou seja, o homem e seus atos tornam-se idênticos e ele se insere totalmente neles (AV, 122 [211]). Em tais sociedades, todo indivíduo tem determinado papel e *status* dentro de um sistema bem definido e determinado de papéis e *status*. Um indivíduo é responsável por fazer ou deixar de fazer o que qualquer pessoa que ocupe sua função deve aos outros e essa responsabilidade só termina com a morte. Ademais, essa responsabilidade é particular: é para, por e com indivíduos específicos que tenho de fazer o que devo fazer e é perante esses mesmos indivíduos, membros da mesma comunidade local, que sou responsável (AV, 126 [217-218]). Assim, “qualquer explicação adequada das virtudes nas sociedades heróicas seria impossível se as divorciasse de seu contexto em sua estrutura social” (AV, 123 [213]).

²⁰ MacIntyre é um dos principais responsáveis pelo reavivamento da ética das virtudes, uma forma de abordagem ética que tem sua origem no mundo antigo, particularmente nos escritos de Aristóteles, e que, depois da publicação do artigo *Modern Moral Philosophy* (1958) de G. E. M. Anscombe, passou a ocupar um amplo espaço nos debates morais recentes. Para uma visão mais detalhada desse movimento ver: ANNAS (2004); SLOTE (2000).

²¹ Carvalho comenta que “o levantamento das concepções de virtudes em cada um dos estágios dessa tradição clássica de pensamento e ação revela as raízes dos elementos fundamentais que a compõem: das sociedades heróicas advém o vínculo visceral entre virtude e estrutura social; de Atenas e seus poetas e teatrólogos, a visão de conflito como central à vida humana [...]; de Aristóteles advém o esquema teleológico das virtudes, o vínculo com a *polis*, o nexos entre inteligência prática e virtude, o caráter do raciocínio prático e a superioridade da virtude sobre as regras; e, por fim, do período medieval, a componente propriamente histórica que é acrescentado a esse esquema narrativo de compreensão da vida humana como um todo (1999, p. 91-92).

²² Ver Carvalho (2010, 2003).

A estrutura moral das sociedades heróicas contém um esquema conceitual com três elementos inter-relacionados: uma concepção do que exige o papel social que cada indivíduo representa, uma concepção das virtudes como as qualidades que capacitam o indivíduo a fazer o que seu papel exige e uma concepção da condição humana como frágil e vulnerável ao destino e à morte (Cf. AV, 128-129 [221]). Para MacIntyre, a confluência desses elementos nos proporciona um duplo aprendizado: primeiro, que toda moralidade está sempre, até certo grau, amarrada ao socialmente local e particular, e que as aspirações da moralidade moderna à universalidade liberta de toda particularidade é uma ilusão; e, em segundo lugar, que não há como possuir virtudes, a não ser como parte de uma tradição na qual as herdamos, e nosso entendimento delas, a partir de uma série de predecessores na qual as sociedades heróicas estão situadas em primeiro lugar na série (Cf. AV, 126-127 [218]). Das sociedades heróicas advém, portanto, um dos elementos fundamentais da tradição clássica: o vínculo visceral entre virtudes e estrutura social.

De acordo com a argumentação macintyriana, nas sociedades gregas clássicas, a estrutura social heróica sofreu profundas mudanças e transformações, deslocando-se o foco da moralidade das relações parentescas para a cidade-estado, na qual a concepção de virtude se separa de qualquer papel social particular (Cf. AV, 133 [228]). A discordância moral na cultura ateniense dos séculos V e IV a.C. se caracteriza não apenas porque um conjunto de virtudes se contrapõem a outros, mas também por existirem concepções rivais da mesma virtude. O que há de comum entre essas concepções rivais de virtude é a crença aceita por todas, sem questionar, que o meio no qual as virtudes são exercidas e segundo o qual devem ser definidas é a *pólis*, pois ali é o lugar que o conflito ocupa na ordem social grega clássica (Cf. AV, 135 [232])²³.

Quanto ao conflito entre concepções de virtudes rivais e incoerentes na Atenas do século V, MacIntyre considera que foi Sófocles quem melhor explorou esse assunto, contribuindo, assim, decisivamente para a tradição clássica ao estabelecer a compreensão das virtudes e da vida humana como possuindo a forma narrativa dramática (Cf. AV, 142 [242]). Por conseguinte, para MacIntyre, o ensinamento proporcionado pela Atenas dos séculos V e IV a.C., que se torna claro mais fortemente em Sófocles, e que constitui um componente fundamental da concepção de virtude da

²³ Convém acentuarmos que MacIntyre opera sua filosofia em um contexto de radical desacordo moral e busca oferecer uma alternativa para isso em bases contemporâneas. Ver Carvalho (2001a).

tradição clássica, é a ênfase no elemento narrativo como correlativo de uma definição das virtudes.

Porém, é de Aristóteles a teoria das virtudes que constitui decisivamente a tradição clássica como tradição de pensamento moral, pois, segundo MacIntyre, o Estagitira estabelece com firmeza uma grande parte do que seus antecessores poéticos só conseguiram afirmar ou insinuar, transformando assim a tradição clássica numa tradição racional (Cf. AV, 147 [251]). A teoria aristotélica das virtudes pressupõe a ideia de que os seres humanos, bem como os membros de todas as outras espécies, têm uma natureza específica e essa natureza é tal que eles têm certos objetivos e metas, de modo que se movimentam pela natureza rumo a um *télos* específico. Assim, o bem humano consiste na *eudaimonia*, um estado de estar bem e fazer bem ao estar bem para o qual a posse das virtudes é fundamental (Cf. AV 148 [252-253]).

MacIntyre adverte, contudo, que as virtudes não devem ser entendidas simplesmente como *meio* para alcançar o bem do homem, pois o que constitui o bem para o homem é uma vida humana completa e o exercício das virtudes é uma parte necessária e fundamental de tal vida e não um mero exercício preparatório para garanti-la (Cf. AV, 149 [254]). A teoria aristotélica das virtudes pressupõe, portanto, uma distinção fundamental entre o que qualquer indivíduo em determinado momento acredita ser bom para ele e o que é realmente bom para ele como homem.

No entanto, MacIntyre reconhece que existem limitações e problemas na teoria aristotélica das virtudes, alguns dos quais não comprometem seriamente a consistência de suas posições desde que sejam corrigidos, mas outros cuja manutenção ameaça seriamente a validade dessas posições e, por isso mesmo, sua retomada. Dentre os do primeiro tipo, MacIntyre destaca a crença aristotélica na unidade das virtudes que tem na sua base uma posição de negação do conflito, a existência de certa tensão entre as concepções de Aristóteles do homem como essencialmente político e do homem como essencialmente metafísico e o lugar da liberdade na estrutura social e metafísica concebida por Aristóteles, mais especificamente o tratamento dado aos bárbaros e escravos, considerados não-gregos (Cf. AV, 157-159 [267-270])²⁴.

MacIntyre ressalta, entretanto, que existem pelo menos três tipos de questões que, se não forem resolvidas satisfatoriamente, põem em risco toda a estrutura

²⁴ Ver Carvalho (2001b).

aristotélica e sua reapropriação contemporânea. A primeira diz respeito à íntima relação entre a teleologia aristotélica e a sua biologia metafísica. Para MacIntyre, qualquer explicação teleológica adequada deve oferecer uma explicação clara e defensável do *télos*, o que implica dizer que qualquer explicação adequada, de caráter aristotélico, deve oferecer uma explicação teleológica que possa substituir a biologia metafísica de Aristóteles (Cf. AV, 163 [276])²⁵. A segunda diz respeito à relação da ética com a estrutura da *polis*. MacIntyre questiona como seria possível reformular o aristotelismo de modo que este se torna relevante moralmente num mundo onde não existem cidades-estados, se grande parte dos detalhes da explicação aristotélica das virtudes pressupõe o contexto das relações sociais das antigas cidades-estado há muito extinto (Cf. AV, 163 [276]). A terceira e última refere-se à forte presença da herança platônica na crença aristotélica na unidade e harmonia tanto da alma humana como da cidade-estado e a consequente percepção do conflito como algo a ser evitado e controlado. Para MacIntyre, a falta de reconhecimento da centralidade da oposição e do conflito na vida humana esconde de Aristóteles uma importante fonte de aprendizagem humana sobre as virtudes e um importante meio de vivência humana nas virtudes, pois é por meio do conflito e, às vezes, somente por meio do conflito, que descobrimos quais são nossos fins e propósitos (Cf. AV, 163-164 [277-278]).

Por fim, é no período medieval que a tradição clássica, por meio da permanência do esquema aristotélico enriquecido, mas não substancialmente modificado, por uma nova concepção de *télos* e pela descoberta agostiniana da vontade capaz de dar consentimento ao mal, adquire seus contornos finais, tornando-se uma teoria moral consistente. MacIntyre começa esclarecendo em que aspectos a teoria medieval difere da teoria aristotélica: a Idade Média elabora uma estrutura narrativa da vida humana em que o homem é compreendido como essencialmente *in via*, ou seja, “o fim que procura é algo que, se conquistado, pode redimir tudo o que havia de errado em sua vida até aquele ponto” (AV, 175 [295]).

²⁵ No prefácio de *Dependent Rational Animals*, MacIntyre retoma essa questão afirmando que se equivocou ao supor que era possível uma ética aristotélica independente de sua biologia. Segundo ele, essa correção se deve a duas razões distintas: primeiro, nenhuma explicação dos bens, normas e virtudes que definem a vida moral será satisfatória se não conseguir explicar como essa é possível para seres constituídos biologicamente como o ser humano, oferecendo uma explicação do desenvolvimento humano concernente a essa forma de vida e dentro dela; segundo, a incapacidade para entender essa condição e para entender como pode ser iluminado pela comparação entre o ser humano e outras espécies animais inteligentes, deixará escondido aspectos fundamentais desse desenvolvimento (Cf. DRA, p. x).

Essa concepção do fim do homem não é aristotélica em dois pontos: primeiro, a ideia de redenção final de uma vida quase totalmente sem regeneração não tem lugar no esquema aristotélico, já que neste o *télos* de uma vida virtuosa não é algo a se conquistar a certa altura no futuro, mas no modo como construímos toda a nossa vida; e, segundo, a ideia de vida humana como uma busca ou jornada na qual encontramos e superamos uma série de formas do mal requer um conceito de mal do qual existem no máximo indícios nos escritos de Aristóteles.

Desse modo, segundo MacIntyre, embora a concepção medieval das virtudes continue teleológica, é um conceito bem diferente do de Aristóteles em pelo menos dois aspectos importantes, além de sua compreensão cristã e agostiniana do mal. Primeiramente, Aristóteles acredita que infortúnios externos podem frustrar a possibilidade de se alcançar o bem humano, a *eudaimonia*, enquanto que na perspectiva medieval nenhum ser humano está excluído do bem humano por quaisquer condições externas, uma vez que nem mesmo o mal que possa nos acontecer exclui tal possibilidade, desde que não nos tornemos seus cúmplices (Cf. AV, 176 [296]). Segundo, e aqui reside a contribuição fundamental dos medievais para a tradição clássica, a perspectiva medieval é histórica de um modo que a de Aristóteles não poderia ser, pois “ela situa nossa aspiração ao bem não apenas em contextos específicos – Aristóteles situa essa aspiração dentro da *polis* – mas em contextos que também têm uma história” (AV, 176 [297]). Para MacIntyre, portanto, os pensadores medievais tinham uma concepção de vida humana como histórica, na qual as virtudes são qualidades que desempenham a função de habilitar o homem a enfrentar e vencer os males nessa sua jornada histórica na busca do seu bem²⁶.

Assim, ao reconstruir desta forma a história das virtudes, MacIntyre elabora um conceito de virtude de aspecto fortemente histórico-cultural. Para MacIntyre, “a virtude é uma qualidade humana adquirida, cuja posse e exercício costuma nos capacitar a alcançar aqueles bens internos às práticas e cuja ausência nos impede, para todos os efeitos, de alcançar tais bens” (AV, 219 [321]). Essa definição seria complementada em *Dependent Rational Animals*, onde o filósofo as compreende como “qualidades intelectuais e de caráter que permitem a uma pessoa identificar os bens relevantes e usar as habilidades necessárias para conseguir-los” (DRA, 92). MacIntyre

²⁶ Sobre essa dimensão histórica em MacIntyre, ver Carvalho (2005).

demonstra, assim, que as virtudes são necessárias em todos os âmbitos da vida humana, enfatizando a influência das mesmas nas atitudes em diferentes situações.

Quanto à lista das virtudes enaltecidas por MacIntyre, ele nomeia claramente três como sendo as virtudes à luz das quais precisamos caracterizar a nós mesmos e aos outros, seja qual for nossa perspectiva moral ou os códigos específicos da nossa sociedade, quais sejam, a sinceridade, a justiça e a coragem (Cf. AV, 191-192 [324]). Entretanto, MacIntyre amplia e aprofunda consideravelmente essa lista de virtudes quando inclui a romancista Jane Austen como a última grande voz eficiente e criativa da tradição de pensamento e de prática das virtudes que ele vem tentando identificar, uma vez que ela se afasta dos catálogos concorrentes das virtudes do século XVIII, restabelecendo uma perspectiva teleológica a partir da união de temas cristãos e aristotélicos num contexto social específico (Cf. AV, 240 [402]). Seu catálogo de virtudes incluirá, portanto, a partir da apreciação dos romances de Austen, o autoconhecimento (cristão e não socrático), a constância, a paciência e, novamente, a coragem como as virtudes necessárias para sustentar as espécies de comunidades políticas nas quais homens e mulheres podem procurar o bem juntos.

4. Nietzsche e MacIntyre: perspectivas convergentes?

Até o presente momento, procuramos retratar os elementos conceituais que compõem as teorias morais de Nietzsche e MacIntyre, objetivando, com isso, responder à pergunta-título que norteia o presente trabalho. Considerando o que foi exposto até aqui, decorre que esses dois teóricos apresentam fortes convergências no que diz respeito a alguns dos seus temas e problemas. Nietzsche e MacIntyre têm em comum um esquema conceitual que conserva três elementos inter-relacionados: uma concepção de florescimento humano como fundamental à empresa ética, o cultivo de determinadas virtudes como condição necessária para tal florescimento e, finalmente, um conceito de homem como ser inacabado, como algo suscetível de aperfeiçoamento. Não obstante divergirem quanto ao que venha a ser cada um desses pontos, a ideia básica de cada um deles é mantida.

Ora, a possibilidade de tal convergência parece implicar, entre outras coisas, um novo olhar sobre a ideia de uma escolha radical entre Nietzsche *ou* Aristóteles sugerida por MacIntyre (Cf. AV, 109-120 [189-207]). É que, se existem elementos

conceituais comuns entre Nietzsche e MacIntyre, tal como estamos apontando, essa escolha se torna complexa, pois parece não se tratar mais de uma simples escolha disjuntiva, mas de uma escolha aditiva.

MacIntyre parece estar bem mais próximo de Nietzsche do que, aparentemente, se pode imaginar. A retomada do aristotelismo proposto por MacIntyre depende da percepção de que o iluminismo falhou enquanto projeto de descobrir qualquer justificação racional para a moralidade. Segundo MacIntyre, foi devido ao fato de uma tradição moral, da qual o pensamento de Aristóteles foi o principal núcleo intelectual, ter sido repudiada durante as transições entre os séculos XV e XVII, que foi preciso empreender o projeto das luzes e foi porque esse projeto falhou que MacIntyre interroga: será que estava certo rejeitar Aristóteles? (AV, 117 [202-203]).

Para MacIntyre, quem melhor percebeu o fracasso do projeto iluminista de justificar racionalmente a moralidade foi Nietzsche. De acordo com ele, Nietzsche teve o mérito histórico de entender mais claramente do que qualquer outro filósofo que o apelo à objetividade, tão característico do projeto iluminista, não era senão um dos disfarces da vontade subjetiva, bem como perceber a natureza dos problemas que isso representou para a filosofia moral (Cf. AV, 113 [196]). Assim, é Nietzsche quem fornece, de certo modo, uma parte essencial do quadro teórico do projeto filosófico de MacIntyre, pois a pertinência e a plausibilidade da argumentação macintyriana dependem da validade da crítica nietzschiana à condição moral moderna. Sem a percepção nietzschiana da inevitável falha do projeto iluminista, MacIntyre não teria ferramentas conceituais importantes para elaborar seu diagnóstico da situação em que se encontra o discurso e a prática morais atuais e, conseqüentemente, não poderia propor retomar Aristóteles nas bases que o fez em *After Virtue*²⁷.

Em *First Principles, Final Ends* (1990), MacIntyre chega mesmo a condicionar a provisão de uma teoria aristotélico-tomista à genealogia nietzschiana (Cf. FP, p. 57). Segundo ele, a contemporaneidade é um cenário filosófico de problemas não-solucionados e de desacordos sem resolução. Os conceitos que produzem estas divisões divergem radicalmente com os modos de pensamento característicos da modernidade, de modo que não é surpreendente que apareçam como não-funcionais ou desorientados ou ambos, com respeito a esses modos de pensamento, apesar de continuarem

²⁷ Ver também Carvalho (2010b).

aparecendo e dando continuidade a suas antigas funções, produzindo assim fortes tensões (FP, 55). Frente a essas dificuldades em que se encontra a filosofia contemporânea, MacIntyre sugere que a elaboração de uma teoria nos moldes aristotélico-tomistas pode dar conta de explicar não apenas as dificuldades em que se encontra a filosofia contemporânea, mas como essas dificuldades foram produzidas e em que condições podem ser evitadas e superadas (FP, 57).

MacIntyre ressalta, entretanto, que a provisão de uma teoria semelhante requer a construção de algo análogo ao que Nietzsche chama de genealogia (FP, 57). De acordo com MacIntyre, a narrativa genealógica não argumenta, mas desvela algo acerca das crenças, pressupostos e atividades de alguma classe de pessoas, logo, ela teria a vantagem de explicar, de um modo característico, como se chegou a essas dificuldades e porque não se pode reconhecer ou diagnosticar adequadamente a natureza dessas dificuldades a partir dos recursos conceituais e argumentativos da própria filosofia contemporânea.

Para MacIntyre, construir a genealogia da filosofia contemporânea, ou pelo menos de boa parte dela, desvelaria três aspectos dessa filosofia que estão, de alguma maneira, ocultos: primeiro, o entendimento de como se constituiu a problemática característica da filosofia contemporânea e qual sua relação com os momentos iniciais da história da filosofia moderna; segundo, as teses e argumentos, usados no interior da filosofia contemporânea acerca da verdade e da racionalidade enquanto tal, advêm do contexto aristotélico-tomista dentro do qual são ou completamente inteligíveis ou adequadamente defensáveis; e finalmente, o fato de o defensor da filosofia contemporânea encontrar-se em certo dilema, pois se, por um lado, ele a entende como um progresso rumo a um maior esclarecimento, o que ele está oferecendo é algo muito semelhante ao tipo de explicação narrativa cuja estrutura pressupõe justamente o tipo de ordenação teleológica amplamente recusado pela filosofia moderna; mas, por outro lado, se ele não a entende assim, então a filosofia não pode ser compreendida como possuidora de história acumulativa inteligível, exceto no tocante à compreensão dos detalhes de pontos de vista diferentes (Cf. FP, 62-63).

Para MacIntyre, é adotando esse tipo de história genealógica que o tomismo será capaz de abrir possibilidades de diálogo e debates filosóficos com posições das quais ele não compartilha premissas ou pressupostos comuns de justificação racional. O

tomismo nos habilitaria, portanto, a escrever um tipo de história da filosofia moderna e contemporânea que essa filosofia não pode oferecer por si mesma. Assim, visando iniciar uma conversa filosófica entre tomistas e representantes de posições filosóficas contemporânea, MacIntyre recorre a meios não tomistas com o intuito de conseguir fins tomistas (FP, 57).

Todavia, esse tipo de acordo entre teorias rivais é perfeitamente possível, pois, segundo MacIntyre, a incomensurabilidade entre tradições não é algo permanente e absoluto. Pode até mesmo acontecer que duas tradições, até então independentes e mesmo antagônicas, passem a reconhecer certa possibilidade de acordo fundamental e se reconstituam como um debate único e mais complexo (JR, 24). Só podemos identificar adequadamente nossos próprios compromissos e os dos outros nos conflitos argumentativos do presente se os situarmos dentro das histórias que os fizeram ser o que são; não há nenhuma base, nenhum lugar para a pesquisa, nenhum modo de se avançar, avaliar, aceitar e rejeitar argumentações racionais que não seja fornecido por uma ou outra tradição (JR, 376). Contudo, MacIntyre ressalta que isso não implica, necessariamente, que o que se diz numa tradição não possa ser ouvido por outra. Segundo ele, tradições que diferem radicalmente sobre certos assuntos podem, quanto a outros, compartilhar crenças, imagens e textos, pois é no enfrentamento e na abertura às posições rivais que uma tradição se desenvolve (JR, 376-377)²⁸.

Assim, apesar de reconhecer a incomensurabilidade entre tradições rivais, MacIntyre também admite a possibilidade de que existam semelhanças entre essas tradições, pois, na medida em que uma tradição de pesquisa racional se constitui como tal, ela tende a reconhecer elementos comuns com outras tradições, o que implicaria certa comensurabilidade²⁹. MacIntyre ressalta, contudo, que a possibilidade de acordo entre tradições conflitantes só pode ser captado por alguém que vive e fala os dois esquemas conceituais alternativos, quem reside na fronteira, ou seja, quem participa das duas tradições em questão. Um exemplo disso é o que ocorreu com Tomás de Aquino. Segundo MacIntyre, ele vivia na fronteira entre as tradições aristotélica e agostiniana, e, como todo aquele que habita uma fronteira, aprendeu a falar ambas as línguas, de modo

²⁸ Ver Linhares (2011) sobre esse enfrentamento entre tradições em conflito.

²⁹ Izquierdo comenta, numa nota de rodapé, que MacIntyre trata confusamente essa questão, pois reconhece, por um lado, que existem elementos comuns entre tradições rivais – o que denotaria certa comensurabilidade entre elas –, mas, por outro lado, ele afirma que a tradutibilidade de uma tradição por outra não implica comensurabilidade (2007, p. 287-288).

que pôde enxergar bem a crise em que ambas as tradições se encontravam imersas, o que lhe possibilitou fundir essas duas tradições, até então, rivais e inconciliáveis, formando um único esquema conceitual (Cf. TRV, 120-125).

É interessante notar que, nesse ponto, o próprio MacIntyre também parece encontrar-se numa posição semelhante à de Tomás de Aquino. Por um lado, compartilha com Nietzsche o diagnóstico da condição moral moderna, mas reconhece as limitações da posição nietzschiana no que diz respeito aos problemas diagnosticados, apontando a tradição aristotélico-tomista como uma alternativa capaz de resolver satisfatoriamente tais problemas. Por outro lado, ele ressalta que a provisão de uma teoria semelhante requer a construção de algo análogo ao que Nietzsche chama de genealogia. MacIntyre se apropria, deste modo, de recursos teóricos que são inicialmente alheios ao seu esquema conceitual, fundindo, em alguns pontos, a tradição aristotélico-tomista com a genealógica, ou seja, ele recorre a meios não tomistas para atingir fins tomistas. Trata-se, portanto, de teorias que se complementam, o que inviabiliza a ideia de uma escolha excludente entre Nietzsche e Aristóteles e acena para a ideia de uma escolha aditiva.

Nesse sentido, a escolha entre Nietzsche e Aristóteles, proposta por MacIntyre em *After Virtue*, não é algo que se revele como excludente. Pelo contrário, uma simples escolha que deixe intocada cada uma das posições em jogo no embate pelas nossas consciências não se faz mais exequível. Da perspectiva macintyriana, a própria escolha por Aristóteles só se tornou possível pela incorporação de elementos nietzschianos, como vimos, tanto pela assunção da crítica de Nietzsche à moralidade moderna, como pela assunção da genealogia como procedimento metodológico essencial na reflexão sobre a historicidade da moralidade.

Por outro lado, a afirmação pura e simples da perspectiva nietzschiana, desconsiderando completamente elementos postos pela matriz aristotélico-tomista de MacIntyre, implicaria em um mergulho ingênuo daquela na maquinaria capitalista do individualismo consumista, uma vez que não perceberia sua própria inserção histórica no quadro moderno das sociedades burguesas. Com isso, a busca daquele indivíduo aristocrático estaria sujeita à perdição nas malhas da matriz do lucro e do consumo próprio da vida burguesa moderna, que destroem a possibilidade de quaisquer virtudes que não aquelas que denegam a valorização da vida e a condição criadora do homem na perspectiva radical de Nietzsche. A grande saúde nietzschiana não seria possível sem a

quebra dessas estruturas sociais liberais criticadas por MacIntyre e a libertação dos desejos transformados em mercadorias pela maquinaria econômico-política institucional do capitalismo, através da recuperação de certo *télos* da vida.

Em suma, o embate entre essas duas importantes posições alternativas à moralidade moderna revela a complexidade própria da vida moral e de como esse conflito entre posições aparentemente tão distintas revela tantas proximidades e exigências de intercâmbio teórico para que uma delas possa pensar em afirmar ser superior à outra. Portanto, as perspectivas de Nietzsche e MacIntyre não são, a nosso ver, alternativas excludentes à moralidade moderna, mas uma escolha aditiva, em que a assunção de uma implica necessariamente a incorporação de elementos da outra.

Helder Buenos Aires de Carvalho

José Elielton de Sousa

Universidade Federal do Piauí

hbac@ufpi.edu.br

jose_elielton@yahoo.com.br

Referências

- Annas, Julia (2004) 'Virtue Ethics'. IN COPP, David (ed.). *The Oxford Companion to Ethical Theory*. Oxford: Oxford University Press.
- Anscombe, G. E. M. (1958) 'Modern Moral Philosophy'. *Philosophy*. Vol. 33, nº 124: 1-19.
- Aristote (1970) *L'Éthique a Nicomaque*. Introduction, Traduction et commentaire. TI-2. 2ª ed. Ed. GAUTHIER, R.A & JOLYF, Y. Louvain, Paris: Publications Universitaires/Béatrice Nauwelarts.
- Canto-Sperber, M. (2004) *Que devo fazer? A filosofia moral*. Tradução de Benno Dischinger. São Leopoldo: Editora Unisinos.
- Carvalho, Helder Buenos Aires de. (1999) *Tradição e racionalidade na filosofia de Alasdair MacIntyre*. São Paulo: Editora Unimarco.
- _____. (2010) 'Ética das virtudes em Alasdair MacIntyre'. IN HOBUSS, João (org). *Ética das Virtudes*. Florianópolis: EDUFSC, pp.189-214.
- _____. (2010b) 'Alasdair MacIntyre e a ética nietzschiana: apontamentos para uma reflexão crítica'. *Dissertatio* (UFPel). Vol. 32: 277-318.
- _____. (2003) 'Comunitarismo, liberalismo e tradições morais em Alasdair MacIntyre'. IN OLIVEIRA, M. A. de; ALVES, O. S.; SAHD NETO, L. F. S. (org). *Filosofia Política Contemporânea*. Petrópolis: Vozes, pp. 287-310.

- _____. (2001a) 'Alasdair MacIntyre: um kuhniano da filosofia moral? Apontamentos para uma reflexão sobre racionalidade e história na filosofia moral'. *Cadernos de ética e filosofia política* (USP). Vol. 3: 25-52.
- _____. (2001b) 'A contemporaneidade de Aristóteles na filosofia moral de Alasdair MacIntyre'. *Síntese – Revista de Filosofia*. Vol. 28, n.90: 37-66.
- _____. (2005) 'O lugar da história na filosofia moral: apontamentos sobre *Depois da Virtude* de Alasdair MacIntyre'. IN BRANCO, E. A.; NASCIMENTO, F. A.; PINHEIRO, A. V. (orgs). *História: Cultura, Sociedade, Cidades*. Recife: Ed. Bagaço, pp. 153-172.
- Dias, Rosa M. (1993) *Nietzsche educador*. São Paulo: Scipione.
- Frezzatti Junior, W. A. (2006) *A fisiologia de Nietzsche: a superação da dualidade cultura/biologia*. Ijuí: Ed. Unijuí.
- Giacioia Junior, Oswaldo. (2005) *Nietzsche & Para além do bem e do mal*. 2ª ed. Rio de Janeiro: Jorge Zahar Editor.
- Hunt, Lester H. (1991) *Nietzsche and the origin of virtue*. London/New York: Routledge Press.
- Izquierdo, David Lorenzo. (2007) *Comunitarismo contra individualismo: una revisión de los valores de Occidente desde el pensamiento de Alasdair MacIntyre*. Cizur Menor: Editorial Aranzadi.
- Kant, Immanuel. (1980) *Fundamentação da metafísica dos costumes*. Tradução de Paulo Quintana. 2ª edição. São Paulo: Abril Cultural.
- Kaufmann, Walter. (1974) *Nietzsche, Philosopher, Psychology, Antichrist*. 4ª ed. Princeton: Princeton University Press.
- Linhares, João Caetano. (2011) *O teste histórico-dialético das tradições de pesquisa moral na filosofia de Alasdair MacIntyre*. Universidade Federal do Piauí/Mestrado em Ética e Epistemologia, Teresina. (Dissertação)
- MacIntyre, A. (2007) *After virtue*. A study in moral theory. 3ª ed. London, Duckworth. (Ed. brasileira: *Depois da virtude: um estudo em teoria moral*. Trad. Jussara Simões; Revisão Helder B. A. de Carvalho. Bauru: EDUSC, 2001) (AV).
- _____. (1999) *Dependent Rational Animals: Why Human Beings Need the Virtues*. Chicago: Open Court Publishing Company, (DRA).
- _____. (1991) *Justiça de Quem? Qual Racionalidade?* Trad. Marcelo Pimenta. São Paulo: Loyola, (Ed. original: *Whose Justice? Which Rationality?* Notre Dame: University of Notre Dame Press, 1988) (JR).
- _____. (1990) *Three Rival Versions of Moral Inquiry: encyclopaedia, genealogy and tradition*. London: Duckworth, (TRV).
- Marton, Scarlett. (2000) *Nietzsche: das forças cósmicas aos valores humanos*. 2ª ed. Belo Horizonte: Ed. UFMG.
- _____. 'Por uma filosofia dionisíaca'. *Kriterion - Revista de Filosofia*. Vol. 35, nº 89 (1994): 9-20.
- Nietzsche, F. W. (1992) *Além do bem e o mal: prelúdio a uma filosofia do futuro*. Tradução, notas e posfácio: Paulo César de Souza. São Paulo: Companhia das Letras, (BM).
- _____. (2001) *A Gaia ciência*. Tradução, notas e posfácio: Paulo César de Souza. São Paulo: Companhia das Letras, (GC).
- _____. (2008a) *A vontade de poder*. Tradução original do alemão e notas Marcos Sinésio Pereira Fernandes, Francisco José Dias de Moraes; apresentação Gilvan Fogel. Rio de Janeiro: Contraponto, (VP).

- _____. (2008b) *Assim falou Zaratustra: um livro para todos e para ninguém*. Tradução: Mário da Silva. 17ª ed. Rio de Janeiro: Civilização Brasileira, (ZA).
- _____. (1998) *Genealogia da moral: uma polêmica*. Tradução, notas e posfácio: Paulo César de Souza. São Paulo: Companhia das Letras, (GM).
- _____. (2007) *O Anticristo. Maldição ao Cristianismo; Ditirambos de Dioniso*. Tradução de Paulo César de Souza. São Paulo: Companhia das Letras, (AC).
- Slote, Michael (2000) 'Virtue Ethics'. IN LAFOLLETTE, Hugh (ed.). *The Blackwell Guide to Ethical Theory*. Cambridge: Blackwell Publishers, pp. 325-347.