Volume 4 Number 2
2008
ISSN 1807-9792

abstracta

Linguagem, Mente & Ação

ABSTRACTA

Linguagem, Mente e Ação

ISSN 1807-9792

Volume 4 Number 2
2008

TABLE OF CONTENTS

Abstracta 4: 2 pp.74-95, 2008

CAUSAL INHERITANCE AND SECOND-ORDER PROPERTIES

Suzanne Bliss & Jordi Fernández

Abstract

We defend Jaegwon Kim's 'causal inheritance' principle from an objection raised by Jurgen Schröder. The objection is that the principle is inconsistent with a view about mental properties assumed by Kim, namely, that they are second-order properties. We argue that Schröder misconstrues the notion of second-order property. We distinguish three notions of second-order property and highlight their problems and virtues. Finally, we examine the consequence of Kim's principle and discuss the issue of whether Kim's 'supervenience argument' generalizes to all special sciences or not.

1. Introduction

Our purpose in this essay is to defend Jaegwon Kim's 'causal inheritance' principle from a certain objection raised by Jurgen Schröder.[1] The objection is that the principle is inconsistent with a view about mental properties assumed by Kim, namely, that mental properties are second-order properties. The significance of this objection has to do with a different worry about Kim's 'supervenience argument' against non-reductive physicalism. The worry is that, if it is correct, the argument not only shows that psychological properties must be reduced to physical properties; it also shows that biological properties, chemical properties, economical properties and all properties in the domains of the special sciences must be reduced as well. According to Schröder, the right lesson to draw from the supervenience argument, given that the causal inheritance principle fails, is that all these properties are causally idle. We shall argue that Schröder's charge of inconsistency relies on a misunderstanding of the notion of second-order property.

We proceed as follows. In section 2, we describe the causal inheritance principle and we place it against the background of Kim's reductionist program. In section 3, we evaluate Schröder's argument against the causal inheritance principle. Our contention is

[1] Schröder (2002).

that Schröder's argument relies on attributing a certain notion of second-order property to Kim which he does not have. In sections 4 and 5, we step back from the debate on causal inheritance and we explore two broader issues on mental causation that the debate in question leaves open. Section 4 is devoted to the issue of whether the supervenience argument in fact generalizes. We pull apart three readings of this question. Some of these readings, we suggest, warrant a positive answer whereas others do not. Section 5 is devoted to the issue of how exactly we should think of mental properties as second-order properties. We briefly map some of the explanatory virtues and challenges of three notions of second-order properties and conclude that each of them is useful to deal with some of the problems in the mental causation literature. But none of them provides us with a satisfactory account of three of those problems.

2. The causal inheritance principle and the supervenience argument

The causal inheritance principle is sometimes formulated as a principle about the causal powers of mental properties and it is sometimes formulated as a principle about the causal powers of second-order properties. Thus, Kim offers the following two formulations of the principle, which we may label 'causal inheritance for mental properties' (or simply 'CIM') and 'causal inheritance for second-order properties' (or 'CI2' for short):

CIM If mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of this instance of M are identical with the causal powers of P.[2]

CI2 If a second-order property F is realized on a given occasion by a first-order property H (that is, if F is instantiated on a given occasion in virtue of the fact that one of its realizers, H, is instantiated on that occasion), then the causal powers of this particular instance of F are identical with (or are a subset of) the causal powers of H (or of this instance of H).[3]

[2] Kim (1993a), p. 326.
[3] Kim (1998), p. 54.

The reason why it is not problematic that the principle is offered in two forms concerns the role that it plays within Kim's overall reductionist project. Let us therefore turn our attention to that project briefly.

The causal inheritance principle is part of a broader argument against non-reductive physicalism, sometimes labeled the 'causal exclusion' or 'supervenience' argument.[4] This well-known argument is meant to show that non-reductive physicalism is inconsistent given some weak assumptions. More specifically, the argument is meant to show that the following five theses are inconsistent:

1. Anti-reductionism: Mental properties are not physical properties.
2. Supervenience: Mental properties supervene on physical properties. Necessarily, for any object x and any mental property M, if x has M, then there is a physical property P such that x has P and, necessarily, if any object y has P, then y has M.[5]
3. Causal closure of the physical domain: Every physical event that has a cause at time *t* has a physical cause at *t*.[6]
4. Mental causation: Mental properties are causally efficacious.
5. Causal exclusion principle: If an event E has a sufficient cause C, then no event distinct from C can be a cause of E (unless this is a genuine case of causal overdetermination).[7]

Claims 2 and 3 are supposed to capture the minimal commitments of physicalism and, consequently, claims 1-3 are meant to describe the non-reductive physicalist's commitments. Claim 4 is grounded on the intuitive idea that the instantiation of a mental property may cause the instantiation of a different mental property, in the way in which

[4] See Kim (1997, 1998) for two versions of the supervenience argument.
[5] Strictly speaking, this is the definition of 'strong' supervenience. The question of whether this formulation of supervenience is essential to the argument is relevant to the debate on whether the argument generalizes or not.
[6] Kim (1993b), p. 280.
[7] See Kim (2001) pp. 276-278. For a defence of this principle see Kim (1993c).

being in pain causes one to lose concentration, for instance (hereafter 'mental-to-mental' causation). And it may cause the instantiation of a physical property, in the way in which being in pain may cause one to wince (hereafter 'mental-to-physical' causation). The denial of claim 5 is supposed to lead to the counter-intuitive result that, systematically, human behavior is causally overdetermined by neurological and mental causes. Thus, Kim takes claims 4 and 5 to be plausible assumptions. He then proceeds to derive a contradiction from all five claims, namely, that there is and there is not mental-to-physical causation. Basically, the contradiction is obtained by arguing for the following conditionals:

C1 If (mental-to-mental causation and supervenience), then there is mental-to-physical causation.[8]

C2 If (anti-reductionism, causal closure and causal exclusion), then there is no mental-to-physical causation.[9]

The bottom line is that 1-5 constitutes an inconsistent set of claims. The moral of the argument is meant to be that the best way of securing mental causation within a physicalist framework is by embracing the reductionist view that mental properties are identical to physical properties. Basically, we must drop claim 1 in order to hold on to claim 4.

Evaluating the supervenience argument is not our concern in this essay. However, for the purposes of assessing Schröder's attack on the causal inheritance principle, it is worth pointing out that this principle is grounded on the supervenience argument.[10] If the right lesson to draw from the argument is that mental properties must be identified with physical properties, then it is not surprising that the causal powers of the former turn out to be identical to the causal powers of the latter. This explains that the causal inheritance principle is sometimes presented as CIM and it is sometimes presented as CI2. The

[8] Kim (1993d), pp. 350-353 argues for this conditional.
[9] An argument for this conditional is provided in Kim (1993d), pp. 353-357.
[10] In (1993d), p. 355, Kim says of those considerations that support the supervenience argument that they 'point to' a picture of the world that can be stated in a version of the causal inheritance principle.

reason for this is that the principle is grounded on a *reductio* of the view that theses 1-5 above are consistent. And a common point often raised in support of thesis 1 is that mental properties are second-order with respect to physical properties. Thus, Kim's discussion of the causal inheritance principle assumes, for the sake of a *reductio*, the following 'mental as second-order' view (or, for short, M2):

M2 Mental properties are second-order properties.

As a fist approximation to this view, we can say that, according to M2, all physical properties are, in some sense yet to be specified, ontologically basic or 'first-order.' And, for each mental property M, there is a physical property P with respect to which M is 'second-order.' (We may call this physical property P the 'first-order property of' M.) If one assumes that M2 holds, then CIM simply emerges as a more restricted version of CI2. For CI2 clearly entails CIM assuming M2. The upshot is that the central version of the causal inheritance principle is CI2. Following Schröder, we shall concentrate on that formulation of the principle for the remainder of our discussion.

Now, a concern that a number of philosophers have had about the supervenience argument is that it may show too much. If the argument is right about psychological properties, some have argued, then all properties in the domains of any special science must reduce to physical properties for analogous reasons.[11] Schröder shares the view that the supervenience argument generalizes in that way. This is why he thinks that, given that Kim's causal inheritance principle fails, the supervenience argument delivers the result that all second-order properties are causally inert.[12] If the causal inheritance principle fails, then giving up claim 1 above is not an option as a reaction to the supervenience argument. If the principle fails, then the causal powers of mental properties are different from the causal powers of physical properties. This means that we cannot identify mental properties with physical properties (unless we give up the indiscernibility of identicals, that is). Thus, the supervenience argument seems to force us to give up some claim

[11] For expressions of this worry, see Van Gulick (1992), p. 325, Burge (1995), p. 102 and Baker (1995), p. 77.
[12] Schröder (2002) p. 322.

among 2-5. The natural candidate to turn to is claim 4. The upshot is that we need to come to terms with the fact that mental properties are not causally efficacious. But if the supervenience argument generalizes, then Schröder's attack on the causal inheritance principle seems to deliver the stronger result that, for analogous reasons, biological properties, chemical properties and economical properties are not causally efficacious either. A lot seems to hang on Schröder's attack on the causal inheritance principle, then, if the supervenience argument generalizes. We will return to the generalization issue in section 4. First, let us examine Schröder's attack on the principle.

3. The argument against causal inheritance

What exactly is wrong with causal inheritance according to Schröder? Schröder's objection is ingeniously raised as a worry about the metaphor of inheritance. Thus, Schröder claims:

> If the metaphor of inheritance is to be any guide in these abstract matters there must be an heir and a testator. But we only have a testator, the physical property, and no heir. This is because the functional property is second-order. It is not a property on the same footing as the realizer property. Just as a policeman cannot inherit the fortune of Jones if Jones *is* the policeman, i.e., if Jones fills the role of a policeman, so it is impossible for a second-order property to inherit the causal powers of its first-order property.[13]

Schröder's argument seems to come down to an argument with the following structure:

a. Let us suppose, with Kim, that CI2 is correct.
b. Second-order properties inherit the causal powers of their first-order properties. (From assumption a.)
c. But second-order properties are identical with the causal powers of their first-order properties.
d. Thus, second-order properties inherit themselves. (From b and c.)

However, d is absurd. Therefore,

[13] Schröder (2002) p. 321.

CI2 is incorrect.

Naturally, the question is what Schröder's grounds for premise c are. Presumably, premise c is meant to be analytic. This means that, in order to evaluate Schröder's argument properly, what we need to determine is whether the notion of second-order property assumed in CI2 (and therefore used in premise b) is the same as the notion spelled out in premise c.

The idea that mental properties are second-order properties arose within a functionalist understanding of the mind. On this approach to the mental, the attribution of mental states to a given system only makes sense under the assumption that the system in question has a set of states that are causally related to perceptual inputs, behavioral outputs and each other in a certain way.[14] (To abbreviate: They have a certain 'causal role'.) Now, we can distinguish at least three senses in which a given property may be said to be a 'second-order property'. All of them involve the notion of causal role, and some of them will be familiar from the literature on different varieties of functionalism:

2P A property Q is second-order in that:
Q is the property that plays a certain causal role.[15]

2C A property Q is second-order in that:
Q is the property of playing a certain causal role.

2M A property Q is second-order in that:
Q is the property of having some property that plays a certain causal role.[16]

[14] See Putnam (1991), for example.

[15] This notion is used in the kind of functionalism known as 'realizer functionalism', which has been defended by Lewis (1991). Notice that calling a property a 'second-order' property in sense 2P is a little misleading if one is assuming a 'sparse', as opposed to a 'latitudinarian' or 'abundant', conception of properties. There are, in the 2P sense, no sparse second-order properties over and above first-order properties. There only are first-order properties, some of which can be referred to by second-order predicates.

[16] This notion is used in the kind of functionalism known as 'role functionalism'. See Block (1990).

To illustrate, consider the following picture about the relation between a subject's being in pain and the rest of her properties. The picture in question involves the following three properties of the subject. There is, first of all, the property of having C-fibers firing. This is a property of the human subject. Let us call it P (for 'physical property'). Then, there is the causal role that property P plays within the subject's cognitive economy. To simplify, let us assume that this causal role simply comes down to the property of being typically caused to be instantiated by tissue damage and typically causing wincing. Let us call it C (for 'causal role'). Clearly, C is a property of P so it is different from P. Finally, there is the property of having some property or other with C. Let us call this property M. M is, like P, a property of the human subject so it is different from C. In addition, M is a property that the human subject could have even if she did not have P (so long as she had some other property with C). So M is different from P as well. Now, how does the subject's being in pain fit in this picture? If we try to make it fit by construing the property of being in pain as a second-order property, then it will fit in different ways depending on what notion of second-order property we use. If being in pain is second-order in sense 2P, then it is identical to P. For P is the property that plays the appropriate causal role, namely, being caused by tissue damage and causing wincing. However, if being in pain is second-order in sense 2C, then it is identical to C. For C is the property of being caused by tissue damage and causing wincing. Finally, if being in pain is second-order in sense 2M, then it is identical to M. For that is the property of having some property that is caused to be instantiated by tissue damage and whose instantiation causes wincing.

Generalizing on the example above, we can distinguish three readings of the claim that mental properties are second-order properties, depending on whether the locution 'are second-order properties' in M2 is understood along the lines of 2P, 2C or 2M:

M2P If a property Q is mental, then:
Q plays a certain causal role.

M2C If a property Q is mental, then:
Q is identical to the property of playing a certain causal role.

M2M If a property Q is mental, then:
Q is identical to having some property that plays a certain causal role.

Let us now return to the question of whether the notion of second-order property used in CI2 is the same as the notion used in premise c. Clearly, the notion of second-order property used in premise c must be the notion captured by 2C. Otherwise, premise c is patently false: Second-order properties in sense 2P *have* causal powers and, therefore, they are not identical to those powers. And second-order properties in sense 2M are not properties of *other properties*, which means that they cannot be identical to the causal powers of their first-order properties. Thus, the notion used in premise c must be 2C.

What notion of second-order property is assumed in CI2? The answer to this question concerns the above-examined role that the principle plays in Kim's reductionist program. As we have seen, CI2 is grounded on a more general argument against non-reductive physicalism. Recall that the argument in question is meant to be a *reductio* of the view that a certain collection of theses is consistent, one of which is property dualism or anti-reductionism. We have mentioned that the reason usually produced by non-reductive physicalists in support of the anti-reductionist claim 1 is precisely that mental properties are, unlike physical properties, second-order properties. So the notion of second-order property assumed in the causal inheritance principle must be the same as the notion that non-reductive physicalists are operating with when they defend the anti-reductionist thesis.

Now, the notion of second-order property that non-reductive physicalists operate with when they argue in support of anti-reductionism is 2M. The idea that non-reductive physicalists seem to have in mind when they claim that mental properties are second-order properties is best captured by M2M. For their view is not that mental properties are causal powers of other properties, but that mental properties are identical to having some properties with certain causal powers. Ned Block, for instance, specifies the relevant notion of second-order property for functionalists (or, more specifically, those functionalists who are non-reductive physicalists) as follows:

> A property that consists in the having of some properties or other (say first-order properties) that have certain causal relations to one another.[17]

As a matter of fact, Kim is very explicit about the fact that this is the notion of second-order property that he is assuming in his argument:

> Let D be a set of first-order properties: a second-order property over D is the property of having some property in D satisfying a certain specification C. Where C involves causal relations (that is, C specifies a causal role) we may call the second-order property a functional property. Properties in D satisfying C are the realizers of the second-order property in question. […] Notice that a second-order property and its realizers are had by the same entities.[18]

Thus, the notion of second-order property assumed by Kim in the causal inheritance principle is 2M. Where does this leave us *vis à vis* Schröder's argument? It means that the notion of second-order property used in premise c must also be 2M in order for the argument to have some bite. But, as we have seen, it is not. The upshot is that the argument relies on an equivocation of the expression 'second-order property'. For that expression is used in sense 2M in premise b whereas it is used in sense 2C in premise c. There does not seem to be a way of fixing this problem. If Schröder uses the 2C notion of second-order property in premise b (and, therefore, in the initial assumption a), then the principle reduced to the absurd is certainly not Kim's causal inheritance principle, and it is hard to see why such a principle would have any intuitive appeal in the first place. And if he uses the 2M notion of second-order property in premise c, then premise c turns out to be obviously false. Either way, the causal inheritance principle does not seem to be vulnerable to Schröder's argument.

4. Does the supervenience argument generalize?

Let us now step back from Schröder's challenge to Kim and consider whether the supervenience argument actually generalizes. So far we have granted, for the sake of the argument, that it does. This is why Schröder's attack on the causal inheritance principle

[17] Block (1990) p. 45.
[18] Kim (1997) p. 290. See Kim (1998) p. 20 and p. 82 as well.

seemed to have substantial consequences. But an interesting question about the supervenience argument on its own right, one might argue, is whether the argument generalizes or not. Our contention in this section is that this question turns out to be ambiguous. In the background of this question, there seems to be a certain picture of the world according to which the world is somehow layered. Essentially, the idea is that the world is divided into different groups of entities organized hierarchically. They correspond to the domains of the various sciences, beginning with the elemental particles which physics studies at the bottom, all the way up to social groups studied by a discipline such as economics or sociology. But this picture collapses an important distinction.

Kim has pointed out that two different hierarchies are implicit in this model.[19] First, there is a hierarchy of objects divided into 'levels' according to the part-whole relation: Objects belonging to a given level are parts of those objects that belong to higher levels. This mereological hierarchy begins at its lowest level with the elemental particles of physics, moving up to atoms, molecules, cells, organisms and social groups. In addition, there is a hierarchy of properties stratified into 'orders' according to the realization relation. The properties that we take to be 'first-order' properties are those individuated by their causal powers. Examples of first-order properties include, at the level of organisms, the above-mentioned property P of having c-fibers firing. Then, there are second-order properties in the 2M sense. Continuing with the example above, suppose that having c-fibers firing typically causes wincing and the firing is typically caused by tissue damage. Being in pain can then be construed as a second-order property, that is, the property of having some property or other that it is typically caused by tissue damage and it typically causes wincing. Notice that this property is at the same level as P, namely, the level of organisms. We can now appreciate that the original generalization worry involves, at least, two different concerns. The first one is a concern about the drainage of the causal powers of all second-order properties (in the 2M sense) to their first-order realizers. The second is a concern about the drainage of the causal powers of all properties at high levels to properties at lower levels.

[19] In Kim (1998), pp. 80-82.

Kim grants that the supervenience argument yields the result that the second-order properties of objects at a given level do not cause anything over and above the first-order properties of objects at that level. We side with Kim here. After all, the supervenience argument can be generated for any second-order property, not only for psychological properties. To the extent that one intuitively thinks of psychological properties as properties with causal powers that are distinct from physical properties (even though they depend on them), it seems that one will also construe biological or chemical properties as properties with those same features. And one will then be able to run a version of the supervenience argument for those properties as well. (Notice that two of the five claims in the argument do not even mention psychological properties.) Thus, the coherent reductionist position to take here is to accept that reduction generalizes to all second-order properties. This is precisely the result that the causal inheritance principle is meant to capture.

What about generalization across levels? The question is now whether the supervenience argument shows that first-order properties of objects at a particular level do not have causal powers over and above first-order properties of objects at lower levels. The debate on what the right answer to this question is concerns the relation that holds between those properties. Kim claims that a property at some level is 'micro-based' on certain properties at lower levels:

> Property P is micro-based on properties $P_1 \ldots P_n$ and relation R when the object *a* that has P fully decomposes into parts $a_1 \ldots a_n$ which have $P_1 \ldots P_n$ in configuration R. We may call $P_1 \ldots P_n$ the 'micro-properties' of P and $a_1 \ldots a_n$ the 'micro-constituents' of *a*.[20]

The issue is now whether one can run the supervenience argument to reduce micro-based properties to their micro-properties, and considerable attention has been devoted to the status of claim 2 (the supervenience claim) in that hypothetical argument. Kim claims that a micro-based property does not supervene on its micro-properties, since

[20] Kim (1998), p. 84.

supervenience is a two-place relation between properties of the same object.[21] Thus, if we try to run the supervenience argument for micro-based properties, we will not be able to get it started.

The first reaction in the literature has been the following. Some philosophers have replied that strong supervenience is not necessary to get the supervenience argument started. Other relations of dependence between properties, which hold between micro-based properties and their micro-properties, can be used to run it.[22] The suggestion is to replace the supervenience claim 2 with a different thesis that preserves the intuitive idea of dependence between properties: There cannot be a difference between the properties of two whole objects without a difference in either the properties of their parts or the ways in which those parts are put together. It has been proposed that this one-to-many relation of dependence holds between micro-based properties and their micro-properties, and it is therefore possible to run the supervenience argument by reading the supervenience thesis 2 in this way.[23] Now, instead of trying to settle the issue of whether the strong supervenience thesis is essential to the causal exclusion argument, we would like to highlight a different line of defense against the generalization worry which Kim himself has sometimes pursued. We propose that a case could be made for dropping the causal closure principle.

Consider a micro-based property P at some level L and its micro-properties P1 … Pn at a lower level L*. What would it take to show, by using the supervenience argument, that properties at L such as P must reduce to properties at L* such as P1 … Pn? We would need to show, at least, that a conditional equivalent to C2 holds for micro-based properties just like C2 held for mental properties. For that purpose, we would need a principle that requires the instantiation of properties at L* to be caused by the instantiation of other properties at L* (if that instantiation has any cause at all). But micro-based properties can be at any level above the level of elemental particles. Thus, what we would ultimately need is a principle stating that each level is causally closed. Presumably, the overall principle would be that, for any level Λ, if a property at Λ is

[21] Kim raises this point in an exchange with Paul Noordhof in (1999), p. 117. On the same point, see (1998), pp. 85-86 as well.
[22] See Bontly (2002), Gillett and Rives (2001) and Noordhof (1999).
[23] For details, see Bontly (2002), pp. 83-84.

instantiated and its instantiation has a cause, then it is caused by the instantiation of some property at level Λ. But such a principle seems highly counter-intuitive. It would require, for instance, that if I wave my hand to hail a taxi and I displace some molecules of air in the process, then my hand did not displace those molecules of air in virtue of its having a certain mass. Instead, the principle requires that the molecules of air must have been displaced by the instantiation of properties in the *molecules* that constitute my hand, which is counter-intuitive. This means that if we try to run the supervenience argument to reduce micro-based properties to their micro-properties, it will put no pressure on us to drop claim 1. We can drop causal closure instead.

However, Kim's appeal to micro-based properties has provoked a second reaction in the literature on mental causation. And this reaction needs to be addressed differently. Ned Block has objected that a micro-based property of some object is related by Kim's own notion of supervenience to those properties that compete with it for causal powers. It is just that none of the relevant causal competitors is one of its micro-properties. Let us explain. Consider an object O at some level L, and the following property of O:

> Fully decomposing into two non-overlapping parts, Part 1 and Part 2, such that Part 1 has a property P1, Part 2 has a property P2, and Part 1 is in a particular configuration R to Part 2.[24]

This is a micro-based property of O (call it 'micro 1-2'). Thus, micro 1-2 is at level L.[25] Now, suppose that Part 1 fully decomposes into two non-overlapping parts, Part A and Part B, put together in configuration R1. Part 2 also fully decomposes into two non-overlapping parts, Part C and Part D, put together in configuration R2. Suppose that Part A has property Pa, Part B has property Pb, Part C has property Pc, and Part D has property Pd. Consider the following property now:

[24] By 'non-overlapping', what we mean is that Part 1 and Part 2 do not have parts in common. More generally, for any objects x and y, let us say that x 'overlaps' y just in case there is an object z such that z is a part of x and z is a part of y.
[25] We will assume that Part 1 and Part 2 are not at the level of elemental particles.

> Fully decomposing into, on the one hand, Part A (which has Pa) and Part B (which has Pb) put together in configuration R1 and, on the other hand, Part C (which has Pc) and Part D (which has Pd) put together in configuration R2.

This property is a micro-based property of O as well (call it 'micro A-D'). So micro A-D, like micro 1-2, will be at L. Notice that properties such as micro 1-2 meet the necessary conditions to supervene on properties such as micro A-D. If an object has micro 1-2, then its parts decompose into some parts with their own properties. Object O, for instance, instantiates micro 1-2 by instantiating micro A-D. Furthermore, any object that has the latter property should have the former property as well. It is hard to see how there could be an object such that it decomposes into Part A and Part B (with their respective properties) put together in configuration R1 as well as Part C and Part D (with their respective properties) put together in configuration R2 and yet that object fails to instantiate micro 1-2. How could such an object fail to have, for instance, Part 1 as one of its parts once it has Part A and Part B put together in R1? It seems that, once a certain fine-grained decomposition of an object is in place, any coarse-grained decomposition of the relevant object is fixed as well.

The new generalization worry raised by Block is the following. It seems that properties like micro A-D preempt properties like micro 1-2 as causes. These are properties of the same object so they are at the same level. Furthermore, they are all properties of the same order, since none of them is construed by existential generalization over the other one. So this concern remains in spite of the levels/orders distinction. Basically, the worry now is that the causal powers of, let us say, my having c-fibers firing may drain away to my having some chemical configuration, and further to having some atomic configuration, and so on.

To address this worry, Kim has suggested that we can identify properties such as micro 1-2 with properties such as micro A-D:[26] If having parts 1-2 in configuration R *just is* having parts A-D in configurations R1 and R2, then the latter cannot threaten to take away the causal powers of the former. Block, however, has replied that there can be no such identities due to the possibility of 'multiple decomposition': An object may be

[26] In Kim (2003).

decomposable into non-overlapping parts in two or more different ways.[27] The thought is that if Part 1 and Part 2 can be decomposed in different ways, then micro 1-2 is not identical to micro A-D, since O can have the former property without having the latter. Our position is that the right answer to the question 'Do micro-based properties reduce to their supervenience bases if the supervenience argument is right?' may simply be 'It depends on the micro-based property.'

Two different kinds of properties are put forward as examples of micro-based properties in the literature on generalization. Properties like being a certain temperature in a gas, or having a mass of 1 kg, seem to be of one kind whereas properties like being a water molecule (or, perhaps, constituting a certain pattern) seem to be importantly different. Consider, for instance, having a certain temperature in gases. Such a property cannot be identified with any particular configuration of micro-constituents and micro-properties: No particular kinetic energy in each of the molecules in a gas, or relation among them, is necessary for it to have a determinate temperature. On the other hand, the property of being a water molecule is identical to decomposing into certain proper parts, namely three atoms bonded in a particular way, where each part has its own properties (being an oxygen atom and being a hydrogen atom). Call properties of the second type 'relationally structural properties' and properties of the first type 'non-relationally structural properties'.[28] It may be possible to push the generalization worry against Kim when it comes to non-relationally structural properties because multiple decomposition would be a possibility for the relevant objects. On the other hand, it may not be possible to push it for relationally structural properties because it seems that the relevant object can only be decomposed in a unique way. Thus, we should focus on what kinds of properties the special sciences are concerned with. To the extent that those are relationally structural properties, the supervenience argument will not threaten the causal import of explanations that appeal to those properties in the special sciences.

To conclude our discussion on generalization, we have seen that the concern that reduction generalizes to all the special sciences if the supervenience argument is right about psychology is ambiguous. The argument does entail that all second-order properties

[27] Block (2003), pp. 145-146.
[28] We borrow this terminology from David Armstrong, who draws the distinction in (1978), p. 71.

should be reduced to first-order properties. But that would not show that higher-level properties should reduce to lower-level properties. The argument does suggest that non-relationally structural properties in the domain of any special science are causally inefficacious. But that would not show that relationally structural properties are.

5. Higher-order properties and mental causation

Let us keep zooming out, as it were, from Schröder's argument against Kim's inheritance principle and consider now a different issue that our assessment of that argument left open. We claimed that Schröder attributes the wrong notion of second-order property to Kim. But the more interesting issue about second-order properties and mental causation, one might argue, is what notion of second-order property is the right notion to adopt. Our contention in this section is that, interestingly, each of the three notions above will serve different explanatory purposes quite well. But they will do so at the cost of facing different difficulties. This should become apparent by considering the virtues and defects of M2M, M2P and M2C in order.

Suppose that we endorse M2M. The view that if a given property Q is mental then Q is identical to having some property that plays a certain causal role has an obvious virtue, namely, it can nicely account for the intuition that multiple realization is possible. Multiple realization is taken to occur when two systems have a certain mental property by having different physical properties that realize or implement it.[29] The intuition that a given mental property M can be multiply realized in two different physical properties in different systems S1 and S2 is easy to account for if we take M to be the property of having some property or other that plays a certain causal role in the system that has M. For it certainly seems possible that, in S1, the property that plays the appropriate causal role is different from the property that plays that causal role in S2. As we have seen, though, the advocate of M2M is not going to have an easy time accounting for the intuition that mental properties are causally efficacious. The reason why Kim's supervenience argument is interesting is precisely that, provided certain assumptions, it

[29] Putnam (1991).

constitutes a serious challenge to the view that M2M is consistent with the intuitive causal efficacy of the mental.[30]

Suppose that we endorse M2P. Then, our predicament will be the converse to that of the advocate of M2M. The view that if a given property Q is mental then Q is the property that plays a certain causal role in the relevant system squares with the causal efficacy of mental properties neatly. Quite simply, mental properties are identical to physical properties if M2P is correct. Thus, mental-to-physical and physical-to-mental causation turns out to be not more problematic than everyday physical-to-physical causation. Assuming M2P, if there are reasons to be skeptic about mental causation, then there are reasons to be skeptic about causation *simpliciter*. And this is something that reductivists and non-reductivists alike will not easily concede. The problem for the advocate of M2P is multiple realizability. The intuition that a given mental property M can be multiply realized in two systems S1 and S2 is not easy to account for if we assume M2P. How could a single mental property M be realized by two physical properties in different systems S1 and S2 if M is the property that plays a certain causal role in a certain system? If S1 and S2 are such that the property that plays C in S1 is different from the property that plays C in S2, then it is impossible that both S1 and S2 have M. (If M is the property that plays C in S1, then S2 does not have M, and *vice versa*.)[31]

Suppose that we endorse M2C. Is there any advantage to that view over M2P and M2M? We believe that there is. It is sometimes argued that a further challenge to M2P is that it needs to face the so-called '*qua* problem'.[32] The objection is sometimes put in terms of whether, assuming that a mental property is a physical property that plays a certain causal role, the mental property will be causally efficacious *qua* the mental property that it is or not. The claim is that, even if we identify mental property M with the physical property P that has a certain causal role in the relevant system, the question whether this property causes what it does in virtue of its being a mental property or in virtue of its being a physical property remains open. This is the interesting question to

[30] For a different sort of worry about the causal efficacy of mental properties assuming M2M, see Block (1990).
[31] The advocate of M2C may, of course, try to either individuate mental types more finely or individuate physical types more broadly. These two moves are explored in Kim (1993a). For difficulties with them, see Fodor (1974).
[32] Jackson (1996), p. 25

address in the mental causation debate, the objection goes, and the advocate of M2P has made no progress towards an answer.[33] As we see it, the only way to make sense of the question at issue is by assuming that property M/P itself has properties, some of which are causally relevant for certain effects of its instantiation, some of which are causally relevant for other effects of it. If we think of M/P this way, then it makes sense to ask whether a certain property of M/P, say, its being mental, is causally relevant for a certain effect of its instantiation. Thus, the advocate of M2C has a way of making perfect sense of the *qua* challenge. For she is working within a framework where second-order properties are properties of other properties, which is what we seem to need in order to make sense of the question that is raised in the *qua* problem. As far as we can see, neither the advocate of M2P nor the advocate of M2C will be able to make sense of that question easily.

This means that, in so far as one sees the *qua* problem as a legitimate worry that can be raised against M2P, one should feel some intuitive pull towards the view of second-order properties that Schröder seems to have in mind. Nevertheless, it is not clear to us that M2C really does justice to the intuition that mental properties are causally efficacious. If M2C is correct, then the generalizations of folk-psychology, where we invoke such properties as 'being in pain' to causally explain someone's wincing, are fundamentally wrong. For properties like being in pain are not really causally efficacious for such effects as wincing. Properties of those properties, such as being a mental property, are causally efficacious for effects like wincing if M2C is correct, which seems quite counter-intuitive. Neither is it clear to us how the advocate of M2C would handle multiple realizability. Thus, the advocate of M2C seems to be better off *vis à vis* the *qua* problem but she does not seem to be better off *vis à vis* accounting for the intuitions of multiple realizability and the causal efficacy of the mental. The advocate of M2P can easily account for the latter intuition, but not the former. And the advocate of M2M can easily account for the former intuition, but not the latter.

[33] It is interesting to note that the move here is analogous to the traditional move according to which overcoming substance dualism does not solve many of the interesting problems of mental causation, since those problems reappear at the level of properties. According to the *qua* theorist, overcoming property dualism does not solve them either. For they reappear at the level of meta-properties.

6. Conclusion

Let us take stock. We have examined the causal inheritance principle, its role in Kim's reductionist program, and Schröder's argument against it. We have distinguished several notions of second-order properties and we have argued that Schröder's critique of Kim's causal inheritance principle relies on an equivocation of two of them. We have then examined the issue of whether, independently of Schröder's argument, the supervenience argument generalizes. We have distinguished three versions of this worry. Regarding generalization across orders, we have suggested that the argument does generalize. Regarding generalization across levels, we have also agreed with Kim that the argument does not generalize. The interesting worry about the generalization of the supervenience argument, we have seen, is the generalization to all micro-based properties. We have proposed that, in some cases, the micro-based property of having a certain decomposition will indeed have to be reduced to the micro-based property of having, so to speak, a more fine-grained decomposition but, in other cases, this will not necessarily need to happen. The crucial point, in each case, will be whether the object has a unique possible decomposition or not. Finally, we have taken a look at the reasons in favor and against each of the notions of second-order properties that we have distinguished while discussing Schröder's argument. And we have concluded that each of them will be quite useful to deal with some of the issues in the mental causation literature but none of them seems to be able to handle three of those issues satisfactorily.[34]

Suzanne Bliss *Macquarie University*
Jordi Fernández *University of Adelaide*

sbliss@scmp.mq.edu.au/ jorge.fernandez@adelaide.edu.au

[34] We are very grateful to Peter Menzies for comments on earlier versions of this paper. One of us also wishes to acknowledge support from a grant from the Spanish Ministry of Science and Technology for the project *Modal Aspects of Materialist Realism* (HUM2007-61108).

REFERENCES

Armstrong, D. M. (1978) *A Theory of Universals*, Cambridge: Cambridge University Press.

Baker, L. (1995) 'Metaphysics and Mental Causation', in J. Heil & A. Mele (Eds.) *Mental Causation*, New York: Oxford University Press, 75-97.

Block, N. (1990) 'Can the mind change the world?', in G. Boolos (ed.) *Meaning and Method.* Cambridge: Cambridge University Press.

----- (2003) 'Do Causal Powers Drain Away?', *Philosophy and Phenomenological Research* 67: 133-150.

Bontly, T. (2002) 'The Supervenience Argument Generalizes', *Philosophical Studies* 109: 75-96.

Burge, T. (1995) 'Mind-Body Causation and Explanatory Practice', in J. Heil & A. Mele (Eds.) *Mental Causation*, New York: Oxford University Press, 97-121.

Fodor, J. (1974) 'Special Sciences, or The Disunity of Science as a Working Hypothesis', *Synthese* 28: 97-115.

Gillett, C. and Rives, B. (2001) 'Does the Argument from Realization Generalize? Responses to Kim', *The Southern Journal of Philosophy* 39: 79-98.

Jackson, F. (1996) 'Mental Causation', *Mind* 105: 377-413.

Kim, J. (1993a) 'Multiple Realization and the Metaphysics of Reduction', in *Supervenience and Mind*, Cambridge: Cambridge University Press, 309-335.

----- (1993b) 'The Myth of Nonreductive Materialism', in *Supervenience and Mind*, Cambridge: Cambridge University Press, 265-284.

----- (1993c) 'Mechanism, Purpose, and Explanatory Exclusion', in *Supervenience and*

Mind, Cambridge: Cambridge University Press, 237-264.

----- (1993d) 'The Nonreductivist's Troubles with Mental Causation', in *Supervenience and Mind*, Cambridge: Cambridge University Press, 336-57.

----- (1997) 'Does the Problem of Mental Causation Generalize?', *Proceedings of the Aristotelian Society* 97: 281-297.

----- (1998) *Mind in a physical world*. Cambridge, MA: MIT Press.

----- (1999) 'Supervenient Properties and Micro-Based Properties: A Reply to Noordhof', *Proceedings of the Aristotelian Society* 99: 115-118.

----- (2001) 'Mental Causation and Consciousness: The Two Mind-Body Problems for the Physicalist', in C. Gillett & B. Loewer (Eds.) *Physicalism and its Discontents*. Cambridge: Cambridge University Press, 271-284.

----- (2003) 'Blocking Causal Drainage and Other Maintenance Chores with Mental Causation', *Philosophy and Phenomenological Research* 67: 244-248.

Lewis, D. (1991) 'Psychophysical and Theoretical Identifications', in D. Rosenthal (Ed.) *The Nature of Mind*. New York: Oxford University Press, 204-210.

Noordhof, P. (1999) 'Micro-based properties and the Supervenience Argument: A response to Kim', *Proceedings of the Aristotelian Society* 99: 109-114.

Putnam, H. (1991) 'The Nature of Mental States', in D. Rosenthal (Ed.) *The Nature of Mind*. New York: Oxford University Press, 197-203.

Schröder, J. (2002) 'The Supervenience Argument and the Generalization Problem', *Erkenntnis* 56: 319-328.

Van Gulick, R. (1992) 'Three Bad Arguments for Intentional Property Epiphenomenalism', *Erkenntnis* 36: 311-332.

Abstracta 4: 2 pp.96-122, 2008

RULES, GAMES, AND SOCIETY

Martin A. Bertman

Abstract

'Game' means 'play within the construction of rules'. The sub-category 'sport' considers *play* as competition (in classical Greek, 'athletos' means 'competition for the sake of victory') where the rules are known to the audience, under the following divide: fundamental constructive rules about the game's structure and less important or flexible rules facilitating and monitoring play. These provide athletes and audience with stable knowledge. The excitement of play comes from the vagaries of the actual engagement of the rules in the action of play. The social order can use this as a metaphor of its ideal of civil law (and less sharply, of cultural custom) in relation to the citizen.

I. Description

Sport is a physical game, but the game structure can be applied to human activities aside from sport; specifically it can be applied to a legal system or order. Therefore, sport's defining characteristics must be given beyond the formal characteristic of being a game or universe of rules within which its activity necessarily takes place; further sport must be considered beyond its efficient characteristic of competition, since that also applies to activity within a legal system and many other systematized social activities. An appropriate characterization of sport should provide what sort of competition it is. A material characteristic is not helpful in this task since it is not the having of equipment that defines use, though it relates to a game in its specificity; but also, the material conditions, including equipment, vary greatly from game to game: Wittgenstein has brought to attention that games have their activity in a variety of formalized conditions, i.e. as board games or on some sort of field, say a track or a ring or a rink: a designed or ordered physical area; and also as composed of various numbers of persons, using living and not living equipment, etc. Consequently, the mentioned characteristics are too wide to define sport since competition within a universe ordered by rules includes other activities aside from sport.

Let us narrow the universe of discourse somewhat by considering sport needs judgments that warrants the action of play. Of course this still seems too large a universe of discourse to merely include sport, it still includes a legal order. The narrowing means that many games like chess or card games are no longer considered, since in such games the rules are known in principle to the competitors and no authority is necessary to judge any action within the game. Further, by the introduction of a judgment for a warranted action of a game, this means that any sport or, better, sport preparatory activity, that does not have a judge is a preparation or an exercise or practice for sport rather than sport in its primary sense as a competition among individuals or teams and let us now add, for the sake of victory, expressed by the superlative, the best, as measured in quantitative term by some measurement in points or a unit of length or time in a specific play of the sport. The Greek etymology of the word 'athlete' is competition for victory and it suggests the importance of the addition of the goal of victory to competition. Notably, the rather dull-minded Latin slogan of the Olympics, "*citius, altius, fortius*" (faster, higher, stronger), does not present a more informative value than the betterment of physical performance.

What is important is to differentiate the sort of judgment involved in sport from other judgments in systematic or quasi-systematic activities. The judge in sport is an referee or umpire, whose judgmental task is not primarily qualitative since this person who warrants the play must judge whether a particular play is in concert with the rules that constitute the universe of the sport; only in a secondary sense, when and if penalties are part of the task of the referee may there be a qualitative aspect to his judgments, that is, what "penalty" is deserved for a particular infraction of the constitutive rules. For the evaluation of the normative qualities of a particular sport it is well to apply the principle of John Dewey that means are not to be separated from ends. Consequently, since there are a number of ends that may be related to the mental and physical abilities tested by a sport, within the determination of its universe of rules – its constitutive basis – the means of the sport are the basis for value considerations of various sorts, even beyond the aptness of the sport to test particular natural skills; it is the basis for considering various social goals that the sport might habituate, like cooperation and competition, and it is the

basis for relating its formal understanding of play to the general understanding of mankind as cultural creature.

Preliminarily, it is to be noticed the purpose of sport as a competition between individual or of groups or teams of individuals with a referee to judge the play within an already asserted group of rules is a highly organized or political matter, where the framework for a sport is conventionally accepted conditions. Both sport and the legal order are political contexts which always in their construction rely on society; thus, one may expect the discussion of a particular sport to move to consider the specific culture that has created or adopted it.

But, now, we must proceed with the purpose of sport that differentiates it from other political contexts, including the jurisprudence. This can only the enjoyment of exercising the set of capacities involved in a particular sport, under the specification of competition leading to victory which is warranted by a referee. Games of blind chance are certainly not sport since there is no exercise of ability; on the other hand, card games, chess, etc. do exercise skill capacities, are competitions and have victory as a goal but they are not decided by a referee. Differentiated merely by a consideration of a referee, the similarity of games like chess and card games to sport suggests that they also are pursued under a condition of leisure and for the sake of refreshment. Games and sport are not work; the legal system part of the work of the political order. The reflexive nature of any skill activity should be noted, even when winning is a goal of an activity, the better one is at the play of these the more one enjoys playing them. This is true both of work as it is of entertainment but for entertainment the enjoyment of the activity is its uppermost consideration rather than its social affect; e.g., contrariwise, the work of the legal order is for the sake of the health of the social order: here, the primary or final normative measure is outside the activity itself.

There seems to be evidence for sport as work for professional sports where those so engaged are paid and do it not or primarily not to entertain themselves but as their work? Certainly no sociological classification would discount the business of sport with its various promotional and financial activities. These activities, however, are tangential to the activity itself in the way that being a physician, whatever his personal motivations, is

tangential to being paid. The actors in a sport – and one can even assume human being usually have various motivations for their doing something –may have various "personal" motives for engaging in their sport that are tangential to the exercise of capacity, for example, fame, wealth, social popularity, etc. Nevertheless, they are tangential to the activity like the physician wanting and/or receiving payment or social status is tangential to the activity of doctoring.

This is true in a definition of the activity even when the sport person or physician is engaged in sport or doctoring for the sake of wealth or social status. Other matters can be asserted of sport in a sociological description of its character and tendencies in a particular society that in some way are in tension with an ideal descriptive approach to the activity. The ideal is normative; it provides the best conceptual classification of sport in the essential clarity of what it is in itself and in distinctness from other activities.

Therefore, let us return to the character of the referee who as the guardian of the constitutive rules as they are projected into the unclarity of play, translated the essential character of the game into its temporal instantiation. At first sight, the referee may be taken as a minor consideration because he is not a player, especially when the referee is compared with the striving of the competitors in the sport. But he is the ideal spectator; he strives to assert the character of the sport instantiated in the play of a particular game. From the standpoint of general dignity, an individual strives for excellence in whatever he or she does. But athletic excellence, -- at least in the sense of striving to maximize one's abilities, -- is within the universe of rules of a sport that define allowable competition. The referee – who strives to translate the ideal into a particular instantiation: the game -- is necessary to that universe of rules that determines athletic striving because the referee is the authority that decides what is allowed in the play within those boundary rules; further, he provides penalties.

Penalty, like the western religious concept of sin, involves recognition of a fall from the ideal and, as justice, a return to the continuance of activity despite of this lapse. Sometimes the penalty is implicit in the disallowance of a play and sometimes the referee explicitly imposes a penalty: these matters depend on the sport or some condition within the activity of a sport. They are the details of conceptualizing justice within the conditions

created by the constitutive rules. The referee does not engage in the sport as a competitor yet his judgment can determine how various aspects of the game is characterized and even who has the victory. For the universe created by the constitutive rules he has a divine but not an all powerful or decisive role, in that sense he is a rather like a pagan deity who shares the universe with human actors and, though he can be decisive on an occasion of play, he must allow fateful competition: the total situation includes his role but his role, like that of the competitors, is determined by the ideal whose servant he is.

In this sense, he decides between two contending parties before him; like the judge when the judge is bound by rules. These binding rules assert the obligatory framework for play whereas his decision within that constitution, which has a qualitative freedom, may be called the regulative or particularized determination of a concrete sport's action. The regulative rules are less formal and are a response to the many possible occurrences of an actual game. The referee has a unique relation to the game by being the final authority for interpreting the actual action of play within the constitutive rules, thus his verdict is like that of a supreme court. It is noteworthy that even a photographic replay that shows the referee's judgment to be wrong does not (ideally) change the authoritative force of his decision since, human beings not technology are the instruments of its universe or ideal. (It takes a special pleading to relinquish judgment to instruments as a human product.) In sum, the referee does not determine the constitution of the sport, the constitutive rules, he applies them. But, this application demands a judgment about justice and judgment must have a definite end in the practical application of any system of rules. Noteworthy, for this symbolic consideration, a national legal system, -- except one given by God as claimed by Israel, -- can be considered to be, in a degree, a rather corrupt or distanced assertion of a monotheistic God's justice to a fallen world His creation but, sport justice is determined by a universe of rules created by man and, thus presenting itself as isolated, self-empowered or idolatrous universe, is isolated from political justice which is still within the monotheist God's domain.

From a less symbolic viewpoint, the constitutive rules that create the formal universe of the game are a social adaptation in the manner of a contract. They assume agreement between the parties competing in a game. Thus they must be clear and internally

consistent, like a law code (at least over a part of the jurisdiction of the law, say the matters of divorce or inheritance), setting up the universe within which the skill and competence of the players is exercised. Competition may be called *private* when it is with one's own self as a competition against one's past performance. Even the performance of a team or a player within the course of games, can be a private when it is merely a historical or statistical complementation against the performances of another individual person or team's past performance. Competition against one's own past performance is more an aspect of practice than of play. It is not the competition of an actual game. This is not to say that in some sense private competition may not have a higher value but merely to say that it does not have any value for sport, which is a public competition. *Public* competition is with another individual or team and it needs a referee to make judgments at the time of play. An individual runner, in actual competition, may lose the private competition to better the time of his run but win the race: actual competition is public. The goal therefore of the frame of a contract of play, the constitutive basis of a concrete action of sport, is to win. This is a zero-sum game in terms of outcome of actual games but a mutual benefit of contracting on the ideal level, that is, aside from or before the play, the joy of the play is offered by the ideal without being diminished by the result of victory which necessarily involves defeat. The ideal in this sense offers less to the victor and more to the defeated of the competition.

Supremacy in an instance of the artificial universe of sport is symbolic of human achievement through competition, even where, unlike sport, there are no constitutive rules. The enforcement of rules in many political aspects of society, as in the legal order instills the beneficial socialized habit where competition is responsive to the social will; it is the beneficial ground of a legal contract, where the parties are overseen by and participate in social agreement. The social will is idealized in a sharpened form in the formative contract or constitutive rules of creating government that creates and enforces a derivative legal code; just so, there is a need for a sovereign or referee to judge the non-ideal aspects that are necessary in the play under the original contract within the competition of play. The referee is a sort of "mortal God," in the sense that he can, as a

mortal, make mistakes but for the sake of sustaining the universe of the game he is authoritative in judging play.

Play without authority is warfare or chaos since there can be no settlement of concrete actions in relation to an ideal structure. The contracted structure in such a condition is a normatively inoperative contract or, better, treaty since it has no enforcement to settle disagreements. It is not a game since it has no ideal rules to be enforced but merely an idiosyncratic agreement between parties. It lacks a normative context. What Hobbes called a state of war among individuals is just this: the inability to settle disagreements without an authoritative decision. Also, in Hobbes, rules that define one nation's civil laws can not be applied to another nation: they are so to speak different games or universes; this is a condition of war among states, in principle. Without a referee who is the authoritative power over contracts, there are no operative constitutive rules. Competition continues and may be called mindless because there is no judgment. It is reduced to private practice, a sort of competition with oneself.

How is sport different as a competition than say a musical competition since both have competitors and also judges? The difference is that there are no constitutive rules for the musical contest. Thus even when it is a contest to see who can better play a piece of music composed by Chopin, where the musical score provides something of a constitutive context, there is a creative dialogue between the musical score and the interpretation of the player that escapes the notion of constitutive rules: it is creative within its limits. Here that interpretation is not, as in sport, a matter of action within regulative rules since the interpretation takes the place of regulative rules though it acts somewhat as if it were regulative rules in such competitions. Therefore, competition in the arts is different than in sport. Thus one has a judge but not a referee. Though a referee exercises judgment and in that sense is a judge, the judge has a wider function, as in a musical contest, he has the specific task of judging quality. Depending on the discretion given, a lower court judge in civil matters is either a referee or judge; but, a sovereign or supreme-court, the maker of final decisions is always a judge since quality – the ideal good of the political system or civil game – is his final prerogative.

In the arts, qualitative judgment involves creativity or some wide awareness of say musicality. Human beings always combine possibilities and this makes it difficult to fit essential definitions to the empirical evidence. In sport quality can be combined with quantity so in gymnastics and some ice-skating competitions there are set points given for certain physical actions combined with more subjective notions of grace, costume, and ingenuity or creativity of the composition of the activity, this last is more important in skating than in gymnastics. It is difficult therefore to consider these either as sport or not sport since the judgment that establishes victory is both referral and a qualitative judgment.

In the legal system as well there is usually a combination of judgments; there is a qualitative discretion of the judge or, in some cases, the jury, – sometimes jury and judge divide qualitative judgments of say guilt and penalty. In sum, the judge always has the referee's responsibilities to exercise a qualitative judgment within the constitutive rules of law, though he usually has a limited discretion in imposing penalties or a limited discretion in making an exception to a procedural or a substantive rule. The more such rules, however, are imposed on a judge by the legal order the more he or she is like a referee. When judgment is open to qualitative considerations beyond the constitutive rules, say the general social welfare or the ethical considerations of bringing a verdict, the judge is less like the referee. In physical competitions like dance-skating without a referee, a judge, though the judge may have some refereeing aspects in terms of a conventional view of performance, primarily assesses qualities in the performance and skills. Such competitions are unclear in principle in what makes for victory, despite broad conventional agreement and, therefore, they are sport only in a secondary sense, despite having the characteristics of competition. Of course, there is some qualitative judgment in sport, say in the assessment of a penalty, but the limited sphere of such judgments allows one to "civilize" the qualitative judgment under the hegemony of the definitional clarity of constitutive rules.

In certain activities taken as sport that especially involve equipment in the outcome of victory, e.g., hunting, car or bicycle racing, etc. the factor of quality shifts to the equipment so that as there is a competition about the superior equipment combined with

the skills in using them, from the essentialist viewpoint, there is a limitation on sport: it is no longer (just) a physical competition of human capacity. This ambiguity also occurs when there is a natural factor that is important in the outcome, e.g., in horse racing, water races in rapid and current filled waters, etc. These combine the fortune of having the best animal or the best natural condition with skill and as such it introduces variables into the ideality of the essential understanding of sport in the classical mode of a physical competition of human capacities, per se. All matters that depend on equipment and natural fortune of the course of the race are modifications of sport compared to the simplicity of a foot-race and most other of the original Olympic competitions.

Considering such "simple" Olympian sports as paradigmatic, if a defining characteristic of sport is having a referee; it seems that not every sport, particularly these quintessential ones, has a referee, e.g. a foot-race. Invariably, group or team sport has a referee. This suggests any organization composed of many individuals needs management in relation to the order for play made by rules, particularly, team play tends to have complex rules and, therefore, is particularly open to the judgment of a referee. This suggests a political principle: complexity demands authoritative determination for stabilizing and containing a determinative order. It also suggests the "gravity" of certainty to be achieved through constitutive rules though not the "gravity" of appropriateness in the political sphere where constitutive rules are related, beyond their own "game," to fundamental human needs.

In any case, a simple sport like running to arrive first at a goal needs little management. If the race is short, say a dash or a mile run, one might need to have a referee for deciding who in a "split-second" has broken the tape that marks victory. In a longer race, say a marathon, a "split-second" victory is rare to the point of vanishing; here one may need some inspection about the competitors running rather than traveling over some of the course by some other means. Policing action is a part of the referees function; it includes catching inadvertent and intentional breaking of the rules and, in consequence, he gives penalties. Nevertheless, in some very simple sport referral functions are absorbed by the spectators of the play. The community of spectators is usually enough to do the policing function and to monitor a sport where play is obvious; this situation is most

likely to occur when one particular physical attribute is decisive. But even here the referee exists in principle. The referee is the community that polices the play and authorizes the victory, if not a designated person has such authority. This is the democratic referee rather than the autocratic one. In the Ancient Olympics, as a religious ceremony, it is the divinization of democracy.

In sport, some analogies to the political universe of the state are obvious; especially important is the analogy to a legal system in the important and necessary (though not sufficient) condition of working within constitutive rules. One distinguishing characteristics of sport, and other games, however, is that competition within universe of the sport is done for entertainment or enjoyment. The Olympic games of Ancient Greece located the divine power of nature in all activity, including entertainments: sport, comedy, tragedy, musical and rhetorical contests, etc. This suggests that sociological investigation is necessary to understand how a particular society considers entertainment and sport as entertainment; and, further, the value a cultural context to consider competition and victory. Very competitive societies usually emphasize, in their way, the importance of victory and, for sport; this may be dysfunctional in terms of the ideal. Obviously every society needs cooperation. In team sports, there is not only competition between teams but also cooperation among the members of the team for the sake of victory. In a larger sense, a team includes trainers and coaches, as well as players. Preparation for play is an education or training like any preparation for a work task. Cooperation or collegiality, in its widest sense, is not only necessary but, it may be argued, that it can be more valuable as a human quality than victory and more valuable to social stability and even prosperity. Yet, victory is the supreme consideration in terms of sport. Sport reflects much of the heroic culture of Ancient Greece. Consequently, cooperative play and preparation -- team training and care -- enter into a consideration of sport, beyond a team's "Achilles," the outstanding player.

Further, something human is lost if the goal of victory is emphasized in a manner to obscure the fact of the grounding of it in a social order and its value within a range of human activities. Human beings are engaged in sport and, as human beings, as particular individuals, participation in sport is merely one sort of socialized activity expressing

human abilities and qualities. This is true of the state as well; the state relates to human needs but not to all the fundamental needs of human beings. Certainly, the state exists in a more complex and broader universe than sport; the state expresses and manages some aspects of the individual's humanity but it is not a complete determiner of the individual; its universe is not totalitarian in relation to an individual's humanity. Thus, the state ought not to be the referee or judge of all human concerns; similarly, the legal order ought to be interrogated ethical probity or the fundamental normative interest of the individual.

Now with a clear view of sport; it is appropriate to consider the definition of sport in terms of distinctness, even beyond what has been offered in the comparisons of sport with the state and with the arts. We must seek its place in the complete functioning of human beings. One question for this investigation is: can the values in sport of competition and cooperation be replaced by other activities that are superior to it, from a social view and from a view of the best functioning of an individual? Another is that granted sport has qualities useful to a social order how ought it to stand to other social activities? Answering such normative questions is difficult since one must relate sport to a theory of appropriate human action, *per se*. Before offering a few remarks about such normative considerations, since also they will follow an ideal rather than sociological or historical approach, let us consider the primary objections to the method we have employed so far, viz., that it is not empirical enough. In a word: that the analysis has not investigated how sport is pursued and considered in different places at different times by different groups of people. If objections to the essentialist method are weak and alternative methods of approaching sport are shown to be too limited, that strengthens the desirability to search for an appropriate theory of human nature.

Let us begin to respond by some methodological remarks. First, we have approached sport rather than a particular sport to find a conceptual frame to include any activity to be called sport whether it is known to us or not, whether it has existed or might exist. This level of abstraction strives for an essential knowledge of sport. The essentialist approach is a disputable one; it is primarily considered a disadvantage because 1) its generality is a normative abstraction which, aside from the gravity of cultural relativism and, also, the assumed objectivity of the social sciences, seems to halt further empirical

evidence, even for the sake of a descriptive reconsideration and 2) implicitly, its normativity demands the not very practical task of considering sport in relation to a theory of the ideal nature of human beings: such a theory being difficult and itself burdened by numerous if not all of the problems of philosophy.

Nevertheless, to pursue matters in terms of a social perspective (a) would be to accept that the conventional view of what is called sport. There is a loss of formal characteristics for the concept of sport. Consider: the Olympics has recently viewed ball-room dancing as part of its events; should the conventional authority of the Olympics make for the acceptance of ball-room dancing as a sport? Obviously, to give the Olympics such an authority of inclusion, per se, would halt asking for a definition: one would have to accept a social designation for what is called a sport rather than a conceptual ground. Since the conventional differs from culture to culture, a cultural designation would be time and place bound. Indeed, without rising to a level of abstraction that provides a unified conceptual understanding of sport, when questioning a perspective or cultural designation about sport, one encounters either a chaotic or a provincial intellectual condition. This approach soon finds the activities called sports cannot be brought together since they would share some but not all characteristics, and in various combinations, but not in any determinative manner. This precludes the finding of an adequate norm for sport in the activities themselves that are so designated by the culture. Consequently, the inclination is fostered to disregard the ideal or essential classification and to retreat into the particular for some sense of descriptive firmness whether of an activity designated as a sport or a social attitude about it. Ultimately, the avoidance of an essential norm is driven by a value skepticism about identity, an intellectual despair about finding an ideal definition; the social science approach to the identity of the activity, as it is committed to describing some general social reactions or beliefs, becomes a value relativism where the social characterization of sport is not in itself evaluated.

(b) The advantage of an essential definition, however, is to offer a determinative understanding, distilled from empirical material combined with thought about categorization: an essential definition should be seen as provisional rather than unchangeably determinative. But reform is an intellectual task depending upon better

conceptual procedures and an increased understanding of the empirical possibilities, rather than a mere change of conventional opinion. Further, an essential definition operates as a schema that clarifies the meaning of words. Logically, the appropriateness of the schema or system of concepts is tested by a coherent and systematic presentation of the subject. Practically, this is tested by making sense of factual matters. Definitional reform depends on being cautious about entrapment in habitual and socialized considerations; particularly those that are formally asserted. Often a criticism against the essentialist approach argues against it in terms of an overly formal use of abstraction. Yet abstraction is something that always occurs; the opponents of an essential approach also abstract, consequently this criticism is vulnerable. Too much abstraction is often an accusation of the lazy mind. The opponent's position in relation to the essentialist approach is analogous to the method of chaos theory, which recedes to the always smaller until the momentary focus in this infinite process also becomes an "abstraction" whereas the essentialist method is similar to geometry with its firm definitional lines for abstraction. Chaos theory is asymptotic: any presentation of coastal outline, by an exclusionary focus on a detail of the presentation, can be further considered as a whole new geographical region, – what I liken to a new abstraction – and this is open to being done without end. When the empirical is not considered by an essential definition that evaluates and classifies an actual activity this makes a theoretical consideration fragile in its openness to the relativity of choice. Thereby, it is a mere perspective classification that is rudderless in the theoretical sense and such an evaluation is open to skeptical erasure because it is grounded in a subjective viewpoint for the classification. Sociology tries to avoid this skepticism by a conventional or statistical attitude of sport in terms of some choice of a "universe" or group; large or small. But because the choice is arbitrary in a larger sense than a nominal essential definition by not seeking a coherent and consistent presentation possible to its subject, it can always be disputed by an alternative selection or subjective fiat: there is no stable conceptual norm to determine any choice.

How sport has a relevance to society is a legitimate question; it asks for the norms properly to consider sport within a social context. This provides the temptation for sociology to take a specific cultural universe and find classifications merely in the

conventions of that universe. It is like taking merely the sociology of jurisprudence in a particular state to stand for justice in an ideal sense. The question of the relation of sport to a specific culture is better answered when one has an essential definition of sport since that avoids merely measuring social attitudes. Yet, when one probes the relation of sport to other social and human concerns, one must be content for a preliminary reasonable rather than a rational determination. The enormously complex task of essential definitions of culture or of the political order or of human activity in relating these to sport moves one toward provisional assertions. Provisional and incomplete in terms of an adequate discussion, yet the essential characterization of sport provides some observation on culture and the human.

Some sports depend primarily on physical endurance and others on physical dexterity. In fact, the simpler competition, the testing of a particular physical capacity, the closer the competition is to the ideal of sport. From the point of view of exhibiting and perfecting some quality or skill each sport must be considered in terms of its demands upon players. Yet, considering the characteristics that have been described of sport there is some psychological matters that are general to sport. One is competition for the sake of victory. Of course a person may engage in a sport and not care if he is victorious. Under our characterization of sport as competitive, the individual is taking the activity as a sort of practice, whether or not the activity results in victory for him or her.

There is a distinction, however, between competition and rivalry. Rivalry, as I contrast it to sport competition is an outdoing of competitors without any appreciation of them, that is, an attitude of considering them to be an enemy. Roman gladiatorial contests because there is a seriousness involving extreme bodily injury or death certainly and appropriately falls under the consideration of enemy. Interestingly, it was engaged in by slaves rather than free men whereas the requirement in Greek athletic contests was that one must be free. In a gladiatorial mode of rivalry one can appreciate the competitive skills and capacities of the opponent but that appreciation is overwhelmed by the awareness of the possibility of serious harm to oneself. Rivalry is such a competition or battle but a vice when it occurs in sport. It is a vice of the extreme sort by overemphasizing the worth of the outcome. Yet not caring at all about the outcome is

when competition is abrogated; an attitude of exercise is the "vice" or limit on sport in the opposite direction. Here there is no concern with victory through competition, which under emphasizes the worth of the outcome from the viewpoint of sport.

The psychology of competition must therefore be examined. Proper sport competition appreciates the skills of the opponent and, further, appreciates the activity; thus, a sport person forms a bond with the opponent in the sense that both engage in the sport with its particular difficulties and demands upon one's skills and qualities. Indeed, there is usually a community formed for a sport that includes in the first instance other participants, including trainers and coaches as well as the athletes. And second, followers of the sport, particularly knowledgeable ones like those that report about the preparation for and the action of sport. Thus there are various communal and cooperative endeavors for sport. The competitive attitude is modified within these contexts. Further, the social psychology of the audience in terms of competition is a factor that stands aside from the formal order of a sport. How is soccer received in Afghanistan or baseball in Japan?

A competitive attitude is a habit that relates to a social context that encourages or discourages the habit, competition must be considered in a cultural context and this consideration must be examined in relation to such aspects as appreciation of physical skills, entertainment and, especially, the social view of cooperation and community. It should also be considered in terms of a societies appreciation of sport as a symbolic activity, that is, the sophistication of appreciating activities as an expression of some other value, whether it is nature or – obviously as it has been the origin of many sports with its training and honor codes – war. In a sociological description of a particular sport, the habit of competition is encouraged in preparing the individual for competition but also there are habits and attitudes about cooperation and community that actually limit what a culture considers sportsmanlike behavior. It may be that some cultures only have a sense of rivalry and no sense of sport; here competition has the gravity of victory at any price. Indeed, as social honors and wealth for the victor become unbalanced in their importance to the activity relative to other activities in a culture, rivalry takes the upper hand; sport becomes serious "business" in terms of the appropriation of social goods. On the other hand, there are cultures where sport is not considered at all because of a general non-

competitive ideal or a dismissal of the "non-serious" character of such activities; here the activity of sport, if any such exist, is considered to be something like exercise.

Many benefits in life seem determined by competition and the habit of competition fostered or strengthened by sport is valuable. Aristotle put it: "The race does not go to the swiftest, but to the swiftest of those that compete." Further, the cooperative and community aspects of sport obviously have a preparatory social benefit as well. Certain skills and physical strengths were as well particularly useful for societies that often fought wars in terms of hand to hand combat and especially primitive weaponry. In conclusion, by a preliminary understanding of the relation of sport to culture, it seems that sport provides a balance between the individual as a person who strives for victory and an appreciation of the social context of that striving, especially the elements of cooperation and the rules of the game. The latter especially institutes an appreciation of lawful behavior under some form of governance.

***

II. Sport: Society and the Law

Let us keep in mind a useful distinction about sport competition, as victory depends on the lawful condition of constitutive rules that define a game within which one strives against other human beings. Unlike sport, there is no lawful attitude in rivalry. I call rivalry, the striving against others, even animals, for victory at any cost, even outside rules of play, if there are any. This distinction presents two opposing directions toward proper and improper concepts of sport. Rivalry is broadly exemplified by the gladiatorial contests of Rome[1], where the gladiators were slaves, formally owned by the Emperor. These gladiatorial contests were embellished by increasingly extravagant confrontations, even men against animals. Here competition is against the rawness of nature, even ending in death, and reminding one that death is what in the end defeats all human aspirations. It is the clown-show of tragedy in a culture molded by the tragic sense of life, though its presentation may be in the direction of a circus extravaganza. It suggests an aspect of

[1] The *numera* (contests) first appeared in 246 BC when Marcus and Decius Brutus, in honor of their father, had three duels. Julius Caesar in 65 BC to celebrate his aedileship had 320 duels. The Flavians, Vespasian and Titus, with the gold taken from Jerusalem, built the Coliseum in 80 AD, seating 50, 000 and held contests for 100 days. Occasionally freeborn men participated in the duels but that was relatively rare.

culture that is in opposition to the cooperation and appreciation of free men in the qualities that are open to the vigor of life. Free men, even in the Roman period, were the only permitted competitors in the Olympics. Freedom is the moment of power in its exercise and competition is a demand of community, symbolized by the referee, the enforcer of the rules or laws of the game, especially in team sport where cooperation is necessary in the play for victory. Competition, to assert skill and human physical strength that has been trained or cultivated, is the endeavor of nature combined with artifice in harmony. The distinction between sport competition proper and rivalry suggests the difference between a culture of freedom and a culture of slavery, between a celebration of human power within its limits and a direction toward brutality.

Sociologically, this conceptual tidiness is merely a cultural direction; actual sports are not so clearly distinguishable. Yet, it is a helpful measure for understanding even today's actual sport situation, one not attached to the ancient Greek or Roman pagan religion. For the modern spectator of sport, the spectator that leans toward rivalry is moved by the gravity of a fanatic fan where victory is everything and taken to be a personal triumph. Perhaps one that "spits in the face" of loss and destruction: that asserts the battle of man against the forces that oppose his continuance, especially nature. And, it takes the completed game, its victory as triumph. But also there is the other sort of fan, I shall call this spectator the connoisseur, to whom the competition is appreciated in the skill and qualities of those that engage in the sport. Interestingly he actually must watch the sport event rather than merely get a report of its result; this suggests intelligence rather than emotion.

Though what is called a sport may lean toward one or the other sort of spectator, soccer provides an instance of both sorts of fan; some of the audience behavior recalls a gladiatorial contest and yet the game itself is highly cooperative under constitutive rules, without much openness toward brutality. With the rival partisan and brutal spectator, we have an example of a cultural assertion toward violence to even law-bound activities. When victory is the primary aspect of success, say also in such social activities as politics or business, then in the atmosphere of rivalry one is open either to break the law, the Mafia response, or corrupt it in one's favor, the "special interest" or advertising response.

In rivalry the referee, as a guarantor of the constitutive laws in relation to the games action is diminished in importance. When the games rule limitations is not important, which is a contradiction to play, it is used as the mere momentary condition for victory at any price. Of course, as I shall discuss it further, the culture or the social order's relation to the legal order is essentially more open and ambiguous than the constitutive basis of a sport game but the psychological attitudes between excelling within rules and winning at any price may be a social psychological atmosphere of all human activities.

Should one consider culture's where sport does not exist or where it is given a very marginal value as poorer because of that? This is not an easy question to answer. There are many who argue that sport is a catharsis for emotions or a preparation for life or war in a variety of ways or an artistry of the skilled body, etc. These can be considered in terms of social usefulness. The body of course is a condition of human life and, therefore, an aspect of our humanity open to deep consideration and sport is captured by the evaluation of the natural and supernatural descriptions of value. In Indian culture, yoga involves discipline of the body but it is not competitive and consequently is not sport; the culture of yoga is discipline to evolve the individual consciousness to grasp the unreality of embracing the self as a competitor. Hebrew Biblical culture is not opposed to sport but it certainly is considered not to be a serious matter and hardly mentioned: Solomon Schechter writes a book called *The Jew and the Gentleman* which contrasts the English Eton sport attitude with Jewish seriousness about rules as God's commandments. The holiness of the person asked by monotheistic Western religions denigrated the pagan elevation of sport within its natural religious orientation. Origen, living in the sphere of Roman culture, (d. 257) used the notion of rivalry in a Christian opposition[2] to Satanic and natural bodily temptations, calling the martyr "God's athlete." For the most part, Christian cultures as they now exist in the West give a greater allowance toward sport and

[2] The Christian position was expressed by Tertullian in *De spectaculis* which condemned the gladiatorial contests for its cruelty and its pagan religious association. Novatian called it "*Idolatria . . ludorum omnium mater est*." The gladiatorial schools were closed by Honorius in 399 A.D. Among the Romans Cicero favored them for encouraging fortitude among the audience and Ovid for their beauty but Seneca wrote of them that they were "*crudelior et inhumanior.*"

therefore are culturally more dependent on the pagan attitudes of the Greek Olympian[3] and the Roman Gladiatorial traditions. A delicate sociological appraisal cannot dismiss these historical strains as they affect modern cultural attitudes toward sport in their diverse and tidal presence.

The pagan religious culture of Greece and Rome made nature divine and consequently elevated the possibilities of the human body. Like its sculpture it saw in sport an ideal of the functioning of man's natural capacities; further, nature provides much evidence of competition, especially of rivalry. As I have suggested, rivalry is more appropriate for the animal world, despite striking instances of social cooperation in species like the bees and the ants, and sport competition is uniquely a possibility for a rational being that understands competition and cooperation within the lawful as he creates it for himself. As said, it is the realm of free men; its ideal is open to rationality qua autotelic orientation. Yet, some religious cultures denigrate the natural by subordinating it to a supernatural creation and a supernatural and/or an extraordinary spiritual destiny for the individual. In such a culture sport is at best tolerated as a natural aspect of man. But also it can be considered a distraction from more serious "spiritual" pursuits or even a secular expression of psychological energy and the loyalty that follows the ebbing of the supernatural religious orientation.

The rules of sport are after all invented by men and this may be seen as a rivalry with say the laws given by God in the Hebrew tradition or the divinity even of the laws of nature as they are imputed to be God's instrument. For the Christian the competition of sport is a pride of self and in opposition to the love among human beings; certainly a life in "imitation of Christ" seems at best tangential to sport: though Christian Universities in America have sport teams, religious orders and monks do not. That the rules are invented by man may be seen as speaking for a certain sort of secular humanism; a respite from nature and an expression of the capacity to organize and direct one's human capacities, in their testing. It can be seen as a paradigm for a rather Hobbesean politics where agreement to play is an agreement to obey or abide by the referee's interpretation of the

[3] In the *Iliad*, games are given in honor of Patroklos at his funeral. However subdued, there is a religious quality to sport among the Romans and the Greeks; noteworthy, as well, is the ritual sacrifices after the ball games of the Aztecs and Mayas.

rules that constitute the realm of the game. Sport is fundamentally a secular condition, though since nature beings engage in it, so by an interpretation of man's natural capacities, sport can come into a relation to some sort of view of the divine.

To speak in a general way, within Western Culture, with its tension between a secular naturalism and a supernatural religious orientation, sport is taken seriously and not seriously by in some measure reflection deeper cultural tensions. When it is not taken seriously, but allowed, its gravity is toward entertainment. Perhaps in the wide sense of Aristotle's view of art as a catharsis of certain emotions or a sort of social training. This viewpoint arises from time to time but is psychologically dubious; for example, the famous ethnologist Konrad Lorenz writes in 1966: "While some early forms of sport, like the jousting of knights, may have had an appreciable influence on sexual selection, the main function of sport today lies in the cathartic discharge of aggressive urge."[4] Yet, in 1974 Lorenz is quoted as saying something quite other, "Nowadays, I have strong doubts whether watching aggressive behavior even in the guise of sport has any cathartic effect at all."[5] The psychological value of sport seem to me to be insecurely considered because it does not distinguish between sport in the proper sense of competition as a proper function of the individual in his relation to society and sport in the improper sense of rivalry, with its outer edge of brutality. Consequently, without such a rather Aristotelian distinction in mind, that of proper functioning from a naturalistic view, the social sciences commit themselves to describe the values of that arise from a variety of improper social functioning. It is the same reason that if one takes merely the workings of any legal system as all there is to define justice; one cannot speak about unjust laws, except in the trivial sense of being not properly forged and promulgated. But, if it is appropriate to speak about the "unjust legal laws" in terms of another sense of justice, as well it is appropriate to speak about certain sport or a certain reaction to sport as not defining the proper functioning of those sorts of things.

Further, this holds true for any entertainment; an entertainment often is related to the sport simply or to some aspect of it, especially in activities that are marginally sports, like costumes and musical choices in dance-skating competitions. Entertainment from a social

---

4 Lorenz (1966:242)

5 Evens (1974:93)

description of some activities, in some cultures, suggests a directional emphasis on spectacle; the gladiatorial combat as circus, which now is represented by "commercial" wrestling and mixed forms of fight competitions. The sport of rivalry is quick to attack entertainment in this sense to what it considers sport. Entertainment, in the Aristotelian sense as a respite and preparation of the emotions for work is open to various interpretations yet, conceptually, it is other than work. It is unclear whether sport, in the aspects of rivalry, is a catharsis of certain emotions or a practice for them, if it releases anti-social energy or it is a practice of that energy, perhaps it is both in the sense of disciplining aggressive energies for war. Confusingly for a clear line between sport and rivalry, the commercialization of entertainment has made it a work for those that engage in it. In the sense of livelihood its participants become inclined, say as in soccer, to have an appetite, like some of its audience, for victory at any price. The activity becomes orientated to the collection of statistics as an indication of job performance rather than toward the game's aesthetic.

When work in a culture is or is considered drudgery entertainment has a heightened value, however, when work is considered a "*Beruf*" or "*confessio*" that is, some secular or a religious expression of the way the individual defines himself or herself, than entertainment is taken less seriously or confusedly considered to be a sort of work. In any case, there is certainly some truth in the Marxian viewpoint, historically evidenced by the Imperial Roman use of physical contests, that, for the same reason of social control, sport like religion can be considered the opiate of the people. These are overarching cultural considerations that need a fine analysis, but I simply say that the characteristics of sport in general and particular sports indeed have many interesting relations to culture. And here I take culture not in an ideal or essential sense, which is possible yet very difficult, but merely in the historical and sociological sense.

The relation of sport to the legal order is an aid to understanding how modern culture takes sport because it shows certain limits that respond to sport situations, especially aspects of violence. Hannah Arendt says, "our terminology does not [often] distinguish between such key words as 'power,' 'strength,' 'force,' and finally 'violence'

–all of which refer to distinct phenomena."[6] This reminds us to carefully segregate the concept of violence from other assertive concepts. In practice this can be difficult in sports like football, hockey, and boxing where the very action of the sport involves highly aggressive behavior. In these "contact" sports the concept of violence seems an aspect of play. It seems to relate this to the legal sphere one must think of a further distinction where violence degenerates to something aside from what is acceptable to the sport, where it degenerates to brutality. It is here that revocation of a license to play and criminal penalties can be best considered; and, less seriously, where the prerogatives of a referee exists to sideline players for a period of time or to impose fines on the team.

Let us consider some actual legal decisions in relation to this matter. In the case of Regina v. Bradshaw, in 1878, a British soccer player was charged with manslaughter when after charging and colliding with an opposing player the man died. The defendant was acquitted but Lord Justice Bramwell instructed the jury in a way that was cited in similar cases. He said, "No rules or practices of any game whatever can make that lawful which is unlawful by the law of the land; and the law of the land says you shall not do that which is likely to cause the death of another. Therefore, in one way, you need not concern yourselves with the rules of football."[7]

Interestingly, the orientation of Lord Justice Bramwell does not take into account the cultural allowance for a measure of violence that can cause harm in terms of "contact' sports. A short time later, in 1894, in the United States' city of Syracuse, Robert Fitzsimmons in an exhibition match struck his sparring partner with a blow from which the man died. The judge in this case directed the jury quite differently than in the previous British case: "if the rules of the game and the practices of the game are reasonable, are consented by all engaged, are not likely to cause serious injury, or to end life, if then, as a result of the game, an accident happens, it is excusable homicide."[8] Fitzsimmons was acquitted. This brings forward the issue of consent involved in the engagement of such sports. With this in mind, it is a practical measure which was adopted as a principle of law rather than the previous British example. Of course this seems apt if the culture

6 Arendt (1969:7)
7 Hechter (1977: 425-33)
8 Ibid.

tolerates the sort of sport that allows the possibility of serious injury. On the other hand, the notion of accident is the key concept and it limits the aggressiveness of play aside from an intentional criminality or uncustomary brutality coupled with the intention to maim or kill.

Sometimes there is a policing by a sport association to change rules that seem excessively brutal or leading to serious injury, for example, the rule change outlawing the V-formation in American football. This reflects a cultural pressure on sport. Yet, the cultural relation to the sport often allows a certain measure of brutality, which certainly is the case of Canadian ice-hockey. A few extra-play punches are tolerated. Yet in the 1969 game in Ottawa when Wayne Maki assaulted Ted Green with his stick injuring him to the point of near death there was something of an outcry which nevertheless did not result in a criminal penalty. That is evidence that the cultural strand I have called the gladiatorial attitude has particular force in sports like hockey.

Another legal aspect involves responsibility. In 1979, Rudy Tomjanovich of the Houston Rocket basketball team was seriously injured by a punch given by Kermit Washington of the Los Angeles Lakers. Not only was Washington suspended for 60 days and fined $10,000 but the Lakers as an organization was deemed negligent because they failed to adequately supervise Washington while knowing his tendency toward violence and they had to pay Tomjanovich 3.3 million dollars.[9] This asserts there are considerations that present responsibility for violence beyond play or the players.

Indeed, much violence occurs by fans and this extends the involvement of the legal system because of sport. Though there is no attempt, as is usual with such typologies, to discriminate between proper and improper reactions, the following typology given by Vamplew[10] is suggestive: violence may result from (1) frustration, for example with a referee's verdict, (2) criminality from anti-social activities, (3) remonstrance or the crowds use of a sport event for expression of political grievances, (4) confrontation where rival religious, ideological or ethnic groups riot, and (5) expressiveness due to intense emotional arousal, as in defeat. Since these are the behavior of non-players the anti-social behavior is not related specifically to the play of sport but rather to the cultural and moral

[9] Horrow (1981:9-18)
[10] Vamplew (1980:5-20)

standards of a specific cultural order. Yet, naturally there arises a query from sport crowd violence, i.e., does the specific sport, at least in terms of some of the above typologies, exemplify a culture's social psychology for this behavior or even promote it. This is a question from a descriptive social science viewpoint and not from the ideal position that discriminates between sports properly so called and not properly so called on the basis of a view of human psychology and society that has a decisive value measure. In this perspective, the application of rules, whatever they are, whether in sport or in other aspects of society may not be or appear to be fairly applied. How a culture responds to the frustration of perceived injustice is always an interesting issue. In the case of sport, the emotional stake may be related to betting on the outcome, adding a financial loss to other enthusiasms such as a fan's association with the destiny of a team. Financial loss is a thread that reaches throughout professional sports both to the players, teams and sport associations and, consequently, relates them in this aspect to many of the workings of a particular society, which involves the supervision of a state's legal system.

Despite that they seem open to being schematized either to universal of to socially specific conceptions, consent and responsibility or obligation are concepts that are not only important to define in terms of the legal system, as we have seen in the cases presented above, but in any human activity and so both in sport and society. These concepts have a greater firmness for players under the constitutive rules of sport. In sport the playing of the game is unambiguously the consenting to obey the rules of the game. Since the game is constructed through such rules – they are constitutive rules – doubt occurs merely in terms of the application of those rules as in the case of a referee's verdict. Those rules are presumed clear to all players and they are almost always relatively simple. This approaches the notion of rules formally but not functionally in the hypothesis that there can be an ideal of human functioning. This suggests for the legal system as it is involved with sport that it is clear that if the law allows the game it is committed to allow the result of play. Further, it has a possible role in relation to play when the rules of the allowed sport are broken. Of course, that involves a seriousness, – which is a social understanding – that goes beyond the jurisdiction of the referee who properly awards ordinary penalties that inevitably arise in the play of the sport; for

example, the use of a dangerous implement not allowed in play which causes harm is a criminal matter to be tried by the civil law. The responses of crowds to a sport event is purely within the legal jurisdiction of the civil law, since citizens consent and are obliged to avoid anti-social behavior and any excuse is merely to be taken within the openness of the civil law to the excuses that can be given by circumstance.

In terms of the legal system, from the formal notion of rules there is an important difference, in comparison to the rules of a game, of an openness to change that makes the decision less straightforward, and when considering the forces within the society to which the legal rules are applied, the laws of a state are occasionally quite ambiguous in terms of the perceived social goals for the legal system. A modern state's legal system is complex because the social action is complex and, moreover, open to changes that demand the transference of the intention of say a constitution, written many years before to an unanticipated circumstance. It has some of the qualities of a closed system, as in a game, but its function to protect social order, has openness. Thus there is a tension between its attempt to impose immobility on what is essentially a moving and changing reality. Further, since the legal system is responsive to social attitudes, whatever the form of legislative power, whether, to put it in logical formality, in the hands of one, some, or all of the adult members of society, these attitudes change with time because of internal development or external circumstances.

Consent to citizenship is an obligation to obey the law, whether proper or not by some ideal standard of evaluation, but unlike in sport, there is substantial difficulty in knowing the law. The complexity of tort law is notoriously obscure for the average person. Thus it is not the equity of application that is the only issue but the very understanding of the constitutive structure and, frequently, even the jurisdiction that a matter falls under. Sport may respond to the taste of an audience to change rules, say from an allowance of a dangerous strategy, like the V-formation in American football, but the legal orders relation to social pressure is more multifaceted. It is not only in the interpretation of rules but the jurisdiction of cases, the assessment of penalties, understanding liability, etc. and the transformation of the legal order in terms of disregarding certain laws, even when formally in place, and in making new laws.

Therefore consent and obligation for the citizen is divided between the actual civil law and a perception of justice or a proper perception of justice. This complexity indeed, allows the argument for civil disobedience. But there can be no such notion for sport: the referee is sovereign and to use the phrase of Thomas Hobbes, a sort of "mortal god" of the game in the sense that his ruling is decisive no matter any contrary evidence about the play. Therefore, if sport is symbolic of society, as many have argued, say in terms of certain socially perceived aspects of competition and cooperation, it is not to be considered quite like society: either society as an ideal or from the descriptive viewpoint of the social sciences. The problem of the legal system's attempt to conserve the social structure by rules, even in a conservative society, is open to the tension of change in social life facing often new or a rearranged emphasis on one or another competing value and, by time, new conditions to which the law must respond. Games also may change, but their constitutive rules are clear and distinct at any time whereas society reflected in the law has a measure of ambiguity at any moment. Indeed, the jury system is a technical devise that recognizes this difficulty for a socially appropriate verdict since it is a compromise between the formality of law – some procedural aspects being quite clear – and an informal grasp of social equity.

The difference between conservative and liberal attitudes among one's fellow citizens is reflected in how the law is viewed in its constitutive aspect as a system that serves society. Therefore, as well, consent and responsibility in terms of the law becomes a political question. It is a question of how the constitutive aspect of the legal order fulfills the necessary conditions of equitable cooperation, under an ideal or a culturally oriented norm of competition, and is open to the reasons for cooperation in the first place. This may be called a search for justice; which involves middle ranged abstraction in terms of social consensus and, possibly, the higher abstraction of a universal ideal for human beings. There is no analogue for this in sport. In sport the striving for victory is the clear value and the constitutive rules are agreed to. The existential character of society is more ambiguous. After all one is thrown by fate into a certain situation to which one must respond: one is not an athlete within a range of competence for the playing the economic and social "games" of one's society. As Anotole France put it in 1907: "the law, in its

majestic equity, forbids the rich as well as the poor to sleep under bridges, to beg in the streets, and to steal bread."

Martin A. Bertman
Helsinki University

References

Arendt, H. (1969) *On Violence,* New York: Harvest Books.

Evens, R.I. (1974) 'A Conversation with Konrad Lorenz', *Psychology Today* 8.

Hechter, W. (1977) 'Criminal Law and Violence in Sports', *Criminal Law Quarterly* 19 (3/4), 425-453.

Horrow, R.B. (1981) *Sports Violence: The Interaction between Private Lawmaking and the Criminal Law*, Arlington: Carrollton Press.

Lorenz , K. (1966) *On Aggression,* New York: Routledge.

Vamplew, W.(1980) 'Sports Crowd Disorder in Britain, 1870-1914:Causes and Controls', *Journal of Sport History* 7 (1), 5-20.

Abstracta 4 : 2 pp.123 – 147, 2008

BRAIN AND BEHAVIORAL FUNCTIONS SUPPORTING THE INTENTIONALITY OF MENTAL STATES

João de F. Teixeira and Alfredo Pereira Jr

Abstract
This paper relates intentionality, a central feature of human consciousness, with brain functions controlling adaptive action. Mental intentionality, understood as the "aboutness" of mental states, includes two modalities: semantic intentionality, the attribution of meaning to mental states, and projective intentionality, the projection of conscious content into the world. We claim that both modalities are the evolutionary product of self-organized action, and discuss examples of animal behavior that illustrate some stages of this evolution. The adaptive advantages of self-organized action impacted on brain organization, leading to the formation of mammalian brain circuits that incorporate semantic intentionality in their *modus operandi*. Following the same line of reasoning, we suggest that projective intentionality could be explained as a result of habituation processes referenced to the dynamical interface of the body with the environment.

1. Semantic and Projective Intentionality

Although discussions concerning the notion of intentionality have predominantly been part of the agenda of philosophers of mind, philosophers of language and cognitive scientists, the interest in the nature of intentional states has recently called the attention of cognitive neuroscientists and neurobiologists. One of the most influential neuroscientists of the XXth century, Walter Freeman, elaborated a philosophical view of brain functions based on an inner disposition for perception and action in the environment (Freeman, 1995).

In philosophical literature, the problem of intentionality can be dated back to Aristotle and Thomas Aquinas, who attempted to understand the inner dispositions of living beings to direct actions toward intended goals, instead of a sheer reaction to causal forces of the environment. The modern formulation of the problem was made by Brentano (1925/1973): "Every mental phenomenon is characterized by [...] what we could call, although in not unambiguous terms, the reference to a content, a direction upon an object (by which we are not to understand a reality in this case) or an immanent objectivity". This view is often characterized as the "aboutness" or "of-ness" of mental states.

The concept of intentionality derived from Meinong, Brentano, Husserl, and Frege is present in the contemporary philosophy of mind and language (Dennett, 1969; Searle, 1983). We call it *semantic intentionality*, since it is related to the meaning attributed to mental states, including the cases of linguistic meaning. For more than two decades it has played a central role in the epistemology of cognitive sciences, as a keyword for discussing the reaching and limitations of computational strategies in the characterization of genuine cognition and mentality. John Searle (1980) has forcefully called attention to the serious obstacle posed by the philosophical problem of intentionality for artificial intelligence. The celebrated "Chinese Room Argument" aims to show that the states of a computational system do not have meaning except when interpreted from the outside, i.e., by those who interact with it. The more recent version of the argument (Searle, 1992) emphasizes the role played by the brain. Semantic intentionality is considered to be a product of the causal powers of the brain; if no such causal powers were replicated, no intentional state would ever be produced.

In the context of the ongoing debate about consciousness and the brain, some authors question why conscious states refer to the world external to the brain, but not to the internal states of the brain assumed to produce them (Velmans, 1993). This is also mentioned as the problem of knowing why “when I observe an object in the world the object as I observe it is entirely a product of my sensory (and conceptual) mechanisms, [but] it appears, however, to exist entirely independent of me; even its color appears as an intrinsic property of the object rather than as a mode of my perception” (Ellis & Newton, 1998: 421). Following Velmans’ terminology, we call this phenomenon *projective intentionality*.

What kind of brain function could support mental intentionality? Would a naturalized (neurobiological) account of intentionality meet the demands of a philosophical theory of intentionality? In this paper, our objective is to pursue an account of intentionality in terms of brain and behavioral functions. This account is not a full scientific explanation, but an outline of a possible evolutionary explanation. We review results stemming from neurobiology in order to argue that intrinsically intentional systems in the brain are the evolutionary product of functions that support adaptive self-organized actions. The intentional mark of mental states is therefore conceived of as derived from biological functions that can be accounted for in non-

intentional terms. Our account of intentionality begins by analysing two modalities of self-organized action present in living beings, namely *internally generated* and *goal-directed behavior*, evolutionarily leading to the development of specific brain structures, as e.g. networks for visually guided control of movement in the parietal cortex.

2. Internally Generated Behavior

A simple internally generated behavior is ultimately a property of every living cell, but a nervous system is necessary for the control of a multi-cellular organism in the performance of complex behaviors. The notion of an internally directed complex behavior implies that:

a) more than one pattern of behavior is available to the animal;

b) the choice of the pattern to be performed, in a given context, is made by the animal;

c) the operation of the neuronal mechanism that makes the choice is not determined by external stimulation;

d) the animal does not necessarily construct a representation of goals to be achieved.

An example of simple self-determined behavior is the sexual activity of animals. Animals have a behavioral repertoire that contains diverse forms of social behavior. Sexual behaviors occur in situations where other social behaviors are equally possible. It is engendered by electrochemical mechanisms in the nervous system, involving hormones and synaptic facilitation.

In the performance of the activity, the animal does not have to represent a reproductive goal. The activity is perfectly conceivable in terms of a positive feedback loop, where each performed step is rewarded with pleasant sensations. The performance of the sequence of acts does not require the representation of a goal to be attained at the end. Not even a representation of pleasant sensations is necessary, since the reward is closely entangled with the activity itself. The process may be compared to an operant conditioning with a null time delay for the reward.

Simple internally generated behavior is close to the concept of teleonomy (Mayr, 1989). "Teleonomy" means that biological functions are determined by a "program" encoded in the DNA. However, recent developments in molecular biology have shown

that the DNA does not determine completely the function of cells; such functions are better characterized as *epigenetic*, i.e. the result of a dynamic interaction of genes, proteins and environment. Neurons, for instance, are highly sensitive to external information and do not develop all their potentialities in the absence of external stimulation. Therefore, in the context of our discussion we replace the concept of teleonomy by the concept a *self-organizing system* composed by the organism, its genes and the environment.

The neurobiological understanding of self-directed behavior is related to the conception of the brain as a dynamical system. In this view, the control of behavior is described in terms of neuronal activity approaching "attractors". In far-from-equilibrium systems, such as living beings, the increase of entropy is blocked by a flow of external negative entropy. The effect of the second law, in this context, is not driving the system to thermodynamic equilibrium, but to self-organizing processes leading to states where energy acquires a stable configuration. "Attractors" are the most stable classes of states; once the system reaches one of them there is a high probability that it will remain there for a reasonable amount of time.

Conceiving of brains as dynamical systems is a methodological simplification, since the physical components of such a system have not been formally defined. The problem lies in the complexity of brain structures, which encompass multiple levels of analysis. The solution found by Freeman (1995) was to assume a simplified but realistic neural network model of the brain. In the dynamical analogy, the number of attractors in the state space of the model corresponds to the number of different kinds of behavior. The structure of the state space can be designed to explain the transitions between such attractors, assuming a stable environment. As the animal switches from one kind of behavior to another, the trajectory of the dynamical system moves from one attractor to another.

An *orienting mechanism* (Sokolov, 1975) is responsible for the choice among stereotyped patterns of response, in the simplest modality of internally oriented behavior. In reptiles, it is possibly related to the function of basal ganglia, as proposed by McLean (1990) and Panksepp (1998). An example of an orienting system with a small number of attractors is found in the study of the behavior of small fishes. The behavioral state space of a fish or reptile, according to the McLean (1990) classification,

contains only four kinds of behavior: feeding, fleeing, fighting and reproduction. An interesting model of fish behavior, examining the reaction of fish to their own mirror image, goes back to Tinbergen's studies of the innate behavior of the three-spined stickleback (Tinbergen, 1951).

A current experimental setting for the study of fish behavior makes use of Oreochromis Niloticus males, which compete for territory. One male fish is set in a rectangular aquarium with a partition, where one of the walls is a mirror (Volpato, 1997). The subjects are e.g. young (2 months of age) Oreochromis Niloticus specimens. At the beginning of the experiment the partition is down, and the fish explores the half of the aquarium that does not contain the mirror. After five minutes the partition is lifted. As the fish comes closer to the mirror wall, the image of a male fish is produced. When it perceives the image, it displays two kinds of fighting behavior, frontal and lateral. The behavior is, of course, also displayed by the mirror image. The perception that the "other" (the image) is aggressive increases the aggressiveness of the fish. An excitatory feedback between actions and perceptions occurs.

When two agonistic fish struggle, the consequence is the increase of stress, eventually leading to the death of one of the opponents. In the case of one fish and a mirror, there is also stress, but the outcome is an inhibitory turning point: as stress increases, the fish decreases the signals of aggressiveness; this phenomenon is of course reflected in the image, and influences the fish's next reaction. The inhibitory feedback process leads to an extinction of the aggressive behavior, and the fish begins a new exploratory process. Swimming through the aquarium, it may find food, probably changing from exploratory to feeding behavior. The feeding behavior also has two phases, excitatory and inhibitory feedback. The inhibitory phase begins with satisfaction, leading to extinction. At this moment, the fish begins a new exploratory behavior that leads it to the vicinity of the mirror, where another cycle of aggressive behavior begins.

Lorenz and Tinbergen's model of innate behavior, with hierarchically fixed patterns released by specific types of stimuli, is the first candidate for an explanation of the fish's fighting behavior (Lorenz, 1952). In natural contexts, the release of fixed patterns is related to social companionship. In the experimental case above, stimuli are absolutely constant, so any variation in the fish's perception is a consequence of its own

previous action (e.g., if it changes the position of its body, it will see the same environment from a different perspective). The fish has a stereotyped reaction to the image, releasing the fixed pattern of fighting, and after some time (depending on individual variations) the central mechanism that released the pattern inhibits it. After the investigatory behavior, when the fish comes near the mirror for the second time, the fixed pattern is released again, followed by the inhibitory mechanism.

The internal feedback from action to perception presumed to exist in this kind of fish is the corollary discharge (Sperry, 1950). It has only the (hypothetical) role of stabilizing the visual field, and cannot, under normal circumstances, produce a new kind of behavior. In Sperry's experiment, circular swimming was produced, but one eye of the fish was covered, and the other rotated in 180 degrees. Moreover, the cognitive processing in the fish's brain, as in many other species with the same brain architecture, can only relate external signals with categories corresponding to fixed patterns (although it has been argued that it is possible for some fish species to have learning capacities not found in the observed species; see, e.g., Barlow, 1968, and Bitterman, 1968).

In our example of internally generated behavior, we identify a *proto-intentional function* – corresponding to the operations of the orienting mechanism – that only selects an option from a limited repertory, according to the perceived properties of a stimulus. In this case there is no need for "aboutness" of mental states, not even for mental states properly. However, it is necessary for brain activity to be directed towards the stimulus and its dynamical changes that elicit the continuous operation of the orienting mechanism, defining the kind of behavior to be adopted at each moment. This may be the first step in the evolution of intentionality.

3. Goal-Directed Behavior

A second step in the evolution of intentionality was made possible by means of the development of brain structures that provide the basis for a choice of behavior based on evoked emotional responses. In rodents, E. Sokolov (1975) and O. Vinogradova (1975) related the "orienting reflex" with the limbic system. In this expression, the word "reflex" does not mean a response elicited by external conditioning. On the contrary, it

refers to an unconditional instinct that guides behavior in the following kinds of situations:

a) when a stereotyped pattern is initiated, without being triggered by a specific stimulus, e.g., a chicken searching for food, in the absence of any specific food stimulus;

b) when a stereotyped pattern is triggered by a specific stimulus, e.g., a predator-like sound inducing the prey to escape;

c) when there is a stereotyped reaction to a change in a habitual environmental setting, e.g., the miller who wakes up when the noisy mill stops working.

One role of the limbic system in rodents would be signaling changes among types of behavior. Sokolov and associates tested this hypothesis experimentally, focusing on cases b) and c) above (see diagrams in Sokolov, l975, p. 226, 233). An information-processing version of his theory is currently used in psychophysiology (see Hughdal, 1995; for the role of the amygdala, see LeDoux, 1994).

The emergence of emotions provides a basis for an inner selection of behavior (LeDoux, 1994; Panksepp, 1998), but in the absence of the representation of possible goals to be attained and their impact on future emotions, the choice process is largely a trial-and-error one. Not surprisingly, the phylogenetic evolution of the limbic system occurred together with the development of frontal cortical areas able to elaborate on the representation of goals. The mammalian brain, evolving over but preserving the limbic orienting mechanism, shows a gradual development of the associative areas of the neocortex supporting primate (and human) typical functions. These areas compose the *executive system*, involving an attentional subsystem, as well as a large network that supports declarative and working memory (temporal, parietal and frontal associative neocortical regions). The executive system (see D'Esposito and Grossman, 1996) was presumably segregated from and superposed to the earlier orienting mechanism, allowing larger flexibility of responses to environmental stimuli. Corresponding to this evolution, a second step in the evolution of self-organized action is *goal-directed behavior*. This is a special case of self-determined behavior when the animal *represents a goal* to be achieved and directs its activity towards the goal. The category applies only to animals that have the capacity of representing goals and planning a sequence of actions leading to the goal.

A goal is a state of affairs that is not available to the animal at the moment when it is represented. Regarding the relative "non-existence" of the represented goal, this modality of representation is not far from Brentano's concept. For animals that represent goals, our simple dynamical analogy above seems to be insufficient to account for their behavior, since the behavioral repertory becomes as large as their representational and implementational capacities, generating a complex behavioral state space. It is also important to consider that the sequences of actions of the animals are referring to the represented goal. This implies that the animal can perform actions that generate unpleasant sensations, if such actions are conceived as a necessary step to reach the intended goal.

The capacity for directing actions to a goal reveals the existence of a second cognitive function, besides the orienting mechanism. It is the *attentional executive function* (see Posner, 1995), responsible for the guidance of action and subsequent perception, according to intended goals of the organism. In primates, the attentional system involves a large network that encompasses the prefrontal cortex, its connections with parietal and temporal associative areas, and the cyngulate gyrus. Therefore, understanding goal-directed behavior would require the consideration of *two dynamical brain sub-systems that complement and oppose each other* – the orienting and the attentional executive systems (see Grossberg and Merrill, 1996).

A dynamical account of goal-directed behavior requires the consideration of such inter-connected dynamical systems, both opposed and complementary to each other. The distinction between the present state of affairs and the intended one (the goal) can be metaphorically conceived of as a kind of "opponent processing" between the systems (for a review of opponent processing, see e.g. Seymour, O'Doherty, Koltzenburg, Wiech, Frackowiak, Friston and Dolan, 2005). In a very simplified view, the excitation of attentional systems (prefrontal cortex and cyngulate gyrus) focusing on the goal to be attained, generates an inhibition of the orienting system. Such inhibition is necessary because the orienting system is evolutionarily tuned for stereotypical responses to immediate stimuli. The pursuit of a goal requires inhibition of automatic responses in favor of the release of actions that are necessary for achieving the goal. It is possible, of course, that both systems become compromised with each other, when a sub-goal happens to be compatible with the demand of the orienting system. On the

other hand, strategies that systematically impose sacrifices upon the orienting system may disrupt the overall balance and fail.

The consideration of two coupled dynamical systems could therefore help explaining goal-directed behavior, but it is still insufficient to account for one aspect of the cognitive processing involved in the behavior: how is the animal able to represent something inexistent in the external world? This problem directly addresses the issue of "aboutness" of mental representations.

4. From Self-Organized Action to the "Aboutness" of Representations: the Example of Self-Initiated Locomotion

The "aboutness" of mental representations is a disposition of brain states that does not necessarily arise in direct connection to behavior; i.e., the “aboutness” of those states may exist even if a behavioral pattern is not elicited from them. Of course, such aboutness allows us to assign to those states *possible* or *prospective objects* in the world, and to support goal-directed behavior. Explicit memories, for instance, are endowed with aboutness, although not eliciting behavior or not taking part in an actual behavioral cycle.

In order to understand the possibility of the evolutionary transformation of self-organized action into mental intentionality, we avoid a common interpretation of the problem. Intentionality is frequently conflated with mental representation and, in so doing, some philosophers (following Brentano) claimed that no physical thing could be about another thing. Intentional states could not be brain states. However, there is no *a priori* reason to deny that “aboutness” could happen between physical beings. We are in the world, and so are our mental states. Why do the latter happen *in* the world and, nonetheless, are *about* something?

The riddle becomes less mysterious if we initially consider that an account of self-organized action provides us with an explanation for the advent of a distinction between the "inner" and the "outer". Once we are able show how self-organized action could generate such a distinction and how the sense of externality of our experience may arise from habituation and blockage of brain self-reference, we eschew most of the philosophical conundrums concerning the nature of intentionality.

How to explain the intentional character of memories, beliefs, desires and other cognitive states in terms of brain events, even when they are not connected to some kind of behavioral manifestation? An account of the nature of such brain states is likely to require more than what is proposed in the theory of attractors. It requires taking into account possible evolutionary processes that determined the functional architecture of the brain in such a way that its *internal* states are *directed towards the domain of adaptive interaction* with the environment, even when the present behavior is not being controlled by such states.

Therefore, in attempting to tackle the nature of “aboutness”, two issues need to be addressed: first, how brain states not connected to behavior can have “aboutness”, and secondly, how conscious processing, although internal to the brain, generates episodes which are experienced as occurring outside it.

With respect to the first problem, a central modality of self-organized action is *self-initiated locomotion*, an ancient functional motor skill, one that emerged in the evolutionary process as a life-strategy of diverse living beings in order to cope with the fulfillment of their basic biological needs. Self-initiated locomotion is an adaptive advantage, providing the possibility of seeking out new energy sources by changing to different environmental configurations. The important cognitive consequence of this skill is the possibility of the organism drawing a distinction between the "inner" and the "outer" in terms of "what moves" and "what may not move" relatively to a self-initiated motion.

Intentionality in the sense of “aboutness” includes *representing* objects and processes. The notion of *pragmatic representation* implies that perceptual objects are viewed under different modes of presentation, depending on the actions of the organism (Clark, 1996). Such modes of presentation reveal different properties of the object that would still be considered as ways of presenting the very same object. The different modes of presentation are related to different actions of the organism towards the object, i.e., the action of the organism towards an object influences or determines the means whereby the object is cognitively apprehended. The classical example of an intensional (with an s) context provided by the philosophical literature is the Fregean discussion of the "Morning Star" as contrasted to the "Evening Star". They are, of course, different modes of presentation of the planet Venus.

Current literature on this topic holds that one can have different beliefs about the "Morning Star" and the "Evening Star", although such representations pick out the same object in the world. One can entertain beliefs about the Evening Star that would not withstand if they were extended to the Morning Star. In other words, the possible truth of belief sentences about the Morning Star and the Evening Star differ in spite of the fact that they pick out the same object in the world, namely, Venus. The change in truth value of sentences involving the Morning Star and the Evening Star is affected by the presence of a pragmatic representation that mediates our cognitive relationship to an object in the world (Venus).

Some philosophers consider the generation of pragmatic representations a purely linguistic surface phenomenon. There are also philosophers who attempted to relate the intensionality (or multiple modes of presentation of an object) with the intentionality manifested in behavior. This second approach is worth exploring, since there may be a relationship between intensionality (with an s) and intentionality (with a t) involved in behavior. Some philosophers (Dretske, 1981) have maintained that if we explain how and why such intensional contexts are generated we will also have explained how representations are generated. Thus, we would have an explanation for “aboutness” as well, in so far as the latter is a pragmatic representation. The evolutionary principle at the genesis of such contexts is, we submit, the possibility of apprehending an object by moving around it, i.e., by self-initiated locomotion.

The generation of intensional contexts by moving around objects can be appreciated if we take into account a situation in which an organism is viewing an object "a" in a certain context and acquires the belief that "a" is F. When viewing it in another context the organism acquires the belief that "a" is G. However, it might be the case that such an organism does not acquire the belief that "a" is F and G. Certainly an explanation is needed here for the fact that it could believe that "a" is F without being G, even if in another situation it comes to learn the fact that it was indeed one and the same object "a" it came across in both situations. The reason why the organism could have different beliefs about the same object lies in the fact that it was not capable of realizing *a priori* that it was actually relating to the same object.

So viewed, the phenomenon can be understood as fundamentally rooted in differences between modes of presentation *generated by the organism's self-initiated*

locomotion. The latter is what generates multifarious modes of presentation of an object while moving around it. Intensionality can thereby be conceived of as something rooted in the more primitive notions of moving around an object and a "failure" in our processes of inspection of objects in the world. Such an inspection does not allow us to equate *a priori* different modes of presentation as being presentations of the same object in a different pragmatic context.

Self-initiated locomotion allows us to understand the origin of intentionality as “aboutness”, i.e. our brain states being about something in different pragmatic contexts or in the perceptual absence of the denoted object. Could it be sufficient to explain representations that do not denote anything, i.e., representations about an inexistent object? For instance, in the case of the intensional representations "the aunt that died twenty years ago" and "the death of the unicorn", the difference lies in the fact that the former, although picking out something perceptually absent, could be re-enacted in some kind of effective behavior, whereas the second cannot.

The main characteristic of the latter kind of representation is the brain making an inductive (and/or abductive) inference from what exists to what does not exist. The possibility of making such an inference may well be supported by the capacity of representing perceptually absent objects. A possible line of research on this issue is relating the way we come to have meaningful thoughts about inexistent objects and the child's apprehension of the hidden sides of objects. Although such experimental research with infants has been developed since Piaget's pioneering efforts, the relation with philosophical discussions of intentionality still remains to be established.

5. Some Experimental Evidences for Brain Systems Characterized by an Intentional Modus Operandi

In the study of visual perception, the organization of different visual appearances of the same stimulus has been related (Ullman, 1984) to the organism's movements. One possible implication of the influence of action on perception is that the very idea of an invariant object would emerge from distinguishing variations in perception due to bodily movement from variations resulting from the object's movement. So viewed, bodily movement in self-initiated motion is what accounts not only for the concocting

of visual episodes as forming an object, but also their invariant properties identified from the organism's variation of perspectives.

The same role is played by self-initiated motion in auditory perception. Whenever we hear a sound, we hear not only the pitch and the tone but also localize the source of the sound. Such localization depends on binaural mechanisms and also on the organism's possible shifts in spatial location relative to a sound source. Self-initiated locomotion allows to differentiate whether it was the sound source that moved or the organism that changed place in the environment.

Empirical evidence in favor of the view that self-initiated motion plays a central role in the organism's apprehension of physical objects in the world was advanced by a classical series of experiments carried by Richard Held and Alan Hein at MIT (Held and Hein, 1958; Hein and Held, 1962; Held and Hein, 1963). This work points to the importance of self-motion in several cognitive tasks, including the development of motor skills, organization of perception and the individuation of physical objects. For example, Held devised an experiment with humans in order to show that "the importance of body movement and particularly of self-produced movement derives from the fact that only an organism that can take account of the output signals to its own musculature is in a position to detect and factor out the decorrelating effects of both moving objects and externally imposed body movement" (Held and Hein, 1963, p. 378).

Using a prism, Held and Hein (1963) conducted an experiment where the hand of a subject was moved constantly, and such a movement was perceived as independent of actual motion that could be taking place. The subject could not control the changes in his visual inputs, because of the effect of the prism. The result of the experiment pointed to the deterioration in coordination between eye and hand under conditions of active movement. Furthermore, it showed the importance of the correlation between movement and sensory feedback in maintaining accurate coordination. The series of experiments demonstrate a role of motor-sensory feedback in visual adaptation, as well as the role of active movements in facilitating this process. Such experiments with visual adaptation were also expanded to situations involving movements of the entire body and not just the arm or the hand. They reinforce the conclusion of existence of a link between motor and visual mechanisms in the central nervous system (see also

recent studies pointing to the same conclusion, as Gentilucci, Chieffi, Daprati, Saetti and Toni, 1996; Tse, Cavanaugh and Nakayama, 1997).

Since we consider self-initiated motion a biological function selected by its advantages in the fulfillment of basic needs, intentionality would thus arise as a result of primitive non-intentional forms of agency. The circumstances in which such a condition occurs begin with the notion of an organism's attempts to satisfy those biological needs by moving around. As an evolutionary consequence of this habit, the nervous system comes to incorporate a lead to the world that allows the organism to behave successfully. Such interaction between organism and environment could start as a trial-and-error task or as a kind of cycle where the effects of those movements are registered by the nervous system and form a kind of feedback loop that selects actions until behavioral adequacy is attained. The registering of those effects plays an essential role in constraining new cycles of trial-and-error, thus narrowing the set of behaviors to a set of adequate behaviors. Once such adequacy is attained, the trial-and-error cycle may evolve into a disposition to behave in a certain way in the environment.

Initially, dispositions may constitute just a motor skill that defines a trajectory between the informational input and the reaction of the organism – a cycle where informational input triggers a certain pattern of behavior. This is already a gain for the satisfaction of the biological needs of the organism, and the nervous system could pre-empt the trial-and-error cycle by incorporating such a disposition in the form of a reflex arc.

The possible gain in behavioral adequacy is, nevertheless, constrained by a lack in terms of resilience. If the environment changes, the triggered behavior may not be beneficial for the organism. The next step in the evolution of behavior is therefore the induction of brain mechanisms controlling behavior to be *resilient*, i.e., prone to possible environmental changes. Based on the emergence of such resilience, representational “aboutness” can be understood in terms of a *residual directness*, the by-product of an interrupted or incomplete behavior cycle, as classically proposed by Miller, Gallanter and Pribam (1960). These authors suggest that directness is a pervasive property of brain states, ranging from reflex functions to goal directed behavior and, finally, to “aboutness”. Our hypothesis – following the classical

hypothesis of Miller et al. (1960) – is that representational “aboutness” would evolutionarily depend on more basic dimensions of intentionality.

In the present state of evolution, we can have dispositional states in the absence of interrupted behavior, or even in the absence of any ongoing behavior related to such states. According to our hypothesis, the present state of affairs is a result of an evolutionary process that led to the emergence and maintenance of the resilience of brain states in order to support adaptive actions.

This fact implies that an appropriate explanation of “aboutness” should take into account the evolutionary emergence of new brain structures and functions able to represent possible actions independently of ongoing behavior. Cognitive Neuroscience has progressively studied diverse brain systems that fulfill this requirement: neuronal networks localized in the anterior cingulate cortex, premotor cortex, posterior parietal, inferior temporal and prefrontal cortex, with a possible involvement of the cerebellum and basal ganglia.

Advances in the studies of functional anatomy and the proposal of more realistic theoretical models have led to a distinction between brain areas that represent the intention of performing an action and the areas that control the performance of the action (see Mazzoni, Bracewell, Barash and Andersen, 1996; Colby, Duhamel and Goldberg, 1996; Snyder, Batista and Andersen, 1997; Cisek, Grossberg and Bullock, 1998). These findings corroborate observations made by Held, Hein, Ullman and others (see Jeannerod, 1997) about the relation of perception and action. Such a concept of intentional representation was incidentally refuted by the most influential researcher – J.J. Gibson (1979) – who stressed the importance of action in perception.

The consideration of motor influence upon perception, assumed by relatively few researchers in the past, became one of the main areas of research in contemporary cognitive neuroscience. Commenting on a recent study by Ballard, Hayhoe, Pook and Rao (1997) on "deictic codes", Goodale remarked: "It joins nicely the fields of motor control and cognition [...] It also makes evolutionary sense by suggesting that mechanisms which evolved for the distal control of movement might have been co-opted (in both a literal and figural sense) for the computations underlying the cognitive life of the animal" (Goodale, 1997).

Another field of research has recently been developed, focusing on prefrontal areas responsible for an anticipation of possible future events (Hasegawa, Blitz, Geller and Goldberg, 2000). According to Miller and Cohen, "cognitive control stems from the active maintenance of patterns of activity in the prefrontal cortex that represent goals and the means to achieve them. They provide bias signals to other brain structures whose net effect is to guide the flow of activity along neural pathways that establish the proper mappings between inputs, internal states, and outputs needed to perform a given task" (Miller and Cohen, 2001, following the research made by Rainer, Rao and Miller, 1999). The representation of goals is assumed to involve a cognitive operation called *prospective memory*, defined by Burgess, Quayle and Frith (2001) as "the functions that enable a person to carry out an intended act after a delay".

6. The Problem of Projective Intentionality

The problem of projective intentionality (Velmans, 1993) refers to explaining how processes going on in the brain are experienced as occurring in the external world. For obvious adaptive reasons, brain states should not produce sensations of their own workings. Projective intentionality has an important function of allowing the organism to direct action upon objects in the world, and not towards the neuronal machinery that supports cognition.

The generation of an *externally-referenced content* when a stimulus is received, by the exclusion of the inner brain's workings, gives rise to the "first-person" or "phenomenal world" (Chalmers, 1996). The phenomenal world is generated by the activity of the nervous system, but it is experienced as standing outside. Signals impinging on the peripheral sensors are processed by the central nervous system, and nonetheless their generators are perceived as standing outside.

It is generally accepted in neuroscience – following classical studies made by Muller in the XIXth century – that the localization of a percept is determined by the *afferent pathway*. For example, a touch sensation is localized in the region of the skin where contact occurred, and not in the brain receptors of the signal. This assumption helps to describe the phenomenon but does not attempt to formulate an explanation.

From a physicalist perspective, there is an apparent paradox in the fact that sensations are not localized in the place where they are presumably produced: how can a signal that is physically located in the brain *be perceived as* being outside the brain? A biological fact that could help dissolve the paradox is that the brain does not have internal sensors. In other words, all the input to the central nervous system, in normal situations, comes from the peripheral sensors (eye, ear, touch sensors at the skin, etc.). However, the problem that remains to be explained is why percepts are localized at the location where the peripheral sensors are situated (or "in front of" them), and not *where the signal is sent to*.

A possible solution comes from noting that the influx of external signals to the brain is necessary for the production of percepts. Endogenously produced action potentials and rhythmic patterns do not produce any sensations by themselves (except in dreaming and hallucination, two phenomena we will not discuss here). In the case of direct brain stimulation in patients under surgery, or with chronically implanted electrodes (Penfield and Boldrey, 1937), the external stimulation provoked by the neurologist is necessary to produce "phantom" sensations (i.e., sensations of an inexistent external object). Our hypothesis is that recurrent relations maintained by the active organism with the environment indirectly support the projective relation. This hypothesis may be developed in terms of a process of *habituation* that engendered *a peculiar structure of neuronal networks*, such that only inputs that match patterns from the peripheral sensors are able to reach a certain threshold of activity corresponding to conscious perception.

An exceptional situation, in evolutionary terms, is the ingestion of substances – hallucinogenic drugs – or chemical disbalancements – as in schizophrenia – that produce sensations endogenously to the brain. In these cases, projective intentionality is present, since all endogenously produced sensations are subjectively or phenomenically referred to the external world, but the recurrent interaction with the external world is perturbed at the moment when the sensations are generated. However, it is important to consider that the subjects who experience such endogenously-generated sensations interacted with the external world during their entire life, allowing the brain intentional mechanisms to develop.

An analysis of how habituation works in the ontogenetic scale (see Sokolov, 1975; Gray, 1995) suggests an adaptation of brain perceptual mechanisms for the recognition of distal stimuli based on information carried by proximal signals. For instance, when a person holds a tool, (e.g. a hammer), he/she first experiences contact with the tool, but as he/she uses it, he/she becomes habituated and then the attentional focus turns to the interface between the hammer and the part of the environment that offers resistance to it, e.g. the nail.

The signals generated internally to the brain are the most proximal stimulus, the ones to which we are most habituated. Is it possible that in the distant evolutionary past, or in some artificial situation, the brain could perceive its own workings? The answer seems to be positive, since it is possible for the brain to monitor its own activities, by means of a technique called "biofeedback" (see e.g. Sterman and MacDonald, 1978; Hauri, Percy, Hellekson, Hartman and Russ, 1982; Ayers, 1991). Another example is closing the eyes and pressing the eyeballs with a finger; some phantom imagens can usually be seen, presumably produced by remaining excitatory activity of retinal cells. In both examples, the nervous system perceives a fragment of its own workings, but in normal adaptive situations habituation mechanisms preclude brain self-perception. These examples serve to demonstrate that it is not impossible for the brain to perceive parts of itself, hence suggesting that projective intentionality would be generated by habituation mechanisms that *block* such self-perception.

Discussing computational mechanisms in the cortex, Phillips and Singer (1997) wrote: "the local processors are in effect discovering distal variables and relationships [...] these foundations do not constitute intentional representation proper because such local processors do not distinguish between the signals they receive and the distal causes from which those signals arise" (1997, p. 663). This remark provides one clue for the discovery of the brain mechanisms underlying projective intentionality. It could be supported by inhibitory mechanisms necessary to block the perception of signals internal to the brain. For instance, it is well known that the existence of inhibitory interneurons restrict feedback loops in cortical columns. Cortical layer 4 seems to be protected by inhibitory neurons, so as to receive input only from the thalamus; more precisely, it may be inhibited soon (around 12 ms) after receiving an input, as a way of avoiding excitation by the same signal re-entering from layer 5 to the superficial layers.

Such inhibition could favor the tangential spreading of the excitatory potential to other columns, generating sequential cognitive processing.

An interesting consequence of habituation processes is that the perceptual space of animals is progressively shaped by the space of their actions. If projective intentionality is supported by inhibitory mechanisms related to learning processes, then in a familiar environment the space where animal localize the stimuli is learned to be the same space where they act. For instance, a chained dog, although having the same brain architecture as a stray one, *perceives* the world differently – an observation made by Kurt Goldstein and discussed by philosopher Merleau-Ponty (1945). In other words, the *structure of the action space*, defined by the relationship between body and environment, *shapes the structure of the perceptual space*, both in phylogenetic and ontogenetic scales. As a consequence, when we find ourselves in a strange environment a feeling of disorientation emerges, since the perceived space does not match the action space to which our neuronal networks were previously habituated.

Based on such intuitive observations, as well as on the scientific data reviewed above, we suggest the possibility of a neuroscientific approach to intentionality.

João de F. Teixeira and Alfredo Pereira Jr
Federal University of São Carlos, Brasil
São Paulo State University (UNESP), Brasil

jteixe@terra.com.br / apj@ibb.unesp.br

References

Ayers, M. (1991) 'A Controlled Study of EEG Neurofeedback Training and Clinical Psychotherapy for Right Hemispheric Closed Head Injury', *Los Angeles: National Head Injury Foundation Annual Conference*.

Ballard, D.H., Hayhoe, M.M., Pook, P.K. & Rao, R.P.N. (1997) 'Deitic Codes for the Embodiment of Cognition', *Behavioral and Brain Sciences* 20, 723-767.

Barlow, G.W. (1968) 'Ethological Units of Behavior', IN D. Ingle (ed.), *The Central Nervous System and Fish Behavior*, Chicago: University of Chicago Press.

Bitterman, M.E. (1968) 'Comparative Studies of Learning in the Fish', IN D. Ingle (ed.), *The Central Nervous System and Fish Behavior*, Chicago: University of Chicago Press.

Brentano, F. (1925/1973) *Psychologie von empirischen Standpunkt / Psychology from an empirical standpoint*, translated by A. Pancurello, D. Terrell, L.L. McAlister, New York: Humanities Press.

Burgess, P.W., Quayle, A. & Frith, C.D. (2001) 'Brain Regions Involved in Prospective Memory as Determined by Positron Emission Tomography', *Neuropsychologia* 39 (6), 545 - 555.

Chalmers, D.J. (1996) *The Conscious Mind*, New York: Oxford University Press.

Cisek, P., Grossberg, S. & Bullock, D.S. (1998) 'A Cortico-Spinal Model of Reaching and Proprioception Under Multiple Task Constraints', *Journal of Cognitive Neuroscience* 10, 425-444.

Clark, A. (1996) *Being There: Putting Brain, Body and World Together Again*, Cambridge, MA: The MIT Press.

Colby, C.L., Duhamel, J.R. & Goldberg, M.E. (1996) 'Visual, Presaccadic and Cognitive Activation of Single Neurons in Monkey Lateral Intraparietal Area', *Journal of Neurophysiology* 76, 2841-2852.

Dennett, D.C. (1969) *Content and Consciousness*, London: Routledge and Kegan Paul.

D'Esposito, M. & Grossman, M. (1996) 'The Physiological Basis of Executive Functions and Working Memory', *The Neuroscientist* 2, 345-352.

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: The MIT Press.

Ellis, R. and Newton, N. (1998) 'Three Paradoxes of Phenomenal Consciousness:Bridging the Explanatory Gap', Journal of Consciousness Studies, 5 (4), 419-42.

Freeman, W.J. (1995) *Societies of Brains*, Hillsdale: Lawrence Erlbaum.

Gentilucci, M., Chieffi, S., Daprati, E., Saetti, M. & Toni, I. (1996) 'Visual Illusion and Action', *Neuropsychologia* 5 (34), 369-376

Gibson, J.J. (1979) *The Echological Approach to Visual Perception*, Boston: Houghton-Mifflin.

Goodale, M. (1997) 'Pointing the Way to a Unified Theory of Action and Perception', *Behavioral and Brain Sciences* 20 (4), 749-750.

Gray, J.A. (1995) 'The Contents of Consciousness: A Neuropsychological Conjecture', *Behavioral and Brain Sciences* 18, 659-722.

Grossberg, S. & Merrill, J.W.L. (1996) 'The Hippocampus and Cerebellum in Adaptively Timed Learning, Recognition, and Movement', *Journal of Cognitive Neuroscience* 8, 257-277.

Hauri, P.J., Percy, L., Hellekson, C., Hartman, E. & Russ, D. (1982) 'The Treatment of Psychophysiologic Insomnia with Biofeedback: a Replication Study', *Journal of Biofeedback and Self Regulation* 7 (2), 223-236.

Hasegawa, R.P., Blitz, A.M., Geller, N.L. & Goldberg, M.E. (2000) 'Neurons in Monkey Prefrontal Cortex That Track Past or Predict Future Performance', *Science* 290, 1786-1789.

Hein, A. & Held, R. (1962) 'A Neural Model for Labile Sensorimotor Coordinations', IN *Biological Prototypes and Synthetic Systems, Vol. 1,* New York: Plenum Press.

Held, R. & Hein, A. (1958) 'Adaptation of Disarranged Hand-Eye Coordination Contingent Upon Re-Afferent Stimulation', *Perceptual and Motor Skills* 8, 87-90.

Held, R. & Hein, A. (1963) 'Movement-Produced Stimulation in the Development of Visually Guided Behavior', *Journal of Comparative and Physiological Psychology* 5, 872-876.

Hughdal, K. (1995) *Psychophysiology*, Cambridge, MA: Harvard University Press.

Jeannerod, M. (1997) *The Cognitive Neuroscience of Action*, Oxford: Blackwell

LeDoux, J.E. (1994) 'The Amygdala: Contributions to Fear and Stress', *Seminars in the Neurosciences* 6, 213-237.

Lorenz, K. (1952) 'The Past Twelve Years in the Comparative Study of Behavior', IN C.H. Schiller (ed.), *Instinctive Behavior*, New York: International Universities Press.

Mayr, E. (1989) *Toward a New Philosophy of Biology*, Cambridge, MA: Harvard University Press.

Mazzoni, P., Bracewell, R.M., Barash, S. & Andersen, R.A. (1996) 'Motor Intention Activity in the Macaque's Lateral Intraparietal Area I. Dissociation of Motor Plan from Sensory memory', *Journal of Neurophysiology* 76, 1439-1457.

McLean, P.D. (1990) *The Triune Brain in Evolution*, New York: Plenum Press.

Merleau-Ponty, M. (1945) *Phenomenologie de la Perception*, Paris: Gallimard.

Miller, E.K. & Cohen, J.E. (2001) 'An Integrative Theory of Prefrontal Cortex Function', Annual Review of Neuroscience 24, 167-202.

Miller, G.A., Gallanter, E.H. & Pribram, K. (1960) *Plans and the Structure of Behavior,* New York: Rinehart and Winston.

Panksepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions*, New York: Oxford University Press.

Penfield, W. & Boldrey, E. (1937) 'Somatic Motor Sensory Representation in the Cerebral Cortex of Man as Studied by Electrical Stimulation', *Brain* 60, 389-443.

Phillips, W.A. & Singer, W. (1997) 'In Search of Common Foundations for Cortical Computation', *Behavioral and Brain Sciences* 20, 657:722.

Posner, M.I. (1995) 'Attention in Cognitive Neuroscience: An Overview', IN M.S. Gazzaniga (ed.), *The Cognitive Neurosciences*, Cambridge, MA: The MIT Press.

Rainer, G., Rao, S.C. & Miller, E.K. (1999) 'Prospective Coding for Objects in Primate Prefrontal Cortex', *Journal of Neuroscience* 19, 5493-505.

Searle, J. (1980) 'Minds, Brains and Programs', *Behavioral and Brain Sciences* 3, 417-424.

Searle, J. (1983) *Intentionality: An Essay on the Philosophy of Mind*. Cambridge: Cambridge University Press.

Searle, J. (1992) *The Rediscovery of Mind*, Cambridge, MA: The MIT Press.

Seymour, B., O'Doherty, J.P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K. & Dolan, R. (2005) 'Opponent Appetitive-Aversive Neural Processes Underlie Predictive Learning of Pain Relief', *Nature Neuroscience* 8 (9), 1234-40.

Snyder, L.H., Batista, A.P. & Andersen, R.A. (1997) 'Coding of Intention in the Posterior Parietal Cortex', *Nature* 386, 167-170.

Sokolov, E.N. (1975) 'The Neuronal Mechanisms of the Orienting Reflex', IN E. Sokolov & O. Vinogradova (eds.), *Neuronal Mechanisms of the Orienting Reflex*, New York: Lawrence Erlbaum.

Sperry, R.W. (1950) 'Neural Basis of the Spontaneous Optokinetic Response', *Journal of Comparative Physiology* 43, 482-489.

Sterman, M.B. & MacDonald, L. (1978) 'Effects of Central Cortical EEG Feedback Training on Incidence of Poorly Controlled Seizures', *Epilepsia* 19, 207-222.

Tinbergen, N. (1951) *The Study of Instinct*, Oxford: Clarendon Press.

Tse, P., Cavanaugh, P. & Nakayama, K. (1997) 'The importance of parsing in high-level motion processing', IN T. Watanabe (ed.), *High-Level Motor Processing*, Cambridge, MA: The MIT Press.

Ullman, S. (1984) 'Visual Routines', *Cognition* 18, 97-159.

Velmans, M. (1993) 'A Reflexive Science of Consciousness', IN G. Bock & J. Marsh (eds.), *Experimental and Theoretical Studies of Consciousness*, Chichester: John Wiley.

Vinogradova, O.S. (1975) 'The Hippocampus and the Orienting Reflex', IN E. Sokolov & O. Vinogradova (eds.), *Neuronal Mechanisms of the Orienting Reflex*, New York: Lawrence Erlbaum.

Volpato, G. (1997) 'Fighting Behavior of Oreochromis Niloticus (Video)', Botucatu: State University of São Paulo (UNESP).

Abstracta 4:2 pp.148-184, 2008

¿ES INCOHERENTE LA POSTULACIÓN DE MUNDOS POSIBLES?

José Tomás Alvarado Marambio

Abstract:

This work presents the different forms of Cantorian paradoxes that have been proposed against all the variety of forms of Actualism in the metaphysics of modality. These Cantorian paradoxes are presented in the wider context of Cantorian paradoxes directed against the notion of "world". The article presents two main general strategies to deal with the paradoxes: (i) the strategy of "amplification" and (ii) the strategy of "restriction". The first strategy is proposed as the most plausible. As the problems affect in the same way all actualist theories (and even the notion of "world"), they cannot be used against any particular actualist theory of modality. The feasibility of a general amplification strategy, on the other hand, is a reason to suppose that all those theories have not much to fear from those paradoxes.

Resumen:

Este trabajo presenta las diferentes formas de paradojas cantorianas que se han propuesto contra toda la variedad de formas de actualismo en metafísica modal. Estas paradojas cantorianas se presentan en el contexto más amplio de las paradojas dirigidas contra la noción de "mundo". El artículo presenta dos estrategias principales para neutralizar las paradojas: (i) la estrategia de "ampliación", y (ii) la estrategia de "restricción". Se defiende la primera como la más plausible. Como los problemas afectan del mismo modo a todas las teorías actualistas (e incluso a la noción de "mundo"), no pueden ser utilizados en contra de alguna teoría actualista de la modalidad en particular. La plausibilidad de una estrategia general de ampliación, por otro lado, es un motivo para suponer que ninguna de tales teorías tiene mucho que temer de tales paradojas.

Muchos filósofos han sostenido en las últimas tres décadas que la mejor forma de comprender los hechos que hacen verdaderos o falsos a los enunciados modales es como una totalidad de "mundos posibles". Cuando se dice que "p es necesario" se debe entender que p es verdadera en todos los mundos posibles y cuando se dice que "p es posible", entonces se debe entender que existe al menos un mundo posible en el que p es verdadera. ¿Qué motivos hay para pensar que esta forma de comprender las condiciones de verdad de los enunciados modales es aceptable? David Lewis es especialmente claro en este respecto:

> Creo que hay mundos posibles diferentes del que de hecho habitamos. Si se quiere un argumento, es éste. Es verdadero de manera no controvertida que las cosas podrían ser de

> otra forma de cómo son. Creo, y lo mismo usted, que las cosas podrían haber sido diferentes de incontables maneras. ¿Pero qué es lo que esto significa? El lenguaje ordinario permite la paráfrasis: hay muchas formas en que las cosas podrían haber sido además de la forma en que son actualmente. Esto es evidentemente una cuantificación existencial. Dice que existen muchas entidades que satisfacen cierta descripción, esto es, 'formas en que podrían ser las cosas'. Creo que las cosas podrían ser diferentes de incontables maneras. Creo en las paráfrasis permisibles de lo que creo. Tomando la paráfrasis tal como aparece, creo, por lo tanto, en la existencia de entidades que podrían ser llamadas 'formas en que podrían ser las cosas'. Prefiero llamarlas 'mundos posibles'.[1]

El argumento desplegado por Lewis podría ser aceptado por todo filósofo que postule mundos posibles. Por supuesto, Lewis sostiene además que esos mundos posibles diferentes del mundo actual son entidades de la misma naturaleza que el mundo actual, esto es, sumas mereológicas de todas las entidades relacionadas entre sí espacio-temporalmente, pero esto es algo que no es necesario aceptar para lo que se tratará aquí.[2] La cuestión crucial es que parece haber un motivo muy simple para aceptar que, dado que creemos que las cosas podrían ser diferentes, *hay* formas alternativas en que podrían ser las cosas. Si hay formas alternativas en que podría ser una entidad específica, como el gato Micifuz, no parece existir ninguna dificultad en pensar que hay también formas alternativas en que podrían ser *todas* las cosas: mundos posibles.

Comprender qué son los mundos posibles es comprender, por lo tanto, qué es lo que hace verdaderas o falsas a las proposiciones modales. Nos parece obvio que hay muchas proposiciones verdaderas sobre lo que podría suceder, pero si existen tales verdades (y son conocidas por nosotros), entonces debe existir algo en virtud de lo que sean verdaderas. Parece intuitivamente obvio que aquello que hace verdaderas a las proposiciones modales son las formas alternativas en que podrían darse las cosas, todas las cosas, esto es, lo que hace verdaderas o falsas a las proposiciones modales son la totalidad de mundos posibles. Esto es terreno neutral, sin embargo, entre una multitud de teorías que explican la naturaleza de esas entidades. Una forma de realismo modal extremo es la defendida por David Lewis, tal como se ha indicado. Esta concepción se denomina usualmente "posibilismo modal" y se opone a aquellas teorías que sostienen que el mundo actual no se

---

[1] D. Lewis, (1973, 84).

[2] D. Lewis ha complementado esta argumentación con una detallado y complejo desarrollo para justificar que los mundos posibles deben ser entendidos como él los postula y no como los postula un filósofo actualista (cf. D. Lewis, 1986). Es perfectamente posible aceptar el argumento de Lewis citado en el texto y desechar la forma idiosincrática en que Lewis explica qué son los mundos posibles.

encuentra a la par, desde el punto de vista ontológico, que toda la restante pléyade de mundos posibles. Estas teorías se denominan usualmente "actualistas", por la preferencia ontológica dada al mundo actual. Este trabajo tiene que ver con una familia de dificultades que afectan a las teorías actualistas de la modalidad. En efecto, si se postula que, en algún sentido de la palabra, el único mundo "real" es el mundo actual, ¿cómo pueden ser comprendidos los mundos posibles? El filósofo actualista debe, de alguna manera, especificar que "existen" los mundos posibles pero sin negar la preferencia ontológica por el mundo actual. La forma más socorrida de realizar esta especificación es sosteniendo que los mundos posibles son construcciones abstractas efectuadas a partir de elementos que vienen ya dados en el mundo actual.

Por ejemplo, muchos filósofos han pensado en los mundos posibles como "historias completas" sobre cómo podría estar constituida la realidad. No se debe suponer que hay un mundo aparte del mundo actual del que sean verdaderas tales historias, tal como lo puede ser una historia infinitamente detallada y exhaustiva respecto del mundo actual. Esas historias, sin embargo, *son* los mundos posibles. Se puede pensar en tales historias como conjuntos de oraciones de un lenguaje determinado o como simplemente un conjunto de proposiciones. La forma en que estas historias llegan a describir completamente cómo podría estar constituido el mundo es porque son máximamente consistentes. Se dice que un conjunto de proposiciones S es máximamente consistente si y sólo si, para toda proposición bien formulada p, o bien $p \in S$, o bien $\neg p \in S$. La identificación de los mundos posibles con conjuntos máximamente consistentes de oraciones o proposiciones ha sido propuesta por filósofos como Carnap,[3] Jeffrey[4] y Adams.[5] El problema que será presentado afecta de una manera directa a estas concepciones de los mundos posibles asociadas a "historias completas", pero las restantes formas de actualismo también están sujetas a paradojas de este estilo, por lo que el examen que se hará aquí tiene un valor de carácter general.

El problema surge por un resultado básico de teoría de conjuntos. Cantor probó que la cardinalidad del conjunto potencia de un conjunto dado es mayor que la cardinalidad de

---

3 Cf. R. Carnap (1947)
4 Cf. R. C. Jeffrey, (1965)
5 Cf. R. M. Adams, (1979)

ese conjunto.[6] La cardinalidad es la función que asigna a un conjunto su número de elementos. Para un conjunto A, sea su cardinalidad #A. Por otra parte, el conjunto potencia de un conjunto A, sea P(A) es el conjunto compuesto por todos los sub-conjuntos de A. En términos generales, la cardinalidad del conjunto potencia de un conjunto A, tal que #A = n, es de 2^n (esto es $[\#A = n] \rightarrow [\#P(A) = 2^n]$). Cantor estaba interesado en la generalización de este resultado para números transfinitos. Si hay una cardinalidad que pueda ser asignada al conjunto de los números naturales, esto es, $\#N = \aleph_0$, entonces se podrá definir inmediatamente un conjunto potencia del conjunto de todos los números naturales, con una cardinalidad equivalente a $\#P(N) = 2^{\aleph_0} = \aleph_1$. Nada impide ahora que se defina el conjunto potencia del conjunto potencia del conjunto de todos los números naturales, esto es PP(N), y su cardinalidad puede ser definida como $\#PP(N) = 2^{\aleph_1} = \aleph_2$. La iteración del mismo procedimiento permite generar un "paraíso" matemático de entidades todas ellas de numerosidad infinita pero no equivalente, pues, en general $\forall n\ n \leq 2^n$. Todo esto es bien conocido. Lo que tiene relevancia para la cuestión que se discute aquí es que #P(A) > #A.

Pues bien, considérese que en las concepciones actualistas que se están comentando se sostiene que un mundo posible es un conjunto máximamente consistente de proposiciones formuladas en un lenguaje. La cuestión crucial es que por cada sub-conjunto del conjunto de todas estas proposiciones existirá una proposición que o bien formará parte del conjunto, o bien su negación formará parte del conjunto. Entonces parece resultar que el conjunto máximamente consistente en cuestión tiene un número de proposiciones tan grande como su conjunto potencia, lo que va derechamente en contra del teorema de Cantor. En términos más formales, el razonamiento puede formularse así:[7]

(1) $\forall S\ [(S \text{ es máximamente consistente}) \leftrightarrow (\forall q\ (q \in S \vee \neg q \in S) \wedge (S \text{ es consistente}))]$

Esto es la definición de qué significa ser un conjunto máximamente consistente. Se subentiende que se trata de un conjunto de proposiciones.

---

[6] Para explicaciones generales del desarrollo del punto de vista conjuntista en filosofía de las matemáticas, cf. R. Torretti, (1998), especialmente 21-47; M. Potter, (2004) especialmente 153-174, véase el teorema 9.2.6.

[7] Esta explicación sigue de cerca de J. Divers (2002, 243-256).

(2) Hay un conjunto máximamente consistente de proposiciones A

(3) $\#P(A) > \#A$

(2) es la hipótesis que se va a someter a la *reductio ad absurdum* y (3) es el teorema de Cantor, aplicado al caso del conjunto máximamente consistente de proposiciones A. Ahora bien:

(4) $\forall S ((S \in P(A) \rightarrow \exists q$ (q versa sobre S))

Esto es, por cada conjunto perteneciente al conjunto potencia P(A) existirá una proposición sobre ese conjunto. En efecto, si uno de esos conjuntos es $\{q_1, q_2, q_3\}$, entonces, por ejemplo, existirá una proposición que enunciará que $q_1 \in \{q_1, q_2, q_3\}$ y otras muchas. Sea una proposición semejante r. Ahora bien, por (1) se sigue que:

(5) $(r \in A) \vee (\neg r \in A)$

Como esto vale para todo elemento de P(A), resulta que habrá tantas proposiciones en A como conjuntos pertenecientes al conjunto potencia de A, por lo tanto (por (5)) se sigue que:

(6) $\#P(A) \leq \#A$

Pero (6) está en abierta contradicción con (3). Hay, por lo tanto, una línea de razonamiento que conduce desde las premisas (1) a (4) a una contradicción. Debe rechazarse una de estas premisas. No es plausible rechazar la definición de conjunto máximamente consistente de proposiciones (1), tampoco es razonable rechazar el teorema de Cantor (3). Las opciones abiertas son la suposición (2) de que hay un conjunto máximamente de proposiciones A, o bien de que hay al menos una proposición por cada sub-conjunto del conjunto potencia de A, esto es (4). Como (4) es plausible por motivos independientes, parece que se debe

rechazar (2). Esto es, pareciera que la idea de que hay conjuntos máximamente consistentes de proposiciones está en contra de la teoría de conjuntos y debe ser rechazada.

Este trabajo tratará de explorar formas de resolver esta aporía para diversas concepciones actualistas. Para esto, se van a considerar, en primer lugar, las diferentes formas de aporía y cómo es que afectan a las diversas formas de actualismo. En segundo lugar, se van a considerar estrategias generales de solución de las aporías. Se va a argumentar aquí que: (i) las paradojas cantorianas no afectan de manera específica a una u otra teoría modal actualista, sino que se trata de dificultades que afectan por igual a cualquier concepción que utilice –de algún modo– la noción de "mundo"; (ii) dado lo anterior, las paradojas cantorianas requieren un tratamiento general. Para este tratamiento general se va a sostener que lo más razonable es sostener que realmente el "mundo" no es un conjunto; tampoco deben concebirse como conjuntos las otras "construcciones" actualistas que se suelen aducir para entregar mundos posibles. Esto es, la teoría de conjuntos debe ser vista como una teoría matemática interesante, pero no como nuestra ontología fundamental. Junto con considerar la estrategia indicada, se explicará porqué las estrategias de "restricción" no parecen aceptables.

1. Diversas formas de la aporía cantoriana

No existe, en realidad un único problema proveniente de las consideraciones de cardinalidad cantorianas. Se trata de una entera familia de problemas que afectan no sólo a las concepciones actualistas de mundos posibles sino también a postulaciones metafísicas de carácter completamente general. Grim, uno de los filósofos que ha explotado con mayor profusión esta familia de dificultades señala que argumentaciones análogas a la desplegada arriba afectan a la idea de que existe un conjunto de todas las verdades, a la idea de que existe un conjunto de todas las verdades necesarias, de que existe un conjunto de todas las falsedades y de que existe un conjunto de todas las cosas conocidas por un ser omnisciente.[8] Bringsjord ha planteado la misma dificultad respecto de conjuntos de todos los estados de cosas o hechos.[9] Chihara, por su parte, ha desarrollado argumentos contra la idea de que existe un conjunto de todos los estados de cosas posibles y contra la idea de que

[8] Grim ha presentado estas dificultades en una serie de escritos: P. Grim, (1984, 206-208; 1986, 186-191; 1997, 146-151)

[9] Cf. S. Bringsjord (1985, 64; 1989, 186-189).

existe un conjunto de todas las esencias.[10] Será conveniente considerar estas formas diferentes de argumentación.

(A) No existe un conjunto de todas las verdades. Supóngase –para efectuar una *reductio ad absurdum*– que hubiese un conjunto de todas las verdades, sea $A = \{p_1, p_2, \ldots, p_n\}$. Existe una cardinalidad de este conjunto de todas las verdades $\#A = k$. Resulta, sin embargo, que a cada sub-conjunto perteneciente a A puede asignársele una proposición verdadera. Por ejemplo, si $B \subseteq A$, es el conjunto $\{p_{i-1}, p_i, p_{i+1}\}$, entonces existe al menos una verdad como $p_1 \notin B$. Esta verdad ha de formar parte del conjunto A, pues es –en efecto– el conjunto de todas las verdades. Pero entonces, el conjunto de todas las verdades tiene tantos elementos como el conjunto potencia de A, P(A). Así, resulta $\#A \geq \#P(A)$, en contradicción con el resultado de Cantor.

Esta misma forma de razonamiento, que no tiene que ver de manera especial con mundos posibles, permite mostrar que no hay un conjunto de todas las verdades necesarias, no hay un conjunto de todas las falsedades ni un conjunto de todas las verdades conocidas por un ser omnisciente. Para el caso de verdades necesarias, toda la diferencia viene de que el conjunto A estará restringido. Cada sub-conjunto tendrá trivialmente asignada otra proposición verdadera y necesaria, pues la pertenencia de cada proposición verdadera a un sub-conjunto dado es un hecho necesario. Para el caso del conjunto de todas las falsedades, A estará compuesto por todas las negaciones del conjunto que ha entrado en el argumento central sobre las verdades. A cada sub-conjunto se le puede asignar una proposición falsa (por ejemplo, si un sub-conjunto de A es $\{p_1, p_2\}$, entonces se puede formular una falsedad enunciando que $p_1 \notin \{p_1, p_2\}$) y el argumento vuelve a funcionar de la misma manera. Por último, en el caso de todas las proposiciones verdaderas conocidas por un ser omnisciente, el argumento funciona porque el ser omnisciente, precisamente por serlo, deberá también conocer de cada sub-conjunto de A que es un sub-conjunto de A.

(B) No existe un conjunto de todos los estados de cosas o hechos. Se trata éste de un argumento muy semejante a (A), pues los estados de cosas o hechos se conciben ordinariamente como aquello que hace verdadera (o falsa, según sea el caso) a una

[10] Cf. Ch. Chihara, (1998, 126-127; 130-131).

proposición. Si hay un argumento para las verdades, entonces, parece obvio que habrá un argumento para aquellas entidades que están correlacionadas con las proposiciones verdaderas. Pues bien, sea A el conjunto de todos los estados de cosas del mundo. A cada sub-conjunto de A se le podrá asignar un nuevo estado de cosas. Por ejemplo, si hay un sub-conjunto $B = \{S_1, S_2\}$, entonces hay también un estado de cosas S_3 constituido por el hecho de que S_1 pertenece al conjunto B (o, simplemente, al hecho de ser B el conjunto compuesto por S_1 y por S_2). Nótese cómo este argumento parece implicar el rechazo de la idea de un "mundo" constituido por todos los estados de cosas.

(C) No existe un conjunto de proposiciones o de oraciones máximamente consistente. El razonamiento es tal como se detalló arriba. Éste es el argumento que parece atacar directamente a las concepciones actualistas que funcionan con "historias completas".

(D) No existe un conjunto de todos los estados de cosas posibles. Se trata de una variación sobre (B) que está referido al conjunto de todos los estados de cosas actuales. Sea el conjunto de todos los estados de cosas posibles A. A cada sub-conjunto de A podrá asignársele un estado de cosas posible. Por ejemplo, si $B \subseteq A$ está compuesto por $\{S_1, S_2\}$, entonces hay un estado de cosas posible consistente en que $S_1 \in B$. Este argumento parece ir directamente contra la forma de concebir los mundos posibles postulada por Plantinga como estados de cosas posibles máximos, tal como se indicará a continuación.

(E) No existe un conjunto de todas las esencias. Éste es también un argumento que ha sido dirigido contra la concepción modal de Plantinga en la que las entidades posibles no-actuales están representadas por una esencia (que no se encuentra instanciada). La esencia es aquí un conjunto de propiedades que son satisfacidas por un y sólo por un individuo en todos los mundos posibles. Debe suponerse que hay esencias individuales no sólo para objetos sino también para estados de cosas. Esto es, a cada estado de cosas, sea actual o meramente posible, se le asigna una propiedad que es satisfacida por ese estado de cosas y por nada más en todos los mundos posibles. Por ejemplo, si se quiere especificar la esencia individual del estado de cosas consistente en que el gato Micifuz es gordo, se puede perfectamente definir una propiedad consistente en ser algo al mismo tiempo Micifuz y

gordo. Esto es, si M es la esencia individual del gato Micifuz y G es la propiedad de ser gordo, entonces existe la propiedad de $[\lambda x\ (Mx \wedge Gx)]$. Es trivial, entonces, que pueden ser definidas todas estas esencias individuales para estados de cosas si es que hay esencias individuales para los objetos que constituyen tales estados de cosas. Sea ahora un conjunto A de todas las esencias individuales. A cada sub-conjunto de A puede asignarse una esencia individual. En efecto, sea un conjunto $B \subseteq A$, compuesto por $\{E_1, E_2\}$, en que E_1 y E_2 son esencias individuales. Supóngase que $E_1 = [\lambda x\ (Mx \wedge Gx)]$ y que $E_2 = [\lambda x\ (Nx \wedge Fx)]$ (por ejemplo, E_1 es la esencia individual del estado de cosas de ser Micifuz gordo y E_2 es la esencia individual de ser Tom feroz), entonces es trivial que existe una esencia individual $E_3 = [\lambda x \lambda y\ ((Mx \wedge Gx) \wedge (Ny \wedge Fy))]$ (esto es, la esencia individual del estado de cosas de ser Micifuz gordo y Tom feroz). Resulta, por lo tanto, que hay tantas esencias individuales como sub-conjuntos del conjunto de todas las esencias individuales. Esto genera la paradoja cantoriana inmediatamente.

(F) No existe el conjunto de todas las entidades. Este argumento no ha sido formulado con anterioridad, pero es una aplicación bastante obvia. Supóngase simplemente que a nuestra ontología se añade un principio de mereología del siguiente tenor: si hay dos objetos diferentes x e y, entonces existe la suma mereológica (x + y). Se trata de un principio plausible intuitivamente. Ahora bien, sea A el conjunto de todas las entidades del mundo, tal que $A = \{x_1, x_2, \ldots, x_n\}$. A cada sub-conjunto de P(A) se le puede asignar una entidad específica constituida por la suma mereológica de todas las entidades que son miembros de ese sub-conjunto. De acuerdo al principio mereológico indicado, esas entidades son también entidades del mundo y han de formar parte del conjunto A. Resulta, entonces, que A parece tener tantos elementos como su conjunto potencia en contra de lo establecido por el teorema de Cantor.

(G) No existe el conjunto de todos los universales. Este argumento tampoco ha sido formulado con anterioridad. Supóngase que existiese un conjunto de todos los universales $A = \{U_1, U_2, \ldots, U_n\}$. Existe un conjunto potencia de A, P(A) cuya cardinalidad debe ser superior a la cardinalidad de A. Pero por cada sub-conjunto de A puede definirse un universal complejo. Por ejemplo, dado $B \subseteq A$, tal que $B = \{U_1, U_2\}$, entonces hay un

universal que consiste en la instanciación conjunta de U_1 y de U_2, esto es $[\lambda x\ (U_1x \wedge U_2x)]$. Sucede entonces que hay tantos universales como sub-conjuntos del conjunto A, pero todos esos universales pertenecen a A, en contra del resultado de Cantor.

Tal como se puede apreciar, entonces, las paradojas cantorianas parecen afectar a las concepciones modales actualistas de diversos modos y también a ciertas ideas fuertemente arraigadas en la tradición filosófica.[11] Pareciera, en efecto, que fuese inconsistente pensar en el *mundo* como el conjunto de todos los estados de cosas o hechos, pensar también en un conjunto de todas las proposiciones que son hechas verdaderas en virtud de esta totalidad de estados de cosas y pensar que estas totalidades de hechos y de proposiciones verdaderas pudiesen ser conocidas por una entidad que conoce "todo". Si se quiere, el problema no afecta a las concepciones modales actualistas por el hecho de tratarse de teorías específicamente modales, sino que el problema parece provenir de la representación de los hechos modales mediante *mundos* posibles. Pruritos semejantes al que parecen plantear estas paradojas cantorianas han hecho que tradicionalmente los lógicos rechacen la cuantificación irrestricta sobre *todas* las entidades existentes. Es un resultado característico de teoría de conjuntos que no existe un conjunto universal compuesto por todos los conjuntos y, del mismo modo, pareciera que estos mismos principios que obligaron al desarrollo de la concepción iterativa de conjunto, conducen a una restricción fundamental en nuestras ontologías, pues parece que debemos aceptar que no existe el *todo*, no hay *universo*.[12] Así, si no resulta coherente hablar de todos los estados de cosas actuales ni de todas las proposiciones actualmente verdaderas, no es nada extraño que tampoco parezca coherente hablar de todos los estados de cosas posibles o de todas las proposiciones pertenecientes a un conjunto máximamente consistente de proposiciones. De esta manera, las paradojas cantorianas parecen exigir un tratamiento unitario y de carácter general, que

[11] Se ha omitido en este conjunto de paradojas cantorianas una forma atribuida a David Kaplan que afecta la suposición de que existe una totalidad de mundos posibles. Esta paradoja está formulada en M. Davies (1981, 262). También se pueden consultar formulaciones en M. Jubien, (1980); D. Kaplan, (1973). El argumento es satisfactoriamente contestado por David Lewis (1986, 104-108).

[12] Por supuesto, esto no impide que utilicemos ordinariamente las cuantificaciones universales en nuestros discursos ordinarios, pero suponiendo siempre –tal como lo han advertido casi todos los lógicos– que esas cuantificaciones universales se encuentran tácitamente restringidas a un dominio de entidades determinado. El problema que se plantea aquí no es que no pueda usarse la expresión "todo" en esos sentidos ordinarios restringidos tácitamente a un cierto dominio sobre el que se habla. El problema surge para un uso irrestricto de "todo".

es lo que se va a proponer aquí. Ha sido frecuente en la literatura dedicada de manera específica a cuestiones de metafísica modal que se proponen soluciones especiales para tal o cual tipo de paradoja cantoriana. Este enfoque es, al parecer, equivocado: no hay teorías actualistas inmunes a los problemas de cardinalidad; no hay tampoco concepciones ontológicas sobre el "mundo" que se encuentren libres de estas dificultades. La solución –o soluciones– deberán ser aplicables a todas estas concepciones por igual. En efecto, todas las concepciones modales que se están considerando explican los hechos modales por apelación a una totalidad de mundos posibles que son "formas alternativas en que podría estar constituido el mundo", aunque difieran entre sí en cómo se concibe la naturaleza de estas "formas alternativas". Siempre, en todo caso, los mundos posibles son especificaciones de cómo podría ser el *mundo*. Luego, si hay mundos posibles hay un mundo actual (en el caso de Lewis, hay sencillamente "mundos" en plural). Si los argumentos (A), (B) y (F) ponen en cuestión la existencia de un mundo como conjunto de todas las verdades, conjunto de todos los estados de cosas o conjunto de todas las entidades existentes, entonces estos argumentos cantorianos serán un motivo para negar que exista una pluralidad de mundos posibles. Tal como se ha indicado más arriba, no porque exista alguna dificultad particular en la forma en que vienen concebidos tales mundos posibles, sino sencillamente porque deben utilizar una categoría conceptual incoherente. Tácitamente las concepciones actualistas (y las posibilistas también) están suponiendo que hay una totalidad, el mundo, que podría ser diferente. Los argumentos cantorianos mostrarían que no hay tal entidad (el mundo) y no puede existir, por lo que tampoco hay formas alternativas en que podría darse esa entidad. El problema, por lo tanto, tiene un carácter absolutamente general, y no está restringido a la metafísica modal.

2. Las aporías cantorianas y las diferentes formas de actualismo

¿Cómo afectan estas distintas paradojas cantorianas a las teorías actualistas? Las teorías actualistas se suelen clasificar en cuatro grandes grupos: (i) la teoría lingüística (o proposicional), (ii) la teoría de Plantinga, (iii) la teoría combinatoria, y (iv) la teoría de los

mundos posibles como propiedades.[13] Se verá aquí que no hay ninguna teoría actualista que muestre ventajas por lo que respecta a los problemas de cardinalidad.[14]

(i) Teorías lingüísticas. Tal como se ha indicado más arriba, es característico de estas concepciones modales que los mundos posibles sean concebidos como "historias completas" de cómo podría estar constituida la realidad. No se trata de aquello que haría verdaderas a tales historias sino de las historias mismas. Si hay una especificación completa de cómo podría estar dado el mundo, descendiendo hasta el más mínimo detalle para cada instante de tiempo, entonces todo lo que se quisiera que estuviese contenido en el mundo, si es que las cosas fuesen diferentes de cómo lo son actualmente, estará reflejado en alguna peculiaridad de la "historia completa". Para construir la historia completa se puede utilizar un lenguaje determinado o bien se puede apelar a un conjunto de proposiciones. En principio, el recurso a proposiciones puede resultar más apropiado pues es dable esperar que las proposiciones no tengan las limitaciones expresivas que pueden afectar a nuestros lenguajes naturales y artificiales.[15] Si se hace apelación a un lenguaje natural como el español para la construcción de historias "completas", por ejemplo, es obvio que resultarán inadecuaciones importantes, pues en español no tenemos nombres para cada entidad existente (y mucho menos para cada entidad posible no existente actualmente).

Tal como se ha indicado más arriba, la forma en que típicamente se consigue especificar una "historia completa" es tomando cada oración o proposición bien formada del lenguaje en cuestión y añadiendo o bien la oración (o proposición) o bien su negación a un conjunto consistente. Este procedimiento generará una totalidad de conjuntos máximamente consistentes de oraciones o proposiciones que serán las entidades que cumplirán aquí el papel de mundos posibles. Un estado de cosas es actual si y sólo si la proposición que expresa el darse de tal estado de cosas es verdadero. El mundo actual viene dado por el conjunto máximamente consistente de oraciones y proposiciones en el que

---

13 Para esta clasificación, cf. J. Divers (2002, 169-180).

14 Si el lector está ya persuadido de que todas las diferentes teorías actualistas están afectadas por igual por paradojas cantorianas puede dispensarse de leer esta sección y pasar directamente a la siguiente donde se tratan las estrategias generales de solución.

15 Puede extrañar el que se incluya entre las teorías lingüísticas las construcciones de conjuntos máximamente consistentes de proposiciones. Esta agrupación está motivada por el hecho de que las "historias completas" formadas por oraciones y las "historias completas" formadas por proposiciones poseen una estructura semejante.

todos sus elementos son verdaderos. Un estado de cosas es posible si y sólo si la proposición que expresa el darse de ese estado de cosas pertenece al menos a un conjunto máximamente consistente. Un estado de cosas es necesario si y sólo si la proposición que expresa el darse de tal estado de cosas pertenece a todos los conjuntos máximamente consistentes. Esta forma de concebir los mundos posibles es simple y elegante y, tal vez por ello, ha resultado atractiva para muchos filósofos que siguen trabajando en la depuración de una u otra forma de teoría lingüística.[16]

Las teorías de este estilo, sin embargo, se encuentran directamente afectadas por las paradojas cantorianas (que pueden ser adaptadas para conjuntos máximamente consistentes de oraciones o para conjuntos máximamente consistentes de proposiciones) de la forma (C). Parece obvio que por cada sub-conjunto del conjunto máximamente consistente que se encuentre describiendo cómo está constituido el mundo, existirá una oración o proposición que esté enunciando algo verdadero o que podría ser verdadero, si ese mundo fuese actual. Las teorías lingüísticas se encuentran, por lo tanto, directamente afectadas por las paradojas cantorianas.

(ii) Teorías combinatorias.[17] En las teorías combinatorias los mundos posibles vienen concebidos como construcciones dadas a partir de un conjunto de elementos que se encuentran ya dados en el mundo actual. Estos elementos son básicamente objetos y propiedades. En el mundo actual los objetos y propiedades constituyen estados de cosas o hechos que hacen verdaderas a las proposiciones que enuncian el darse de tales estados de cosas. La idea general es que esos mismos objetos y propiedades podrían estar combinados de otros modos, esto es, que los mismos objetos que ya existen en el mundo actual podrían instanciar otras propiedades y relaciones, configurando, por lo tanto, otros estados de cosas diferentes de los existentes en el mundo actual.[18] Dado un conjunto de todos los objetos $\{x_1, x_2, \ldots, x_n\}$ y un conjunto de todas las propiedades $\{P_1, P_2, \ldots, P_n\}$ puede definirse un conjunto de todos los estados de cosas posibles como el conjunto de todos los pares ordenados (o n-tuplas ordenadas, según sea el caso) de objetos y propiedades $\{<P_1, x_1>,$

16 Cf. T. Roy (1995); J. Melia (2001), 19-29; Th. Sider (2002).

17 Defensores de concepciones modales combinatorias son D. M. Armstrong (1989), véase en particular 37-53; M. J. Cresswell (1979).

18 Naturalmente, filósofos que se encuentran inclinados a reemplazar objetos y propiedades por tropos especificarán los estados de cosas de manera diferente, pero la idea central será la misma.

$<P_1, x_2>$, …, $<P_1, x_n>$, $<P_2, x_1>$, $<P_2, x_2>$, …, $<P_2, x_n>$, …, $<P_n, x_1>$, $<P_n, x_2>$, …, $<P_n, x_n>$}. Pues bien, dado un conjunto de todos los estados de cosas (independientes entre sí) pueden definirse la totalidad de mundos posibles como la totalidad de todas las combinaciones posibles de estados de cosas. Utilizando un ejemplo elemental se puede apreciar cómo se especifican estos mundos posibles: sean los estados de cosas S_1 y S_2 (si se quiere, los que resultan del poseer un único objeto dos propiedades diferentes), entonces hay cuatro mundos posibles, a saber $\{S_1, S_2\}$, $\{S_1, \text{no-}S_2\}$, $\{\text{no-}S_1, S_2\}$, $\{\text{no-}S_1, \text{no-}S_2\}$. En la concepción combinatoria los conjuntos de objetos y propiedades vienen tomados de los objetos y propiedades actualmente existentes.

Estas concepciones combinatorias resultan directamente afectadas por los argumentos cantorianos de la forma (D), pues deben hacer apelación al conjunto de todos los estados de cosas posibles, en el que deben incluirse estados de cosas correspondientes a cada sub-conjunto de tal conjunto.

(iii) La teoría de Alvin Plantinga. Plantinga define a los mundos posibles como estados de cosas posibles máximos.[19] No da demasiadas indicaciones sobre qué debe entenderse por un "estado de cosas", pero tiene que suponerse que su concepción no puede diferir demasiado de la usual en la que los estados de cosas: (a) vienen dados por la complexión de objetos y propiedades, y (b) son las entidades que hacen verdaderas (o falsas) a las oraciones o proposiciones. Un estado de cosas "posible" ha de ser un estado de cosas que, aun cuando no sea efectivo, podría serlo. Se dice, por otro lado, que un estado de cosas S es "máximo" si y sólo si para todo estado de cosas S*, o bien S incluye a S*, o bien S excluye a S*. S incluye a S*, si y sólo si, necesariamente si S es efectivo entonces S* es efectivo. S excluye a S* si y sólo si no es posible que se den conjuntamente S y S*. Un estado de cosas máximo, tal como lo entiende Plantinga, es un estado de cosas que ha de incluir toda determinación o hecho que podría constituir una forma alternativa en que podría estar constituido el mundo y, por ello, Plantinga lo identifica con un mundo posible. Plantinga añade a la noción de estado de cosas posible y máximo la noción de "libro" de un mundo. Dado un mundo posible existirá un conjunto de proposiciones verdaderas sobre lo que

[19] Cf. A. Plantinga (1974, 44-51; 2003, 103-121).

ocurre en tal mundo posible. Este conjunto de proposiciones verdaderas sobre tal mundo es lo que se denomina aquí el "libro" de ese mundo.

La concepción de Plantinga está afectada por las paradojas cantorianas de varias maneras. En primer lugar, Plantinga parece requerir un conjunto de todos los estados de cosas posibles (es aquello sobre lo que parece estar cuantificando su definición de un "estado de cosas máximo", así como las definiciones de "inclusión" y "exclusión"), lo que hace a su concepción sensible al argumento (D). En segundo lugar, aún cuando Plantinga no concibe a los mundos posibles como conjuntos de proposiciones, sí sostiene que esta asociado a cada mundo posible un conjunto de proposiciones que debe ser máximamente consistente. Los "libros" son susceptibles de caer en paradojas cantorianas de la forma (C) del mismo modo que la concepción modal lingüística. En tercer lugar, la concepción de Plantinga se exime de objetos posibles mediante la utilización de un conjunto de esencias individuales. La postulación de un conjunto de esencias individuales, sin embargo, parece estar afectada por paradojas cantorianas de la forma (E).

(iv) La teoría modal basada en propiedades o universales.[20] En esta concepción modal los mundos posibles son concebidos como propiedades o universales estructurales que especifican cómo estaría constituido el mundo si es que las cosas fuesen diferentes. Un universal estructural es un universal –esto es, una cierta determinación que es apta por su naturaleza para ser predicada de muchos– que surge por la complexión de otros universales más básicos. Por ejemplo, el universal "molécula de agua" se instancia por algo en el que hay tres partes, una de la cuáles instancia el universal "átomo de oxígeno", dos de las cuales instancian el universal "átomo de hidrógeno" y estas tres partes se encuentran relacionadas entre sí instanciando otro universal relacional. La idea general es que un mundo posible –una forma posible en que podría estar constituido el mundo– vendría dado por un universal estructural altamente complejo en el que se encontraría dado con todos sus detalles cada determinación que podría tener ese mundo. En esta concepción, la diferencia entre el mundo actual y los restantes mundos posibles es que el mundo actual es el único que se encuentra instanciado. Los restantes mundos posibles son universales que no están

[20] Cf. R. Stalnaker (2003); P. Forrest (1986) éste es también el tipo de teoría modal que he defendido, cf. J. T. Alvarado (2006).

instanciados (pero que podrían estarlo). Un estado de cosas S es posible si y sólo si, existe al menos un universal estructural máximo U, tal que, si U fuese instanciado, entonces S se daría. Por otra parte, un estado de cosas S es necesario si y sólo si, para todo universal estructural máximo U, si U es instanciado, entonces S se daría. Resulta crucial para la identificación de un mundo posible con un universal estructural que el universal en cuestión sea, en algún sentido de la palabra, "máximo". ¿Cómo se entiende aquí la "maximalidad"? Ya se ha visto cómo nociones semejantes deben aparecer en las teorías modales lingüísticas y en la teoría de Plantinga. En las primeras, porque se deben utilizar conjuntos *máximamente* consistentes de proposiciones (esto es, tales que para todo p, o bien p pertenece al conjunto, o bien ¬p pertenece al conjunto) y en la segunda, porque se debe utilizar la noción de un estado de cosas posible *máximo* (esto es, tal que para todo estado de cosas S, o bien se incluye a S o bien se excluye a S). La noción de maximalidad que aparece en estas teorías genera problemas, como se ha visto. En el caso de la teoría modal basada en universales la idea es que un mundo posible es un universal estructural "máximo" porque describe exhaustivamente como sería el mundo en la posibilidad contemplada. El universal en cuestión es capaz de codificar cómo es que sería el mundo porque: (a) es una especificación de un individuo "máximo", esto es, de un individuo tal que todo individuo es parte de él. Formalmente, se puede definir así un individuo máximo:

(7) $\forall x\,[(x \text{ es máximo}) \leftrightarrow \forall y\,(y \text{ es parte de } x)]$

Recuérdese que en mereología todo objeto es parte de sí mismo, aún cuando no sea parte propia de sí mismo. Trivialmente, entonces el "mundo", que es lo que intuitivamente debe entenderse como el individuo máximo,[21] es parte de sí mismo. La especificación del individuo máximo es (b) exhaustiva, en el sentido de que, dado un conjunto de *todos* los universales $\{P_1, P_2, \ldots, P_n\}$ se atribuirá a cada parte, ya sea el universal P_i, o se le atribuirá que no instancia ese universal P_i (para un universal $P_i \in \{P_1, P_2, \ldots, P_n\}$, naturalmente). La forma, entonces, en que un universal estructural llega a ser máximo es porque especificará

[21] No hay más que un individuo máximo, pues dos individuos son idénticos si y sólo si poseen exactamente las mismas partes. Cualquier cosa que satisfaga en un mundo posible la definición (7) va a coincidir con cualquier otra que también lo haga.

cada determinación que posea cada parte de un mundo posible, mediante cláusulas de este tenor:

(8) $\lambda x \exists y\, [(y \text{ es parte de } x) \wedge (P_1 y \wedge P_2 y \wedge \ldots \wedge P_n y)]$

En otras palabras, un mundo posible viene dado por una descripción exhaustiva de todas sus partes. A su vez, las partes de que se compone un mundo posible pueden ser codificadas completamente con la especificación completa de cada una de sus partes, y luego con la especificación completa de cada una de las partes de esas partes, etc. Es perfectamente aceptable, entonces que en la serie de universales P_1, P_2, …, P_n que aparece en (8) se encuentren otras cláusulas que tengan la misma forma de (8). Como puede apreciarse, podría suceder que la codificación de un mundo posible (como el mundo actual) se realice por un universal infinitamente complejo, por ejemplo, si es que las partes sólo se pueden codificar con la especificación de sus partes y éstas, a su vez, mediante la especificación de sus respectivas partes, y luego éstas otras deben remitir a sus partes, y a las partes de las partes, y a las partes de las partes de las partes, etcétera. Puede ser que el universal estructural máximo sea tan complejo que no resulte expresable por ninguna fórmula de un lenguaje natural o artificial nuestro, incluso siendo infinitario. Esto no tiene ninguna importancia. Un universal es un tipo de entidad que puede ser o no conocida por nosotros y puede o no ser cognoscible por nosotros, sin que esto limite su existencia de alguna manera.

Pues bien, ¿cómo pueden afectar las paradojas cantorianas a la teoría modal basada en universales? En primer lugar, se ve afectada esta teoría por argumentos de la forma de (E), esto es, argumentos que rechazan la existencia de un conjunto de todas las esencias. En la teoría modal basada en universales, en efecto, no pueden aparecer "objetos posibles". Naturalmente que hay objetos existentes en el mundo actual, pero nada más. Los mundos posibles son simples universales no instanciados (con una única excepción) y si un universal es tal que si fuese instanciado resultaría que Micifuz sería gordo, entonces Micifuz debe aparecer ahí representado por su esencia individual y no *propia persona*, por decirlo de algún modo. Del mismo modo, entonces, que en la teoría de Plantinga, en esta teoría modal se tienen que utilizar esencias individuales como el dominio sobre el que se cuantifica en los enunciados modales. Tal como se ha indicado más arriba, la idea general

es que un estado de cosas como el ser Micifuz gordo puede incluirse como una cláusula integrante de un universal estructural máximo, pues hay un universal (estructural, naturalmente) que consistirá en ser el gato Micifuz, sea M, y habrá otro universal estructural consistente en ser gordo, sea G, por lo que puede definirse el universal de ser algo Micifuz y gordo, esto es: $[\lambda x\ (Mx \wedge Gx)]$. La utilización de un conjunto de esencias individuales genera una paradoja cantoriana del tipo (E).

Un segundo problema de tipo cantoriano mucho más obvio para la concepción modal basada en universales es la del tipo (G), pues debe hacerse apelación –o eso parece– al conjunto de todos los universales y a construcciones de universales complejos a partir de otros universales más simples. Parece, por lo tanto, que puede generarse un universal complejo o estructural por cada subconjunto del conjunto de todos los universales, según se ha explicado arriba.

3. Estrategias generales de solución

¿Qué es lo que genera las paradojas cantorianas? En primer lugar se selecciona una cierta totalidad (todas las verdades, todos los estados de cosas, todas las esencias, todas las proposiciones de un conjunto máximamente consistente, todos los estados de cosas posibles, etc.). En segundo lugar, sucede que la totalidad seleccionada es "productiva" en el sentido de que dado un conjunto cualquiera de miembros, siempre pueden añadírsele nuevos miembros. La paradoja cantoriana surge porque se supone que un conjunto determinado posee cierta "fijeza" en sus miembros (sea finito o infinito, y si es infinito, sea numerable o indenumerable). Dado un conjunto "fijo", se determina ahora su conjunto potencia que ha de tener una cardinalidad o numerosidad mayor. El problema aquí es que la totalidad seleccionada, al ser "productiva", no tiene la "fijeza" que se requiere de un conjunto para que entre en la estructura iterativa. El teorema de Cantor permite discriminar entre conjuntos infinitos porque, aun tratándose de "totalidades" infinitas se les puede asignar cierta determinación en sus elementos (lo que en teoría de conjuntos es la determinación del conjunto en cuestión: las condiciones de identidad de un conjunto son estrictamente extensionales, pues dos conjuntos con los mismos elementos *son* el mismo conjunto). Una vez dado un conjunto, esto es, una cierta totalidad de elementos determinada, puede definirse un conjunto de todos los sub-conjuntos del primero. Esto

supone que la totalidad en cuestión está "ya dada". En las totalidades que generan las paradojas cantorianas, en cambio, sucede que la totalidad en cuestión nunca está "dada" completamente. La totalidad en cuestión "genera" nuevos miembros de sí misma. Esto es lo que puede ser denominado su carácter "productivo". Se trata, entonces, de dos supuestos:

(I) Se postula un *conjunto* dado A.

(II) Hay un principio generador tal que a cada sub-conjunto de A se le puede asignar un elemento de A.

La conjunción de (I) y (II) lleva a una contradicción patente con el resultado de Cantor según el cual la cardinalidad del conjunto potencia de A es mayor que la cardinalidad de A, cualquiera sea la cardinalidad de A. Las formas generales de resolver la paradoja son básicamente dos y se encuentran relacionadas directamente con los dos supuestos de tipo (I) y (II) que se encuentran en la estructura de toda aporía cantoriana. Estas estrategias son las siguientes: (a) estrategias de "ampliación" que consisten en la negación de (I), afirmando que existe una "totalidad" determinada A, pero *no* se trata de un conjunto. Se dice aquí que se trata de una estrategia de "ampliación" pues, de alguna manera, busca ampliar las formas en que puede ser dada cierta totalidad de elementos, no restringiéndolas a lo que viene dado en la teoría de conjuntos; y (b) estrategias de "restricción" que consisten en la negación de (II), tratando de neutralizar el principio generador que produce la aparición de elementos "nuevos" de A por cada sub-conjunto de A.[22] Se examinarán ahora por separado cada una de estas estrategias generales y se argumentará que mientras la estrategia (b) no es recomendable como solución general, sí lo es la estrategia (a).

[22] Otros autores han presentado aquí tres estrategias generales de ataque a las paradoja cantorianas. Por ejemplo, J. Divers (2002, 249-256) distingue entre soluciones de "restricción", de "clases propias" (que no son conjuntos) y de "no maximalidad". Aquí se toman lo que Divers denomina soluciones de "restricción" y de "no-maximalidad" como estrategias generales de restricción en el sentido (b) que se ha indicado. Se trata, en efecto, de estrategias que buscan impedir la operación del principio generador.

3.1. *Estrategias de ampliación*

Tal como presentan la cuestión los propugnadores de las paradojas cantorianas, la teoría de conjuntos, un cuerpo de teoría matemática utilizado como "fundamento" en todas las áreas de la matemática, pareciera decirnos que no existe el "mundo", tal como usualmente lo entendemos. En efecto, hay paradojas cantorianas que, tal como se ha visto, muestran, o mostrarían, que no existe el conjunto de todas las verdades, no existe el conjunto de todos los estados de cosas, no existe el conjunto de todos los objetos (y el conjunto de todas las propiedades) y luego, no es extraño que se declaren también inexistentes el conjunto de todas las proposiciones de un conjunto máximamente consistente (que describe cómo podría ser el mundo), el conjunto de todos los estados de cosas posibles y el conjunto de todas las esencias individuales. ¿Qué motivo tenemos para aceptar lo que parecen mostrar estas paradojas cantorianas? Como sucede ordinariamente con las paradojas auténticas, estas líneas de argumentación dejan la sensación de que lejos de mostrar que algo está mal con la idea de "mundo" (y con las variadas formas de comprender a los mundos posibles), muestran que algo anda mal con las premisas que han conducido a este resultado increíble. Pues es ciertamente increíble que no exista el mundo por tal o cual teorema ingenioso de Cantor, tal como resulta increíble pensar que no hay movimiento por las paradojas de Zenón. El desafío, más bien, es detectar qué es lo que está produciendo que nociones perfectamente aceptables para nuestra comprensión ordinaria aparezcan como incoherentes. Siempre aquí quedará abierta la posibilidad de aplicar un *modus tollens* en vez de un *modus ponens* (o al revés), si es que teoría de conjuntos y nuestra noción ordinaria de mundo se muestran incompatibles de manera irreparable. Esto es exactamente lo que se hará aquí.

Las estrategias de ampliación que aquí se comentan han estado inspiradas por el hecho de que en ciertas teorías axiomatizadas de conjuntos expresamente se deja espacio para ciertas totalidades que no son conjuntos y que han sido denominadas "clases". Como es bien conocido, existen dos grandes formas alternativas de axiomatizar la teoría de conjuntos,[23] una de estas formas es obra de Zermelo y Fraenkel (conocida usualmente como ZF) y la otra es obra de von Neumann, Bernays y Gödel (conocida usualmente como NBG). En NBG existe el concepto de una "clase" que puede no ser un conjunto. Una clase que no es conjunto se denomina ordinariamente una "clase propia". Von Neumann, en

[23] Cf. R. Torretti (1998, 71-111); M. Potter (2004, 312-316).

particular define a los conjuntos como una especie de clase y, con ello, deja la puerta abierta para postular clases de objetos que son "demasiado grandes" para ser un conjunto, como la clase de todos los objetos y de todos los conjuntos. Cantor había hablado, en cambio, de pluralidades consistentes y pluralidades "inconsistentes", esto es, exactamente las pluralidades demasiado grandes para las operaciones conjuntistas.

> En efecto, una pluralidad puede ser de tal índole que el supuesto de que *todos* sus elementos "existen conjuntamente" lleva a una contradicción, de modo que es imposible captar esa pluralidad como una unidad, como "una cosa acabada". A tales pluralidades las llamo *pluralidades absolutamente infinitas* o *inconsistentes.* (...)
>
> En cambio, si la totalidad de los elementos de una pluralidad se deja concebir sin contradicción como "estando reunida", de modo que es posible captarla conjuntamente como "una *cosa*", la llamo *pluralidad consistente* o "conjunto".[24]

Las paradojas, que bien podrían ser llamadas "paradojas cantorianas" tal como se denominan aquí los argumentos filosóficos contra la existencia de las totalidades que se han indicado, condujeron al desarrollo de las axiomatizaciones de teoría de conjuntos y, en particular, a la concepción iterativa de conjunto. En esta concepción, un conjunto sólo puede venir dado en virtud de operaciones conjuntistas que, luego, pueden aplicarse iterativamente sobre los conjuntos ya definidos. Una fórmula bien formulada, por ejemplo, no define un conjunto, esto es, el conjunto de todos los objetos que satisfacen la fórmula, si no es suponiendo que esos objetos *ya* forman parte de un conjunto. Por ejemplo, en el famoso axioma de separación de Zermelo (axioma III en su axiomatización de 1908) se establece que dada una fórmula φ se define el conjunto de todos los objetos x que satisfacen φ, $\{x: \varphi x\}$ si y sólo si hay un conjunto y, tal que $(x \in y \wedge \varphi x)$.[25] No toda expresión bien

---

[24] G. Cantor, carta a R. Dedekind del 3 de agosto de 1899. Citado y traducido por R. Torretti (1998, 51-52). Cantor hace esta distinción en vistas a la paradoja de Burali-Forti para sostener que hay totalidades que se encuentran definidas por una expresión verbal, pero que no pueden considerarse conjuntos (para la paradoja de Burali-Forti, cf. R. Torretti (1998, 465-468); la paradoja trata sobre la inexistencia del ordinal del conjunto de todos los ordinales).

[25] Cf. R. Torretti (1998, 78); los axiomas de Zermelo en 1908 son: (I) Axioma de determinación, dos conjuntos son idénticos si y sólo si poseen los mismos elementos; (II) Axioma de los conjuntos elementales, hay un conjunto vacío, $\varnothing$, que no posee elementos; (III) Axioma de separación, tal como se indicó; (IV) Axioma del conjunto potencia, dado un conjunto x, existe un conjunto P(x) que tiene como elementos todos los sub-conjuntos de x; (V) Axioma de unión, si existe el conjunto x, entonces existe el conjunto unión Yx compuesto por todos los elementos de los elementos de x; (VI) Axioma de selección, si hay un conjunto x que posee como elementos conjuntos no-vacíos disjuntos (esto es, sin elementos en común), entonces el conjunto unión Yx incluye un conjunto en el que se contienen como elementos uno y sólo un elemento de cada uno de

formada en algún lenguaje define un conjunto, por lo tanto, si es que el conjunto en cuestión no viene ya dado en la jerarquía conjuntista generada iterativamente por las operaciones especificadas en los axiomas. Así, no basta que exista la expresión "conjunto de todos los conjuntos" para que exista un conjunto de todos los objetos que satisfagan tal fórmula. No existe, en efecto este conjunto.

La forma en que von Neumann aborda esta cuestión es instructivamente diferente. En vez de declarar la inexistencia de ciertas totalidades (por ser inconsistentes), define los conjuntos como entidades que satisfacen ciertas condiciones y a las que les serán aplicables los usuales principios iterativos. Von Neumann no necesita exorcizar fuera de la existencia a las totalidades "demasiado grandes" para especificar la estructura conjuntista iterativa. La noción primitiva con la que se define "conjunto" en su axiomatización es la de función.[26] Se distinguen entre "cosas-I" y "cosas-II" (*Dinge-I, Dinge-II*). Las cosas-I son los argumentos sobre los que se puede aplicar una función, en adelante, los "argumentos". Las cosas-II son las funciones. Una función puede también ser argumento de otra función y se denomina una "cosa-I/II" o "función-argumento". El argumento que resulta del valor de la función f para el argumento 'a', se expresa como [f, a]. Aquí un "conjunto" se define del siguiente modo: sea un "dominio" una función f tal que, para todo argumento x, o bien [f, x] = 0, o bien [f, x] = 1. Aquí 0 y 1 son elementos distinguidos para indicar, intuitivamente que algo satisface la función f en cuestión o no la satisface ('A' y 'B' respectivamente en el original). Se dice de un dominio f que es un *conjunto* si y sólo si, f es una función-argumento. La expresión "$a \in f$" es una abreviatura de $[f, a] \neq 0$, en que f es una función y 'a' es un argumento. Intuitivamente, un conjunto queda seleccionado como: (i) la colección de objetos que satisfacen una cierta función (esto es, la colección de objetos tales que, para una función f determinada, [f, x] = 1), y que (ii) es argumento de otra función. De acuerdo a estas definiciones, cabe la posibilidad de que exista una cierta función f tal que no exista otra función g y $f \in g$. Una función f semejante selecciona una totalidad de elementos que

---

los sub-conjuntos no-vacíos disjuntos de x; (VII) Axioma de infinitud, existe un conjunto Z tal que: (i) $\varnothing \in Z$, y (ii) si $x \in Z$, entonces $\{x\} \in Z$ (lo que genera una secuencia infinita $\varnothing$, $\{\varnothing\}$, $\{\{\varnothing\}\}$, etcétera) (cf. R. Torretti (1998, 76-80, 471-472)

[26] Cf. para esta explicación R Torretti (1998, 90-101). Von Neumann formuló los axiomas de teoría de conjuntos en 1928 y constan de cinco grupos.

no constituye un conjunto. Es crucial para la idea de lo que hoy se entiende como una "clase propia" o "dominio propio" el axioma IV.2:

(IV.2) Una cosa-II *a* no es una cosa.I/II si y sólo si hay una cosa-II *b* tal que para cada cosa-I *x* existe un *y* que cumple las condiciones $[a, y] \neq A$ y $[b, y] = x$.[27]

De acuerdo a las definiciones dadas, el axioma IV.2 expresa que una cierta función f (*a* en la formulación de von Neumann) no es una función-argumento (y, por tanto, no es un conjunto) si y sólo si se cumple la siguiente condición: para cada objeto x, sea conjunto o no, hay una función g (*b* en la formulación de von Neumann) tal que a cada elemento y de f (esto es, a cada objeto y, tal que [f, y] = 1, o en otras palabras, tal que $y \in f$) le asigna x. En otras palabras, se dice de un dominio que es un dominio propio si hay una función que aplica ese dominio en el universo de todos los argumentos, compuesto por objetos y conjuntos. Un dominio deja de ser un conjunto cuando es "tan grande" como la colección de todos los objetos ordinarios compuesta por conjuntos y elementos de los conjuntos.[28] Recuérdese que al decir aquí que "no es un conjunto" se está diciendo que no es argumento para otra función y, por lo tanto, que no es elemento de otro conjunto. En otras palabras, un dominio "demasiado grande" queda automáticamente fuera de la estructura iterativa. No hay aquí ninguna necesidad de declarar a estos dominios "inconsistentes" para elaborar una teoría restringida a cierta clase de objetos matemáticos, a saber, los conjuntos, para los que se definen las propiedades usuales.

¿Qué se puede decir de estas diferentes axiomatizaciones por lo que respecta a las paradojas cantorianas que se están discutiendo aquí? En principio, las dos axiomatizaciones son equivalentes en el sentido de que todo lo que puede probarse con una de ellas, puede

---

27 Citado por R. Torretti (1998, 97).

28 Recuérdese que una "función" o "aplicación" se define como una relación entre dos colecciones de objetos A y B, tal que a elementos de A (llamado usualmente el dominio de la función), que pueden ser uno o varios, les asigna un y sólo un elemento de B (llamado usualmente el recorrido o dominio inverso de la función). Aquí se está diciendo que una función f no es una función-argumento, esto es, no es argumento de alguna otra función (no es conjunto) si y sólo si todos los elementos de f pueden ser aplicados en el dominio de todos los argumentos. No es necesario que la aplicación de los elementos de f en la colección universal de todos los argumentos sea tal que a cada elemento diferente de f le asigne un elemento diferente de tal colección universal (esto es, no es una función inyectiva), pero sí se requiere que cada elemento de la colección universal sea asignado a algún elemento de f (esto es, se trata de una función epiyectiva) Esto garantiza que el dominio definido por la función f debe ser al menos tan grande como la colección universal de todos los argumentos.

también probarse con la otra. No hay razones para preferir una de ellas por sobre la otra en lo que respecta a las cuestiones de teoría de conjuntos. Esto no impide que consideraciones desde un punto de vista externo a las puramente conjuntistas hagan una de ellas más preferible, por ejemplo, por razones de perspicuidad ontológica. La solución obvia sería sostener aquí que el "mundo" (y luego, por extensión, un "mundo posible") es una totalidad demasiado grande para ser un conjunto y debe tomarse como una clase propia, en el sentido que tiene esta expresión en NBG. Esta suposición, sin embargo, ha sido rechazada por algunos filósofos, como Christopher Menzel.[29] La razón fundamental para negar que el mundo pueda ser concebido como una clase propia en el sentido de NBG es que en estas concepciones conjuntistas la diferencia entre conjunto y clase viene dada no sólo por una consideración debida al "tamaño" de uno y otro tipo de totalidad sino también por una diferencia estructural. En la concepción iterativa los conjuntos sólo existen al ser "construidos" por operaciones sobre una colección inicial de elementos básicos (*urelements*: ∅ y cualesquiera otros átomos que sean dados): unión, pareo, conjunto potencia, separación y reemplazo. Dada la colección inicial de *urelements* y estas operaciones, se genera una jerarquía acumulativa de niveles, a cada uno de los cuales se asigna un número ordinal y contiene todas las colecciones que pueden ser formadas en los niveles previos en esta jerarquía. Aquí una clase propia está fuera de esta jerarquía acumulativa precisamente porque es el resultado último de este proceso de construcción de conjuntos. Anota Menzel:

> En términos más formales, V [esto es, la clase propia] es *ilimitada* (*unbounded*), esto es, contiene miembros de rango arbitrariamente alto (el rango de un objeto se define recursivamente como el menor ordinal más grande que los rangos de todos sus miembros); esto es, uno podría decir, de complejidad arbitrariamente alta en relación con las operaciones para la construcción de conjuntos.[30]

Así, una clase propia se habrá de distinguir no sólo por un tamaño desorbitado sino también por una estructura interna desorbitadamente compleja. Una clase propia se encuentra "más allá" de cualquier operación conjuntista con la que pudiese ser construida y, por eso, posee una naturaleza completamente diferente a la de un conjunto. Pues bien, considérese el

---

29 Cf. Ch. Menzel (1986, 68-72).

30 Ch. Menzel (1986, 70).

conjunto de todas las proposiciones verdaderas. Este conjunto está integrado únicamente por proposiciones, que no son conjuntos, por lo que tienen un rango 0. El conjunto de todas las proposiciones verdaderas tiene, entonces, un rango perfectamente claro de 1 (el rango inmediatamente superior al rango 0 en la estructura conjuntista). Resultaría que el conjunto de todas las proposiciones verdaderas no puede decirse una clase propia en el sentido de NBG.[31] Algo semejante podría decirse del conjunto de todos los estados de cosas, el conjunto de todos los objetos o el conjunto de todas las propiedades.

Creo, sin embargo, que estas peculiaridades de una clase propia en las axiomatizaciones de tipo NBG no son realmente un impedimento para el desarrollo de estrategias de ampliación. El problema de fondo que aquí se discute es si los resultados cantorianos deben tomarse como una limitación ontológica sustantiva en el "tamaño" de las entidades que puedan postularse, esto es, si es que el resultado de Cantor debe tomarse como un motivo suficiente para rechazar la existencia del mundo. En particular, ¿en qué sentido se debe decir que una cierta totalidad es "inconsistente", tal como lo ha sostenido Cantor? Pareciera que una totalidad es "inconsistente" cuando, dada esa totalidad y dadas las operaciones conjuntistas habituales –por ejemplo, la operación de conjunto potencia– se puede deducir una inconsistencia explícita. En todos los argumentos cantorianos que se han considerado se ha podido apreciar exactamente la misma estructura. La postulación de cierta totalidad *como un conjunto* permite deducir una conclusión en abierta contradicción con el teorema de Cantor. Cuando se habla de una totalidad "inconsistente" se está hablando de que existe un motivo para rechazar la postulación de tal totalidad. Ya se ha visto cómo von Neumann ha evitado tener que hablar de tales totalidades "inconsistentes" mediante el procedimiento de diferenciar los conjuntos, como parte de la estructura iterativa, de las "clases propias". La forma en que se han definido las clases propias en NBG, sin embargo, está restringida al tipo de paradojas que se han tenido directamente en vista para la axiomatización de teoría de conjuntos. Las dificultades en cuestión no tenían que ver con problemas ontológicos sino con ciertos conjuntos peculiares, tal como el conjunto de todos los conjuntos que no son miembros de sí mismos, el conjunto de todos

[31] Una conclusión semejante en P. Grim (1986, 187-188). Grim hace anotar, además, que una clase propia, que no puede ser incluida en otras clases más "grandes", no es aceptable para un teórico que quiere hablar de la clase de todas las verdades, pues, por ejemplo, este teórico estará interesado en afirmar que la clase de todas las proposiciones verdaderas forma parte de la clase de todas las proposiciones.

los conjuntos, el ordinal de todos los ordinales, etc. Estas totalidades problemáticas para la teoría intuitiva se pueden tratar de una manera satisfactoria con los axiomas de ZF o de NBG, y, por consiguiente, con la concepción iterativa que se expresa en ellos.

Puede afirmarse, sin embargo, de una manera igualmente legítima, que los problemas ontológicos que se discuten aquí, los que se han denominado paradojas o aporías cantorianas, son dificultades que obligan a introducir limitaciones en el alcance de los axiomas. Es efectivo que Cantor pensaba en los principios conjuntistas –los principios intuitivos– como las formas universales en que puede ser pensada cualquier totalidad actual o posible. Esto es algo, sin embargo, que no tiene por qué admitirse si es que hay totalidades que tienen un valor ontológico fundamental y no se adecuan a tales principios. No importa para esto si es que estas totalidades no pueden ser consideradas "clases propias" de acuerdo a lo que se pudo prever en NBG. Esto es, si sostener que cierta totalidad *es* un conjunto lleva a una contradicción, entonces, puede rechazarse que exista tal totalidad, pero también puede rechazarse que se trate de un *conjunto*. Las paradojas cantorianas consideradas hasta ahora pueden pensarse como dependiendo de una premisa de este tenor, que es una paráfrasis de la tesis (I) tal como se indica arriba:

(I') Hay una totalidad A, y

(I'') A es un conjunto

Aquí se rechaza la tesis (I'') en virtud del argumento cantoriano, pero no se rechaza la tesis (I') según la que existe una cierta totalidad A. Ahora bien, podría sostenerse que esta maniobra es poco razonable y *ad hoc*, esto es, motivada sencillamente para evitar conclusiones ontológicas incómodas. En efecto, podría decirse que los principios conjuntistas recogidos en los axiomas son plausibles independientemente o, al menos, parecen ser independientemente plausibles. Parece natural pensar que dada una colección de objetos, se puede definir la colección de todas las colecciones formadas de elementos de la colección inicial. Parece natural pensar también que, no importa el tamaño de la colección de que se trate, la nueva colección definida mediante la operación de conjunto potencia tendrá una cardinalidad mayor. Así acaece cuando la colección es finita y nada

parece impedir que se generalice esta tesis para colecciones infinitas de objetos. ¿Por qué no pueden aplicarse estos principios a totalidades como el mundo? ¿Qué tipo de veda o prohibición puede impedir que tratemos de pensar en el mundo con los mismos principios con los que se piensa cualquier colección?

La respuesta que se está proponiendo aquí es que la negación de que ciertas totalidades sean conjuntos, aún cuando esto no sea una alternativa contemplada por las axiomatizaciones existentes, es la salida más razonable para una situación teóricamente muy insatisfactoria. Considérese nuevamente la situación dialéctica. La postulación de que existe el mundo *como* un conjunto de todas las proposiciones verdaderas, o el conjunto de todos los estados de cosas o todos los objetos, conduce a una contradicción explícita con el teorema de Cantor, dado un principio generador para estas totalidades que permite introducir elementos en el conjunto en cuestión por cada sub-conjunto de éste. Sería irrazonable aquí declarar falsa *toda* la teoría de conjuntos, por haber conducido a este resultado.[32] Sería igualmente irrazonable declarar que no existe el mundo dado este resultado en contradicción con el teorema de Cantor. ¿Qué opción queda? Quedan abiertas las estrategias de limitación o la salida que se defiende aquí: sostener sencillamente que el mundo no es un conjunto. Se verá más adelante que las estrategias de limitación imponen tales restricciones que no es verosímil creer que puedan ofrecer respuestas a todas las paradojas cantorianas y, en especial, a las paradojas cantorianas de alcance ontológicamente más fundamental.

La principal justificación independiente que tiene la solución general que se propone es que los axiomas de teoría de conjuntos, y la concepción iterativa que expresan, no son *naturales*. No se trata de una serie de postulados obvios para el sentido común o, si se quiere, no se trata de una serie de postulados que sean todos ellos obvios para el sentido común. Si el axioma de conjunto potencia puede resultar "natural" en el sentido de parecer

[32] Una idea de este estilo, por ejemplo, ha sido atender a teorías de conjuntos no estándar en las que no vale el axioma de conjunto potencia o, al menos, no vale con toda generalidad (cf. Ch. Menzel (1986, 71); P. Grim (1986, 188-190)). Por supuesto, desde un punto de vista puramente formal tiene interés considerar qué acaece cuando se elimina algún postulado dentro de un conjunto de postulados independientes, pero no hay motivos para pensar que *toda* teoría de conjuntos deba pensarse sin axioma de conjunto potencia. Aún cuando existan estos sistemas no estándar, esto no es impedimento para la existencia de los sistemas estándar en los que sí vale este axioma y se siguen los resultados desastrosos en comento. Para evitar las paradojas habría que sostener que la única teoría de conjuntos válida es sin conjunto potencia y esto es completamente irrazonable. Los sistemas estándar son los que dan el paraíso de Cantor, por lo demás, así es que es ilusorio pensar que los matemáticos van a renunciar a ellos por los pruritos de un metafísico.

un supuesto perfectamente aceptable para cualquier persona racional, sea o no matemáticamente sofisticada (lo que también puede suceder con el axioma de extensionalidad o determinación, con el axioma de pareo o con el axioma de unión), no sucede lo mismo con otros axiomas, como el de separación, el de reemplazo, el de infinito, tal vez también con el axioma de elección, ni tampoco con la postulación de una entidad matemática misteriosa: $\varnothing$ (cuya utilidad formal para que las operaciones conjuntistas queden siempre definidas es obvia, pero que no puede reclamar la misma obviedad para el sentido común). La única razón que parece justificar un axioma como el de separación es la de evitar el surgimiento de paradojas como la de Russell. Recuérdese que el axioma de separación permite introducir el conjunto de todos los objetos x que satisfacen una condición φ $\{x: \varphi x\}$ si y sólo existe un conjunto y, tal que $(x \in y \wedge \varphi x)$. Esto es, un conjunto sólo se da, si es que existe ya otro conjunto. Los conjuntos sólo existen –de acuerdo a la teoría– dentro de la jerarquía conjuntista, como elementos de otros conjuntos. Esto no es un supuesto *natural* de ninguna manera. Es una forma razonable de estipular una teoría consistente, hasta donde sabemos, que recoge todas las intuiciones centrales de Cantor y que permite habitar su paraíso con razonable confianza, pero no es una serie de principios sobre *toda* colección pensable como tal sin contradicción. Las estructuras conjuntistas están diseñadas para ciertas finalidades teóricas específicas. En tales estructuras se hacen pensables totalidades infinitas de diferentes cardinalidades y se hace pensable una aritmética de infinitudes que, de algún modo, son diferentes y conmensurables entre sí. Pero es claro que no se trata de *los* principios que hacen pensable toda colección de objetos en general. De hecho, no pensamos en colecciones de objetos ordinariamente como parte de la estructura conjuntista. Pues bien, si esto es así, ¿es tan difícil de aceptar que el mundo no sea un conjunto, esto es, un objeto de *esta* teoría?

Lo que se está afirmando cuando se dice que el mundo no es un conjunto es sencillamente que no se le debe aplicar el aparato completo de esa teoría. Un "conjunto" es un tipo de entidad que satisface requerimientos bien específicos y no cabe pensar el mundo como una entidad de esa especie característica. No es tampoco el mundo una clase propia en el sentido de NBG, pero puede sostenerse que la negación de que el mundo sea un "conjunto" es una maniobra inspirada en su espíritu en von Neumann, aunque –como es obvio– no pueda atribuírsele esta idea en particular. El punto crucial es que no es razonable

sostener que los axiomas de teoría de conjuntos son la medida de lo que hay en cielos y tierra. No es razonable sostener que una entidad sólo puede existir si es que ocupa un lugar en la jerarquía conjuntista, esto es, si es que ha sido generada por operaciones conjuntistas. No es verosímil sostener que el axioma de conjunto potencia va a ser la clave de bóveda de nuestra ontología fundamental, funcionando como el discriminador de lo existente. La postulación del mundo como totalidad de todos los entes o como totalidad de todos los estados de cosas (y la postulación correlativa de la totalidad de todas las proposiciones verdaderas) tiene suficientes garantías independientes como para que debamos rechazar todo razonamiento en su contra como una *reductio* de los supuestos que conducen a tal resultado.[33] Esto implica que la teoría de conjuntos es una forma de estructura matemática más, formalmente muy simple y poderosa (y por ello, interesante para el matemático), pero no es nuestra ontología fundamental. Por supuesto, esto parece ir en contra del deseo expreso de Cantor, pero el proyecto original de Cantor fracasó ya de todos modos.[34]

3.2. *Estrategias de restricción*

Tal como se ha visto, las paradojas cantorianas surgen de dos supuestos. Por un lado se supone que existe cierta totalidad determinada A (supuesto (I)) y después se indica un principio generador que multiplica los elementos de esa totalidad A para hacerla tan "grande" como el conjunto potencia de A (supuesto (II)). Las estrategias de ampliación buscan negar el supuesto (I), rechazando que la totalidad en cuestión sea un conjunto, sin tener que rechazar la existencia de la totalidad A. Estas otras estrategias de limitación, en cambio, buscan restringir de una manera fundada la operación del principio generador que multiplica los elementos de A. Considérese, por ejemplo, lo que acaece en el argumento cantoriano (A). Se postula la existencia del conjunto A de todas las verdades. Para cada

---

[33] Es instructivo en este respecto comparar estas consideraciones con lo defendido recientemente por T. Williamson (2003). Williamson aboga por la necesidad de la cuantificación irrestricta sobre "todo" como un requerimiento de nuestra racionalidad, a pesar de los pruritos cantorianos de los lógicos. Señala: "¿Qué tiene que ver esto con la generalidad irrestricta? Muestra que si generalizamos sobre todo, no estamos generalizando simplemente sobre los miembros de algún conjunto o clase, tal como los conjuntos y las clases son ahora concebidas usualmente." (425)

[34] Cf. R. Torretti (1998, 63-70). Cantor dejó sin resolver los dos problemas de los que depende la suficiencia de la aritmética transfinita cantoriana para "medir todas las multitudes del universo" (63): el problema del continuo y el problema del buen orden. El primero se ha mostrado indemostrable desde los restantes axiomas conjuntistas usuales, tal como mostró Cohen en 1963. El segundo se ha podido resolver introduciendo el axioma de selección que ha resultado menos intuitivo para algunos que los restantes postulados de la teoría conjuntista.

sub-conjunto de A existe una nueva proposición verdadera que también debe ser una verdad y, por lo tanto, debe ser parte de A. A es tan grande como su propio conjunto potencia. El principio generador aquí funciona formulando proposiciones verdaderas por cada sub-conjunto del conjunto A. El caso del argumento cantoriano (B) es muy similar a éste, sólo que las proposiciones verdaderas deben aquí ser sustituidas por estados de cosas o hechos. En el caso del argumento cantoriano (F), relativo al conjunto de todas las entidades, por otra parte, el principio generador está explícitamente especificado como el principio mereológico según el cual dadas dos entidades x e y, existe la suma mereológica (x + y). Pues bien, las estrategias de limitación son formas de impedir que se agreguen nuevas entidades al conjunto en cuestión por cada sub-conjunto de éste. Para esto es indispensable que la totalidad que se ha definido sea restringida en un sentido preciso.

Las estrategias de restricción buscan típicamente efectuar cierta "estratificación" de los elementos que van a conformar la totalidad en cuestión, de manera que los nuevos elementos que puedan ser generados por cada sub-conjunto queden fuera al corresponder a otro "estrato" más alto. La tradicional teoría de los tipos de Russell es una forma de efectuar esta estratificación. Se considerará la cuestión en primer lugar respecto del conjunto de todas las proposiciones verdaderas. Se define, en primer lugar, un conjunto de todos los objetos $\{x_1, x_2, \ldots, x_n\}$ (recuérdese que esta suposición está ya afectada por el argumento cantoriano (F)) y un conjunto de todas las propiedades de primer orden $\{P_1, P_2, \ldots, P_n\}$ (esta suposición está afectada por el argumento cantoriano (G)). Se entienden aquí como "propiedades de primer orden" a las propiedades, posiblemente n-ádicas, que se atribuyen a objetos. Dados estos conjuntos iniciales de objetos y propiedades se puede definir un conjunto de todas las proposiciones de estrato-0 que sólo versan sobre propiedades y relaciones entre objetos, $\{P_1x_1, P_1x_2, \ldots, P_1x_n, P_2x_1, P_2x_2, \ldots, P_2x_n, \ldots, P_nx_1, P_nx_2, \ldots, P_nx_n\}$. Un sub-conjunto de estas proposiciones de estrato-0 es el conjunto de todas las proposiciones verdaderas. Sea A. Pues bien, es obvio que a cada sub-conjunto de A se le puede asignar una proposición verdadera. La idea es definir el estrato de tales proposiciones de manera que queden fuera del conjunto A. Se definirá ahora al estrato-1 como el estrato de todas las proposiciones que versan sobre proposiciones del estrato-0. Dada la manera en que se formuló el argumento A arriba, a un sub-conjunto dado de A, sea $\{p_1, p_2\}$ se le puede asignar una proposición verdadera como $p_1 \in \{p_1, p_2\}$. Si aquí p_1 y p_2 son

proposiciones de estrato-0, esto es, son proposiciones que están enunciando que ciertos objetos poseen ciertas propiedades de primer orden, la proposición que enuncia que p_1 pertenece al conjunto compuesto por las proposiciones p_1 y p_2 es una proposición de estrato-1, pues versa sobre proposiciones de estrato-0 (nótese que para esto no es necesario decir que el conjunto cuyos elementos son de estrato-0 debe ser de un estrato más alto. La estratificación sólo afecta a las proposiciones y no a las entidades de las que tratan esas proposiciones; aquí un conjunto es sencillamente una entidad más). Luego, pueden ser definidas proposiciones de estratos más altos. Las proposiciones de estrato-2 son proposiciones que versan sobre proposiciones de estrato-1. Las proposiciones de estrato-3 versan sobre proposiciones de estrato-2. Etcétera. Por cada uno de estos estratos, por otra parte, es necesario definir un conjunto particular de propiedades de diferentes niveles. El conjunto de propiedades de primer orden que sirve para definir el estrato-0 de proposiciones está compuesto sólo por propiedades que se predican de objetos, propiedades de primer nivel. Las propiedades que sirven para definir el estrato-1 es un conjunto de propiedades que se predican de las proposiciones de estrato-0, como, por ejemplo, la propiedad de ser un elemento del conjunto $\{p_1, p_2\}$, esto es, $[\lambda x\ (x \in \{p_1, p_2\})]$, que es una propiedad de segundo nivel. Luego habrá propiedades de tercer nivel que se predican de proposiciones de estrato-2 (que a su vez, predican propiedades de segundo nivel a proposiciones de estrato-1). Etcétera.

De esta forma, al parecer, se resuelve la paradoja cantoriana (A). El problema, sin embargo, es que esta solución no sirve de nada si es que no se resuelven también las restantes paradojas cantorianas (B), (F) y (G). Al resolver al problema local del conjunto de todas las proposiciones verdaderas no se ha salvado todavía el "mundo", pues esta solución debe hacer apelación al conjunto de todos los objetos, que está afectada por el argumento (F), y debe hacer apelación al conjunto de todas las propiedades, que está afectado por el argumento (G). Además, subsiste el problema respecto al conjunto de todos los estados de cosas, esto es, el argumento (B). Recuérdese que debe existir esta totalidad de todos los estados de cosas si es que han de existir *truthmakers* para todas las proposiciones verdaderas. Sucede, entonces, que si se quiere limpiar de aporías a la noción de "mundo" no sirve implementar una solución de estratificación *local*. No va a ser de utilidad estratificar los estados de cosas sin estratificar las proposiciones, los objetos y las

propiedades. No sirve, en fin, estratificar a los objetos sin estratificar las propiedades, los estados de cosas y las proposiciones. La estrategia de restricción por estratificación sólo puede ser *global.* De otro modo, las paradojas cantorianas, arrojadas de un área, reaparecerán en otra.

En el caso de los objetos, tendrá que suponerse que hay un conjunto de átomos de estrato-0. Las sumas mereológicas de átomos tendrán estrato-1. Las sumas mereológicas de entidades de estrato-1 serán de estrato-2. Etcétera. Algo análogo tiene que realizarse con las propiedades. Habrá un nivel de propiedades simples, el estrato-0. Las propiedades estructurales formadas por complexión de propiedades de estrato-0 serán las propiedades de estrato-1. Luego, en general, las propiedades formadas por propiedades de estrato [n – 1] serán de estrato n. Cuando se llega a un estado de cosas, por otro lado, habrá que definir un estrato-0 de estado de cosas básicos o atómicos, que han de estar constituidos por objetos de estrato-0 y propiedades de estrato-0. Los estratos superiores estarán constituidos por estados de cosas de estratos más bajos. Etcétera. Todo esto debe luego reaparecer en la estratificación de las proposiciones. El resultado es un esquema general de una inmensa complejidad. Considérese lo que sucede con ciertos tipos de estados de cosas. Por ejemplo, muchos filósofos consideran las relaciones causales como estados de cosas que relacionan dos estados de cosas.[35] El hecho de que un cortocircuito cause el incendio de una casa es una cierta relación entre el estado de cosas consistente en darse tales y cuales eventos en un circuito eléctrico y el estado de cosas consistente en la combustión de una casa. Si el estrato del estado de cosas consistente en el cortocircuito es n, entonces el estrato de la relación causal entre el cortocircuito y el incendio debe ser, por lo menos, de [n + 1]. Esto es importante, porque si se va a indicar un conjunto de *todos* los estados de cosas y los estados de cosas en cuestión van a ser estados de cosas "básicos" o estados de cosas "atómicos", entonces no van a aparecer ahí las relaciones causales. Si el mundo es el conjunto de *todos* los hechos o estados de cosas, y se pretenden evitar las paradojas cantorianas mediante la estrategia de restricción, entonces el mundo descrito estará limitado al estrato-0 de hechos, pero resultará entonces que en el mundo no van a aparecer las relaciones causales. Esto, por supuesto, tiene sin cuidado a filósofos cuya concepción de la causalidad es anti-realista,

[35] Por ejemplo, cf. M. Tooley (1987); D. M. Armstrong (1997, 202-219; 2004, 125-144); D. H. Mellor (1995).

pero muchos otros filósofos sostienen que los hechos causales del mundo son hechos ontológicamente básicos y para ellos esto es profundamente insatisfactorio.

Podría, sin embargo, sostenerse que esto no es una dificultad mayor. Podría decirse, en efecto, que el estrato de estados de cosas al que debe aludirse para la definición del mundo como conjunto de todos los hechos debe ser lo suficientemente alto como para incluir las relaciones causales, lo que no es un nivel arbitrariamente alto. Esto no es tan simple, sin embargo. Cuando se habla de la relación causal entre un cortocircuito y el incendio de una casa, en efecto, se está hablando de un estado de cosas de un estrato inmensamente alto. Una casa es un objeto de un estrato inmensamente alto, pues es la suma mereológica de entidades de estrato inmensamente alto (una molécula es ya probablemente una entidad de estrato inmensamente alto). Para poder incluir en el mundo la relación causal entre el cortocircuito y el incendio de una casa tenemos que elevar el estrato de estados de cosas a un nivel muy elevado. ¿Qué tan elevado? No es nada claro. ¿Qué tan compleja es una casa como entidad? ¿Cuántas partes posee una casa?

Esto es quizás lo que resulta más dudoso de la estrategia de restricción: si ha de tener sentido debe aplicarse globalmente a proposiciones, estados de cosas, objetos y propiedades (y otras categorías ontológicas si es que se dan). Pero para esto debe tener sentido el que exista un estrato de composición mereológica preciso para cada entidad de las que pueblan el mundo y con las que tenemos contacto ordinario. Para asignar tal estrato definido, es necesario identificar un nivel básico en el que sólo se den "átomos" sin partes. De la misma manera, se hace necesario postular un nivel básico de propiedades no estructuradas. Sin embargo, es obvio que éstas son tesis ontológicas dudosas, en el mejor de los casos. No es sensato pensar que para resolver el problema de las paradojas cantorianas tengamos que salir al mundo a buscar "átomos" reales. Si se fuese a seguir esta estrategia, habría que suponer que, dadas las paradojas cantorianas, entonces *tiene* que existir un estrato de átomos en el mundo, pues de otro modo no podemos hacer inteligible la idea misma de *mundo*. Esto es lo que no parece razonable. No se quiere decir con esto que no pueda surgir en el futuro alguna argumentación filosófica definitiva e independiente para la postulación de átomos en el mundo, o que, lo que puede parecer más difícil, se descubra un átomo (esto es, un objeto material sin partes) mediante investigación empírica. La cuestión

es que no parece una estrategia apropiada hacer descansar la defensa de la inteligibilidad de la noción de "mundo" descanse en el éxito eventual de semejantes especulaciones.

4. **Conclusiones**

Se ha mostrado que las paradojas cantorianas no ofrecen ventajas particulares para alguna de las teorías modales. Los problemas que derivan de las paradojas cantorianas, a pesar de que han concentrado bastante atención como dificultades de principio de diferentes teorías actualistas, usualmente contra una u otra teoría en particular, no son cuestiones de especial urgencia. El problema de fondo es de carácter absolutamente general y afecta a cualquiera que pretenda utilizar la noción de "mundo" en su ontología. Incluso un filósofo que rechace toda modalidad metafísica, tendrá que abordar esta cuestión. No se trata, por tanto, de una dificultad específica de las concepciones modales y, además, no es una dificultad que permita otorgar ventajas o signifique desventajas para alguna teoría modal. La metafísica modal, en resumen, bien puede poner entre paréntesis esta cuestión para concentrar sus energías en las cuestiones que tienen verdadera relevancia para su desarrollo.

Por otra parte, cuando se trata de abordar la cuestión general, la estrategia de ampliación es la que parece ser la más recomendable. Esta estrategia postula sencillamente desatender las paradojas cantorianas. La teoría de conjuntos debe tomarse como una teoría matemática interesante, pero no como el criterio ontológico último de lo existente en cielos y tierra. Seguramente es necesario tener más claridad sobre cuál es el verdadero alcance de las estructuras conjuntistas y, también, más claridad sobre qué es una entidad matemática en general, para poder afirmar con seguridad que los resultados cantorianos tienen un valor restringido. Falta mucho para conseguir esta claridad, naturalmente, pero esto no obsta para postular la estrategia de ampliación como la más verosímil, dada el conocimiento que poseemos hoy sobre estas cuestiones[36].

José Tomás Alvarado Marambio
Pontificia Universidad Católica de Chile
jose.alvarado.m@ucv.cl

[36] Este trabajo ha sido redactado en ejecución del proyecto de investigación Fondecyt 1070339 (Chile). Agradezco los útiles comentarios de un evaluador anónimo de esta revista.

Bibliografía

Adams, R. M. (1979) 'Theories of Actuality' en M. J. Loux (ed.), *The Possible and the Actual. Readings in the Metaphysics of Modality,* Ithaca: Cornell U.P.

Alvarado, J. T. (2006), '¿Qué es el espacio ontológico modal?', *Philosophica* 29, 7-44.

Armstrong, D. M. (1989), *A Combinatorial Theory of Possibility,* Cambridge: C.U.P.

Armstrong, D. M. (1997), *A World of States of Affairs,* Cambridge: C.U.P.

Armstrong, D. M. (2004), *Truth and Truthmakers,* Cambridge: C.U.P.

Bringsjord, S. (1985), 'Are There Set Theoretic Possible Worlds?', *Analysis* 45, 64

Bringsjord, S. (1989), 'Grim on Logic and Omniscience', *Analysis* 49, 186-189.

Carnap, R. (1947), *Meaning and Necessity,* Chicago: University of Chicago Press.

Chihara, Ch. (1998), *The Worlds of Possibility,* Oxford: Clarendon Press.

Cresswell, M. J. (1979), 'The World is Everything That is the Case', en M. J. Loux (ed.), *The Possible and the Actual,* 129-145.

Davies, M. (1981), *Meaning, Quantification and Necessity,* London: Routledge.

Divers, J. (2002), *Possible Worlds,* London: Routledge.

Forrest, P. (1986), 'Ways Worlds Could Be', *Australasian Journal of Philosophy* 64, 15-24.

Grim, P. (1984), 'There is no Set of All Truths', *Analysis* 44, 206-208.

Grim, P. (1986), 'On Sets and Worlds: A Reply to Menzel', *Analysis* 46,186-191.

Grim, P. (1997), 'Worlds by Supervenience: Some Further Problems', *Analysis* 57, 146-151.

Jeffrey, R. C. (1965), *The Logic of Decision,* Chicago: McGraw-Hill.

Jubien, M. (1980), 'Problems with Possible Worlds', en D. F. Austin (ed.), *Philosophical Analysis,* Dordrecht: Kluwer, 299-322.

Kaplan, D. (1995), 'A Problem in Possible-World Semantics', en W. Sinnott-Armstrong, D. , D. Raffman & N. Asher (eds.), *Modality, Morality and Belief,* Cambridge: C.U.P., 41-52.

Lewis, D. (1973), *Counterfactuals,* Oxford: Blackwell.

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

Melia, J. (2001), 'Reducing Possibilities to Language', *Analysis* 61,19-29.

Mellor, D. H. (1995), *The Facts of Causation,* London: Routledge.

Menzel,Ch. (1986), 'On Set Theoretic Possible Words', *Analysis* 46, 68-72.

Plantinga, A. (1974), *The Nature of Necessity,* Oxford: Clarendon Press.

Plantinga, A. (2003) 'Actualism and Possible Worlds', en *Essays in the Metaphysics of Modality,* Oxford: Oxford U.P., 103-121.

Potter, M. (2004) *Set Theory and its Philosophy,* Oxford: Oxford U.P.

Raffman & N. Asher (eds.), (1995) *Modality, Morality and Belief,* Cambridge: C.U.P.

Roy, T. (1995) "In Defense of Linguistic Ersatzism" *Philosophical Studies* 80, 217-242

Sider, Th. (2002) 'The Ersatz Pluriverse', *The Journal of Philosophy* 99, 279-315.

Stalnaker, R. (2003) 'Possible Worlds', en *Ways a World Might Be. Metaphysical and Anti-Metaphysical Essays,* Oxford: Clarendon Press, 25-39

Tooley, M (1987). *Causation. A Realist Approach,* Oxford: Clarendon Press

Torretti, R. (1998) *El paraíso de Cantor. La tradición conjuntista en la filosofía matemática,* Santiago: Editorial Universitaria.

Williamson T. (2003) 'Everything', en J. Hawthorne & D Zimmerman (eds.), *Philosophical Perspectives* 17, 415-465