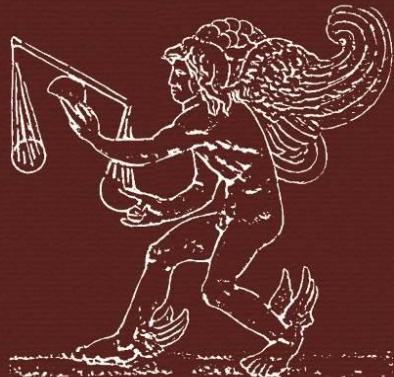


Kairos

Revista de Filosofia & Ciência
Journal of Philosophy & Science



9

Abril / April 2014

Artigos Papers

The Moral Lives of Animals
Michael Bradie

Scientist or Seer? Raymond Aron's Critique of Vulgar Marxism in *Introduction à la philosophie de l'histoire*
Scott Nelson

**Logic and Computers.
A Sketch of Their Symbiotic Relationships**
Jordi Vallverdú

**O problema em torno do conceito de "raça":
será que a "raça" é uma ilusão?**
Joana Inês Pontes

O problema mente-corpo na primeira década do século XXI: visita guiada a pontos-chave da paisagem fisicista
António Zilhão

Kairos. Revista de Filosofia & Ciência
Kairos. Journal of Philosophy & Science
ISSN: 1647-659X

Direcção / Editors

Olga Pombo

(Faculdade de Ciências
da Universidade de Lisboa)

Nuno Melim

(CFCUL)

**Comissão Editorial /
Editorial Board**

Catarina Pombo Nabais

(CFCUL)

Lídia Queiroz

(CFCUL)

María de Paz

(CFCUL)

Nuno Jerónimo

(CFCUL)

**Comissão Científica /
Scientific Board**

Andrea Pinotti

(Università degli Studi di Milano)

Angel Nepomuceno

(Universidad de Sevilla)

Byron Kaldis

(Hellenic Open University)

Francisco J. Salguero

(Universidad de Sevilla)

John Symons

(University of Texas, El Paso)

José Nunes Ramalho Croca

(Faculdade de Ciências
da Universidade de Lisboa)

Juan Manuel Torres

(Universidad Nacional de Cuyo,
Argentina)

Juan Redmond

(Universidad de Valparaíso, Chile)

Marcelo Dascal

(Universidade de Tel-Aviv)

Rudolf Bernet

(Husserl-Archives Leuven:
The International Centre
of Phenomenological Research)

Shahid Rahman

(Université de Lille)

Edição: Centro de Filosofia das Ciências da Universidade de Lisboa

Índice / Index

Resumos / Abstracts	5
Artigos / Papers	
The Moral Lives of Animals	
Michael Bradie	13
Scientist or Seer? Raymond Aron's Critique of Vulgar Marxism in <i>Introduction à la philosophie de l'histoire</i>	
Scott Nelson	29
Logic and Computers. A Sketch of Their Symbiotic Relationships	
Jordi Vallverdú	45
O problema em torno do conceito de “raça”: será que a “raça” é uma ilusão?	
Joana Inês Pontes	73
O problema mente-corpo na primeira década do século XXI: visita guiada a pontos-chave da paisagem fisicista	
António Zilhão	109

Resumos / Abstracts

The Moral Lives of Animals

Michael Bradie

Contemporary discussions of the moral status of animals typically address the key questions from an anthropocentric point of view. That is, in deciding whether non-human animals qualify as moral agents the discussion tends to center around the question of the extent to which members of candidate species are like human moral agents. Similarly, in deciding whether non-human animals qualify as appropriate moral patients, the discussion tends to center around the question of the extent to which members of candidate species have capacities and capabilities that are like the capacities and capabilities that are deemed necessary for human beings to be considered as moral patients. Viewing these issues from the point of view of evolutionary biology, cognitive neuroscience and cognitive ethology suggest a different approach, one which focuses on the lives of animals that takes their characteristic behaviors and capacities as the basis for determining the norms of moral status for them. One might call this approach a 'speciocentric' point of view. The motivation for taking this approach is the evolutionary consideration that the moral capacities and patterns of moral behavior of human beings are evolved capacities and patterns that are rooted in shared homologies that give rise to homologous capacities and patterns of behavior in other species. In this paper, I explore the theoretical and experimental results which make this approach plausible and address two key questions: [1] to what extent is it proper to speak of the *moral* behavior of non-human animals? [2] To the extent that it is proper, what are the implications for our understanding of the nature and function of human morality and of our treatment of non-human animals?

Scientist or Seer? Raymond Aron's Critique of Vulgar Marxism in *Introduction à la philosophie de l'histoire*

Scott Nelson

Thirty years after the death of Raymond Aron there are still many interpretations concerning his critique of Marxism. His primary doctoral dissertation, *Introduction à la philosophie de l'histoire*, is designed to explore the way in which our view of history changes on the basis of the present; however, this work also contains a critique of Marxism interpreted as a science of historical development. Marxism as a science is forced to submit to the rules of causal analysis and part of Aron's dissertation explores causal thought. The purpose of this paper is to investigate Marxism's amenability to causal systematization by way of Aron's analysis of causal thought. Marxism will be probed on both the sociological and historical levels and then finally recovered as more of an existential choice or philosophy than a comprehensive and rigorous science.

Logic and Computers. A Sketch of Their Symbiotic Relationships

Jordi Vallverdú

It is a well established fact that Logic and Computer Sciences have a long and close relationship, which has benefited both sides of this symbiotic relationship. Therefore, this is a case of mutualist symbiosis, a relationship in which both fields benefit. Nevertheless, and according to the main studies about this topic, this relationship could looks like as a parasitist or commensalist one: that is, one field benefits but the other is harmed (in the first case), or one field benefits, while the other remains unaffected (in the second one). Our departure hypothesis is: not only logic has influenced computer sciences development (a common topic on the historiographical approaches to computer sciences), but computer sciences have also changed the logical thinking and practice. This is a mutualist symbiosis. This second idea, the impact of computer sciences into logical field, has been overlooked by most of the researchers who have studied the relationships between both fields. This symmetric symbiosis reaches from practical to very fundamental aspects of the relationship.

O problema em torno do conceito de “raça”: será que a “raça” é uma *ilusão*?

Inês Pontes

Face à recente publicação do genoma humano e ao avanço da possibilidade de distinguirmos geneticamente minorias populacionais, o conceito de “raça” encontra-se, hoje, no centro de diversos debates políticos, científicos e filosóficos. Em consequência, intensificaram-se os trabalhos em metafísica da “raça”, sobretudo, as abordagens semânticas na procura de uma definição esclarecedora do conceito. Embora seja consensual que as “raças” não são tipos naturais (TN), as opiniões sobre aquilo que o termo “raça” significa divergem, não apenas na filosofia, mas na ciência, antropologia e psicologia. Partindo da consideração da tese de Anthony Appiah de que não há tal coisa chamada “raça” e colocando ênfase no problema de saber se as “raças” existem ou não (e se existem, o que são, afinal?) este artigo pretende mostrar que a “raça” existe, é uma entidade real e que Appiah não está correcto ao concluir que a “raça” é uma ilusão: i) Appiah fecha o debate num sistema circular: erra ao negar a possibilidade de desenvolvimento de um conceito de “raça” que não envolva essências; ii) a “raça” é simultaneamente, uma construção social e uma realidade biológica: as “raças” não diferem essencialmente, mas superficialmente, porém, são muito importantes para a formulação e uso do conceito ordinário de “raça”.

O problema mente-corpo na primeira década do século XXI: visita guiada a pontos-chave da paisagem fisicista

António Zilhão

Nos últimos anos do século XX e nos primeiros anos do século XXI assistiu-se, no âmbito da teorização materialista acerca do problema mente-corpo, a um ressurgimento da perspectiva conhecida como ‘teoria da identidade tipo-tipo’ (e.g., Block, Stalnaker, Papineau). Neste ensaio procuro analisar se este ressurgimento veio acompanhado de elementos novos que permitiriam distingui-lo essencialmente das propostas originais dos anos 1950 (e.g., Smart, Place, Feigl). Em particular, procuro determinar se as objecções que levaram ao declínio destas teriam finalmente encontrado uma resposta adequada. A conclusão a que chego é que isso não aconteceu; defendo, por isso, que não devem alimentar-se grandes expectativas em torno deste ressurgimento; o problema mente-corpo continua por resolver.

Artigos / Papers

The Moral Lives of Animals

Michael Bradie

(Department of Philosophy, Bowling Green State University)

mbradie@bgsu.edu

1. Introduction

Do animals lead moral lives? What exactly might be meant by claiming that they do and how might we be able to establish that fact? This is the focus of the following, programmatic, paper. My aim is to establish a framework for answering these questions and suggest a direction for further investigation.

Much of the literature on animals and morals focuses on the moral *status* of animals. Do they need to be considered in *our* moral calculations and if so how? A related concern is the extent to which the moral status of animals suggests or dictates human attitudes towards them and human practices with respect to them. To borrow a phrase from Peter Singer, the question is should the circle that encompasses the moral community of human beings be expanded to include some if not all animals? If so, what criteria are relevant for determining who is or is not to be included in this expanded circle? Typical criteria include the capacity to feel pain, the ability to have and fulfill preferences, evidence of a degree of rationality or reflective capacity, and the capacity for a sense of self, among others. The point is that each of these approaches reflects what might be called an ‘anthropocentric perspective’ insofar as each singles out a capacity or set of capacities that human beings possess to serve as the hallmark of moral status. As such, with respect to other animals, a key underlying question seems to be what the implications of including or excluding animals in the ‘moral circle’ are *for us*? They are anthropocentric in another sense as well in that who counts as morally

relevant is determined by criteria that are set by some understanding of human conceptions of morality.

My approach is somewhat different. The question I am interested in exploring is this: To what extent can we get a handle on the moral lives of animals from the perspective of the animals themselves? Does it make any sense and, if so, what sense, to talk of animals as leading moral lives independently of questions about how and whether to factor them into our moral deliberations? In terms of Singer's 'expanding circle' metaphor we may put the question in the following way: Is there one moral circle that encompasses all those who warrant moral consideration or are there perhaps a number of (possibly overlapping) circles centered around different focal points? Does it, for example, make sense to talk of a moral community of wolves or elephants where the norms of these communities and the criteria for membership are determined by and reflections of the social dynamics of the respective groups? In contrast to the traditional anthropocentric perspective this approach might be labeled 'speciocentric.'

The plan of the paper is as follows. Section 2 is a brief summary of the main empirical and theoretical considerations that suggest that at least some non-human animals lead moral lives that can be appropriately characterized from a speciocentric point of view. Section 3 raises two questions that need to be addressed if the project of attributing moral lives to animals is to get off the ground. Section 4 explores the sense in which animals might be construed as moral agents. This discussion draws on some recent work by Geoffrey Sayre-McCord on the nature of normativity.¹ Section 5 is a discussion of a moral version of what is known as the 'logical problem' in the theory of mind literature. This material draws on some recent work by Robert Lurz.² Finally, section 6 contains some brief remarks on the potential implications for the treatment of animals that may live more or less rich moral lives.

2. The empirical and theoretical background

Here I briefly summarize material that is dealt with more fully in 'The moral life of animals'.³ The general empirical and theoretical support for attributing

¹ Sayre-McCord, 2011.

² Lurz, 2011.

³ Bradie, 2011.

moral sensibilities to animals derives from three sources: evolutionary theory, neuroscience and cognitive ethology.

The argument from evolutionary considerations has its roots in the work of Charles Darwin and George Romanes.⁴ The basic idea is that human beings, other mammals and even more distant lineages have a shared evolutionary history. This history records the development of the shared underlying biological mechanisms that give rise to psychological and affective states. Different lineages may manifest those characteristics in different ways but the implication is that the differences between lineages are differences in degree and not differences in kind. The attribution of mental and affective states to animals was blocked by the rise and dominance of behaviorism in the first half of the 20th century. However, developments in neuroscience in the past 40 years have led some to challenge the behaviorist paradigm that rejects all attributions of mental or affective states to animals as anthropomorphism gone wild.

The evidence from neuroscience is extensive although the implications for attributing mental and affective states to non-human animals are still somewhat controversial. Two of the major figures advancing the view that the neuroscience strongly supports the view that animals do have minds and experience affects are Paul Maclean and Jaak Panksepp. The basic idea of Maclean's 'triune brain hypothesis' is that the evolved mammalian brain can be conveniently represented as the product of 3 developmental stages: A primitive reptilian brain located in the basal ganglia, an old mammalian brain located in the limbic system, and a new mammalian brain located in the neocortex.⁵ The triune brain thesis argues for deep homologies between the brains of animals and the brains of human beings.⁶ Neurological evidence points to deep structural similarities between the ancient brain systems that we share with other animals. In particular, the ancient structures are the neural source of basic qualitative *feels* or *affects*. Jaak Panksepp has identified seven primary limbic emotional action systems which, he argues,

⁴ Darwin 1981, 1998, 2000; Romanes 1882, 1885.

⁵ Maclean, 1990.

⁶ A recent discussion of the philosophical implications of the concept of homology for identifying psychological and behavioral traits across lineages can be found in a special issue of *Biology and Philosophy* edited by Ingo Brigandt and Paul Griffiths (2007, Volume 22, Number 5). Of especial interest are the articles by Marc Ereshefsky (2007) and Paul Griffiths (2007). Additional relevant resources include Griffiths (1996), Griffiths (2006), Matthen (1998), Matthen (2000), Matthen (2002),

are the basis of animal responsiveness and lie at the foundation of both emotional and cognitive states. In addition to this shared affective neurostructure, he has recently argued that mammals share brain structures that constitute what he calls “proto-selves” and “core selves.” Further study, he suggests, may reveal the basis for attributing a sense of self to a wide range of animals. It stands to reason, he argues, that animals with brain structures similar to those in humans not only react in ways that make them appear to have qualitative experiences similar to those of humans when the homologous brain structures are stimulated, but also that they do in fact have those experiences.⁷

The unregenerate behaviorists among you may object that the attribution of affects to non-human animals is unjustified anthropomorphism. Frans de Waal, among others, however, argues that it is not. de Waal argues that there is a double standard at work when charges of ‘anthropomorphism’ are tossed about.⁸ On the one hand, researchers take cognitive differences between humans and other animals to justify the non-attribution of emotional and mental capacities to animals while, on the other hand, they ignore evolutionary evidence that suggests that animals and human beings have shared inherited brain structures associated with emotional and mental capacities. de Waal labels this blind spot “Anthropodenial,” which he characterizes as the *a priori* rejection of the importance of the fact that although non-human animals are not human, humans are animals.

The third line of relevant scientific findings comes from investigations by cognitive ethologists. In their book *Wild Justice*, Marc Bekoff and Jessica Pierce argue from the perspective of cognitive ethology that animals exhibit behaviors that are best interpreted as manifestations of empathy, cooperation, and a sense of fairness. In essence, “animals have morality”.⁹ Bekoff and Pierce understand morality to be “a suite of interrelated other-regarding behaviors that cultivate and regulate complex interactions within social groups”.¹⁰ However, these behaviors do not constitute morality in themselves; a certain level of cognitive and emotional sophistication is necessary. Bekoff and Pierce’s approach is data-driven, and they emphasize the need and importance of expanding research beyond non-human primates

⁷ Panksepp, 2005.

⁸ de Waal, 2006

⁹ Bekoff and Pierce, 2009.

¹⁰ Bekoff and Pierce, 2009, ch. 7.

to other social mammals including hunting predators such as wolves, coyotes, and lions, as well as elephants, mice, rats, meerkats, and whales, among others. In addition, they emphasize the importance of studying animals in their natural habitats and not merely in the confines of laboratories where they are often asked to perform tasks in accordance with the interests and expectations of animal behaviorists, which may or may not reflect the interests of the animals themselves.

Where is the line to be drawn between animals that evince morality in this limited sense and those that do not? Bekoff and Pierce suggest that the line is shifting as more empirical evidence becomes available and as our philosophical understanding of what it means to be moral is modulated by reflection on the accumulation of scientific data especially data derived from field studies of animals interacting in their natural environments. Although their focus is on social mammals, there is a widening body of evidence that suggests that some birds have the wherewithal to constitute a moral community, in the sense of being capable of feeling and expressing relevant emotions, exhibiting co-operation, and the like.

Although they argue that the data strongly support the attribution of morality to animals, Bekoff and Pierce also argue that what constitutes morality has to be understood as species specific. Thus, what counts as morality for human beings may not count as such for wolves, for instance. Nevertheless, they argue, the fact that human standards of morality are not appropriate for wolves does not mean that wolves do not possess some sense of moral relationships that is exhibited in their own manifestations of empathy, cooperation and a sense of fairness. The net effect is that there is not one sense of moral community and that we humans, as allegedly prototypical moral agents, may expand our understanding of morality to include some organisms and exclude others. But the proper way to understand animal morality, they suggest, is to see that there are a number of distinct species-specific moral communities. Within these diverse communities, what counts as moral needs to be attuned to the characteristic features of the species themselves as opposed to being determined by considerations that derive from our own case. Indeed, even within species, different communities may develop different social practices, so that what is acceptable in one wolf pack, for example, may not be acceptable in another.

These considerations, taken together, are compelling support for the claim that at least some animals, especially the social animals, have moral lives.

The evidence is compelling but not conclusive. Putting aside behaviorist qualms there are still significant hurdles to be overcome before we can be confident in concluding that animals are moral creatures in their own right. To these qualms we now turn.

3. Two questions

There is an extensive literature on the dual questions of whether animals have minds and whether, if they do, they have a ‘theory of mind.’ There are two fundamental issues: (1) Do animals have minds?, and (2) Given that they do, are they capable of attributing mental states to others and acting on those attributions? Parallel questions can be raised with respect to the moral lives of animals. (1m) Do animals have moral lives, that is, are they motivated by ‘moral’ considerations, properly understood? (2m) Given that they are, can they attribute moral motivations to others and act accordingly?

Robert Lurz, in a recent book, has identified two fundamental issues that need to be addressed in order to be in a position to answer questions about the mindreading capabilities of animals.¹¹ One is theoretical and one is experimental. Parallel issues have to be addressed in order to be in a position to answer questions about the moral lives of animals.

For our problem, the theoretical issue is this: What does it mean to attribute moral lives to animals? In particular, what does it mean to attribute moral motivations to animals? The empirical issue is this: How best can we test for the existence moral sensibilities and moral motivations in animals?

I do not have a good answer to either of these two questions but I think we can make some headway in identifying the key questions that need to be answered and in identifying what is the proper perspective for answering them.

4. Levels of agency

To the extent that we attribute psychological and moral states to animals they are, in some sense, persons and not merely biological organisms. What, then, does it mean to attribute personhood to animals? We can adopt either

¹¹ Lurz, 2011.

an anthropocentric or a speciocentric perspective on this question. From an anthropocentric point of view, something is a person if it has a sufficient number of properties that make it an entity *like us*. I am not sure what constellation of properties this would include but the fact that in some legal sense corporations can be persons shows that the applicability of the concept is not limited to living beings. The central ideas that legitimate the extension of the idea of personhood to corporations, for instance, are notions of agency and responsibility. Corporations can act as (legal) agents and can be held (legally) responsible for their actions. However, what degree of agency and responsibility they possess is conferred upon them by human beings and their social practices. What about the moral agency of animals? Is that to be construed as merely derivative as well? A speciocentric perspective would reject this way of understanding what it means for an animal to be a moral agent. To the extent that animals lead moral lives (as opposed to being merely factors in our moral calculations) we must be able to construe them as moral agents in their own right. Can this be done?

In a recent paper responding to claims in the literature that attribute moral agency to animals, Geoffrey Sayre-McCord asks ‘Just what is it to be a moral agent?’¹² In effect, what is the nature of normativity? In his analysis, Sayre-McCord identifies several levels of ‘agency’ where an agent is understood to be something capable of representing its environment and acting on the basis of those representations.¹³ These are, in order of increasing sophistication, (1) ‘stimulus-response agents who ‘represent the world as being a certain way and then respond directly’¹⁴; (2) ‘planning agents,’ which are basically stimulus-response agents with the extra capacity to identify alternative courses of action and act in accordance with some plan of action. Sayre-McCord characterizes these agents as ‘decision-theoretical’ agents whose behaviors can be adequately modeled by decision theory¹⁵; (3) ‘strategic agents’ are agents who attribute designs and plans to others and act accordingly. Their behavior can be modeled by game theory; (4) ‘norm-governed agents’ are ‘strategic agents . . . [who] introduce rules for behavior with which they are disposed to conform and disposed to enforce in various ways’¹⁶; and finally, (5) ‘rational agents,’ that is, strategic norm-governed

¹² Sayre-McCord, 2011, 2.

¹³ Sayre-McCord, 2011, 5.

¹⁴ Sayre-McCord, 2011, 5.

¹⁵ Sayre-McCord, 2011, 5.

¹⁶ Sayre-McCord, 2011, 7.

agents who are ‘able to represent the different options as better or worse, as right or wrong, or as justified or not and . . . [are] able to act on the basis of such normative representations.’¹⁷

On Sayre-McCord’s account, truly moral agents need to be able to have a capacity for second-order reflection on first order states. That is, truly moral agents need not only follow norms but be capable of recognizing that they are following norms and be capable of using this reflective insight to guide their actions. This is a high bar for non-human animals to pass. Whether they are capable of passing it depends on how sophisticated their mental and psychological capacities are. Many who are willing to allow that some animals have sophisticated psychological states are reluctant to attribute reflective second-order capacities to them. Sayre-McCord, for one, allows that some animals are capable of rising to the level of norm-governed agents but he resists attributing any rational, and hence, truly moral, agency to them.

The distinction between norm-governed agents and rational agents roughly parallels Kant’s distinction between acting in accordance with duty and acting *from* duty. Indeed, this is the central theme of Sayre-McCord’s analysis.¹⁸ His main project is to provide a Kantian account of rational agency freed from the metaphysical baggage of Kant’s own account.¹⁹ For our present purposes, the question is: ‘Is norm-governed agency good enough for *non-human* moral agency? It is clear that many cognitive ethologists see the structured behavior of social animals as manifesting norm governed behavior. Some, perhaps sympathetic to the idea that such behavior doesn’t rise to the Kantian level of moral agency, are content to qualify such animals as ‘proto-moral’ beings. I don’t want to haggle over labels here but merely want to suggest that the resistance to qualifying animals as ‘truly’ moral may reflect a subtle anthropocentric bias. If we view human morality as one manifestation of a shared evolved set of homologous social enabling mechanisms, then the peculiar feature of rational reflectivity, as Sayre-McCord understands it, looks more like a refinement of a capacity that is shared among many lineages rather than as a defining characteristic. If so then we can tentatively accept norm-governed agency as moral agency enough and move on to the empirical question of how to establish whether any animals do live moral lives, so understood.

¹⁷ Sayre-McCord, 2011, 7.

¹⁸ Sayre-McCord, 2011, 2.

¹⁹ Sayre-McCord, 2011, 2.

At some point on the phylogenetic scale, we reach a point where moral behavior fails to exist. Sayre-McCord's analysis suggests that this point is high up, namely at the level of the human. Those sympathetic to the traditional analyses tend to agree. My analysis suggests that the dividing line lies lower down at the level of manifestations of sociality which are based on shared homologies. Is the difference between us to be attributed to a mere conflict of intuitions or is there something more to be said in favor of one approach rather than another? Sayre-McCord starts from the presumption that humans are the proto-typical moral agents and constructs an analysis on that basis that winds up focusing on the peculiar attributes that constitute the basis of human morality at least as construed from a broadly Kantian perspective. The starting presumption of the present analysis is that moral lives are grounded in the shared evolutionary homologies that manifest themselves in social behaviors across a range of related lineages. From this point of view, human morality is but one manifestation of a set of homologous behaviors and practices shared among a wide spectrum of related lineages.

5. The logical problem

Turning to the question of how to empirically test whether or not animals lead moral lives, we confront what has been labeled in the mind reading literature as the 'logical problem.' The mind reading problem is this: Is there any way to empirically distinguish between (1) animals that are mind readers, that is, animals that act in light of their attribution of intentional states to others, and (2) animals that are acting on behavioral cues but who do not attribute intentional states to others? The problem arises, in part, because, in the absence of language, the attribution of mind reading to animals is determined solely by their behavioral responses to environmental situations. In a recent book, Robert Lurz argues that all previous experimental results that suggest that some animals are mind readers are compromised by a failure to rule out the hypothesis that the observed behaviors can be explained equally well by a 'behavior-reading' hypothesis to the effect that the animals are responding to behavioral cues and are not attributing mental states to either other conspecifics or to the experimenters.²⁰

²⁰ Lurz, 2011, ch. 2.

Some take this failure to be able to discriminate between cases of mind-reading and cases of behavior-reading to be an insurmountable barrier to the unequivocal attribution of mind reading to organisms that cannot communicate their thoughts and intentions through the use of language that is intelligible to us. Lurz, however, argues that it is possible to design experiments that will be able to discriminate between the two hypotheses and he proposes several, as yet untested designs, that he claims will yield different predictions depending upon whether the tested animals are mind-readers or not. I do not want to pursue this here but rather to formulate the analogous problem for determining whether or not non-human animals live moral lives.

The logical problem for the moral lives question boils down to this: Is it possible to experimentally distinguish between animals that are acting in accordance with moral norms and animals that are behaving as *if* they were but for whom no moral considerations, *per se*, are relevant? If we allow, for the sake of argument, that morally motivated animals are norm-governed in Sayre-McCord's sense then what we want to know is whether the behavior of the animals is directed by (first-order) moral motivations or whether the characterization of their behavior as norm-governed is imposed upon their behavior by the ethological investigators.

This problem dogs much, if not all, of the cognitive ethology data that suggests that many social animals exhibit behaviors that can be interpreted as a result of the animals acknowledging and enforcing social and moral norms within their respective communities. Unlike Lurz, I do not have any good sense that these alternative accounts are empirically distinguishable. If they are not, then the claim that animals lead moral lives will remain in limbo despite the suggestive evidence from evolutionary considerations and the neuroscientific data. However, I am persuaded by the work of the cognitive ethologists that any decisive conclusions one way or the other must be the result of investigations *in situ* where experiments and observations are set up to reflect the conditions and expectations of the animals under investigation and *not* the expectations of alien investigators (that is, us).

6. Some moral implications

If we assume, for a moment, that at least non-human animals do lead moral lives of a sort, what are the implications for our understanding of the nature and function of human morality and of our treatment of non-human animals?

If the argument of this paper is on the right track it leads to the conclusion that the capacity for human morality is one manifestation of the social nature of our species, a nature that is, in turn, one manifestation of a shared inheritance that spreads across a wide spectrum of biological lineages and which manifests itself in a variety of ways. Thus, attempts to distance humans from other animals on the basis of the alleged fact that humans, as moral agents, are qualitatively different from the other animals, must be resisted. Even if it is conceded that the human capacity for morality is intimately connected with the human capacity for reason which other animals may or may not possess, this merely goes to show that different organisms manifest their moral proclivities in different ways. The speciocentric point of view suggests that any attempt to isolate the one defining and separating feature of any trait or capacity on the basis of the special circumstances and characteristics of any one species is misguided. This is not to deny the obvious, namely, that sociality and moral behavior manifests itself in different ways in different lineages, but merely to challenge the hubris of adopting one preferred perspective as the correct one. This, I take it, is a corollary of adopting a general Darwinian approach to the evolution of species.

What, then, are the implications for the human treatment of non-human animals? I take it, as a given, that independently of whether animals qualify as moral agents of a sort, given that they are sentient there is a *prima facie* reason for not inflicting gratuitous pain on them. What then does the recognition that they are capable of leading moral lives add to this? For starters, it does not mean that they need to be included in ‘our moral community,’ however that is understood. This is due to the fact that, although our shared evolutionary heritage binds us together, our evolutionary histories are sufficiently different to mean that, in some sense, we are more or less alien to each other. That said, the fact that they are moral beings in their own right means, I should think, that we owe it to them to be more sensitive to their needs and interests and overall well-being. One practical implication is with respect to the construction of zoos and animal parks. These institutions are,

for the most part designed with our interests and not the interests or well-being of the animals in mind. At the very least, the construction of such facilities should strive to recreate as closely as possible the natural environments of the animals. Similar considerations apply to efforts to preserve wild habitats and shelter them from human despoilment. Here the issues are complicated due to the fact that human interests driven by economic and political considerations are often at odds with the interests of populations of wild animals. There is no easy formula for resolving such conflicts but an increased awareness of the moral capacities of animals should weigh in as a relevant consideration in our deliberations about what to do or what to tolerate in particular circumstances. Exactly what further accommodations need to be made in reshaping human attitudes towards and treatment of animals hinges on the results of future empirical investigations that bring to light the hitherto hidden moral lives of animals.

More generally, if the speciocentric perspective is correct, what does the resulting moral landscape look like? From an anthropocentric point of view, there is one moral community and who gets included is a function of a set of centrally determined criteria. From a speciocentric perspective, however, there are any number of different and potentially disjoint moral communities with different criteria for inclusion. How are we to understand the inter-relationships between these different communities? The fact that, as human beings, we may be inclined to be sensitive to the well-being of non-conspecifics by no means entails that the members of other moral communities are or should be so inclined. One might even argue that, given the speciocentric point of view, human beings may have no particular obligations to non-conspecifics.²¹ If this is so, then it may be the case that shifting from an anthropocentric to a speciocentric perspective has no obvious implications for how humans should treat non-conspecifics. I am inclined to think that this is not the case on the grounds that, to the extent these non-conspecifics have more capacities that are similar to our own but manifested in different ways, they deserve some further consideration than traditional approaches to animal welfare would countenance. Is this a retreat to anthropocentrism as one reviewer suggested? I think not, since the key point of difference between the anthropocentric and the speciocentric points of view is not that we should not deem other creatures worthy or not of moral

²¹ The importance of addressing the architecture of the speciocentric moral landscape was urged by an anonymous reviewer.

consideration based on *our* values but rather that we should not decide whether other beings lead moral lives based on criteria which are drawn from our own case and appropriate for us.

7. Conclusion

Three lines of empirical evidence suggest that the capacity for human morality is one manifestation of homologous traits that are shared by other social organisms. The interpretation of the significance of this evidence is complicated by the fact that protocols for evaluating the evidence from the subject's point of view rather than from a human-centered point of view are still somewhat primitive. In addition, there are important philosophical reservations about the very idea of attributing moral sensibilities and moral lives to animals on the grounds that exercising moral behaviors requires mental and rational capacities that are beyond the capabilities of non-human animals. I have suggested that this reflects an anthropocentric bias that needs to be replaced by what I have called a 'speciocentric' point of view. The question of whether animals lead moral lives in their own right and if so how rich those lives are is, nonetheless, far from settled. Reconceptualizing what it means to be a moral being is a necessary precondition for entertaining the very idea that animals are capable of leading moral lives. In the end, whether they are and to what extent will rest on detailed empirical evidence drawn from the study of the activities of animals in their natural habitats.

References

- Allen, C., Bekoff, M., 1997, *Species of Mind: The Philosophy and Biology of Cognitive Ethology*, Cambridge MA, The MIT Press.
- Bekoff, M., Pierce, J., 2009, *Wild Justice: the moral lives of animals*, Chicago, The University of Chicago Press.
- Berridge, K. C., 2003, Comparing the emotional brain of humans and other animals. In: Davidson, R. J., Goldsmith, H. H., Scherer, K. (Eds.), *Handbook of Affective Sciences*, Oxford, Oxford University Press, 25-51.
- Brigandt, I., and Griffiths, P. (Eds.), 2007, *Biology and Philosophy*, 22, No. 5.
- Bradie, M., 2011, The Moral Lives of Animals. In: Beauchamp, T., Frey, R. G. (Eds.), *The Oxford Handbook of Animal Ethics*, Oxford, Oxford University Press.
- Darwin, C., 1981, *The Descent of Man and selection in relation to sex*, Princeton, Princeton University Press.
-- 1998, *The expression of the emotions in man and animals*, 3rd edn., Oxford, Oxford University Press.
-- 2000, *On the origin of species: A facsimile of the first edition*, Cambridge MA, Harvard University Press.
- de Waal, F., 2005, *Our Inner Ape*, New York, Riverhead Books.
-- 2006, *Primates and Philosophers: How Morality Evolved*, Princeton, Princeton University Press, The University Center for Human Values Series.
- Ereshefsky, M., 2007, Psychological Categories as Homologies: Lessons from Ethology. *Biology and Philosophy*, 22, 659-674.
- Griffiths, P., 1994, Cladistic Classification and Functional Explanation. *Philosophy of Science*, 61, 206-226.
-- 1996, The Historical Turn in the Study of Adaptation. *British Journal for the Philosophy of Science*, 47, 511-531.
-- 2006, Function, Homology and Character Individuation. *Philosophy of Science*, 73, 1-25.
-- 2007, Evo-Devo Meets the Mind: Toward a Developmental Evolutionary Psychology. In: Sansom, R., Brandon, R. N. (Eds.), *Integrating Evolution and Development: From Theory to Practice*, Cambridge MA, The MIT Press, 195-225.
- Lurz, R., 2011, *Mindreading Animals: The Debate over What Animals Know about Other Minds*, Cambridge MA, MIT Press.
- MacLean, P., 1990, *The triune brain in evolution*, New York, Plenum Press.
- Matthen, M., 2000, What is a Hand? What is a Mind? *Revue Internationale de Philosophie*, 4, 653-672.
-- 2002, Human Rationality and the Unique Origin Constraint. In: Ariew, A., Cummins, R., Perlman, M. (Eds.), *Functions: New Essays in the Philosophy of Psychology and Biology*, Oxford, Oxford University Press, 341-372.

Panksepp, J., 1998, *Affective neuroscience: the foundations of human and animal emotions*, New York, Oxford University Press, Series in affective science.

-- 2005, Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, 14, 30-80.

Panksepp, J., Northoff, G., 2009, The trans-species core SELF: The emergence of active cultural and neuro-ecological agents through self-related processing with subcortical-cortical midline networks. *Consciousness and Cognition*, 18, 193-215.

Romanes, G. J., 1882, *Animal Intelligence*, London, Kegan Paul, Trench, & Co.

-- 1885, *Mental Evolution in Animals*, London, Kegan Paul, Trench & Co.

Sayre-McCord, G., 2011, Rational Agency and the Nature of Normative Concepts, at:
<http://philosophy.unc.edu/people/faculty/geoffrey-sayre-mccord/on-line-papers/Rational%20Agency%20and%20the%20Nature%20of%20Normative%20Concepts.pdf/view>

Scientist or Seer?

Raymond Aron's Critique of Vulgar Marxism

in *Introduction à la philosophie de l'histoire**

Scott Nelson
(University of Vienna)
nelson.scott.b@gmx.at

It is therefore the task of history, once otherworldly truth has vanished, to establish the truth of this world. And after the sacred form of self-alienation has been exposed, it is above all the task of philosophy, which is handmaiden to history, to expose this self-alienation in its secular forms.¹ (Karl Marx)

The historian encounters the final questions which, posed by theology, are inevitable for the individual. Science is completed in philosophy, or rather the two of them are conjoined: the most positive scientist proceeds spontaneously to the organization of fragmentary regularities, without which history would tend to dissolve into an incoherent plurality and lose the intelligible unity that defines it.² (Raymond Aron)

Raymond Aron's criticisms of Marxism are well known and yet still subject to debate.³ Marx himself and his ideas are enjoying a new lease on life in the

* This text was prepared as part of an unpublished PhD dissertation.

¹ "Es ist also die Aufgabe der Geschichte, nachdem das Jenseits der Wahrheit verschwunden ist, die Wahrheit des Diesseits zu etablieren. Es ist zunächst die Aufgabe der Philosophie, die im Dienste der Geschichte steht, nachdem die Heiligungsgestalt der menschlichen Selbstentfremdung entlarvt ist, die Selbstentfremdung in ihren unheiligen Gestalten zu entlarven." Marx, 2013 [1844], Loc. 1287.

² "L'historien retrouve les questions dernières, issues de la théologie mais inévitables pour l'homme seul. La science s'achève en philosophie, ou plutôt l'une et l'autre se confondent : le savant le plus positif procède spontanément à l'organisation des régularités fragmentaires, sans laquelle l'histoire tendrait à se dissoudre en une pluralité incohérente et à perdre l'unité intelligible qui la définit." Aron, 1986 [1938], 324.

³ See, for example, Colen, 2013; Gordon, 2011; Judt, 1992; Judt, 1998.

fallout of the recent financial crisis.⁴ It behooves the scholar concerned with both of these phenomena to re-examine one of Marx's most attentive critics working at one of his more generous critiques of Marxism. This will provide us with the clarity indispensable for separating the wheat from the chaff in Marxism, as well as shed some light on the ongoing dispute regarding Aron's critique of Marxism.

It was in 1931 during his stay in Germany that Raymond Aron would begin his lifelong dialogue with Karl Marx, whose influence on the young French scholar's intellectual trajectory would be unmatched.⁵ He delved into this great German thinker's works "less in order to arrive at an opinion on the Soviet Union than to mark out the borderline between the analytics and dialectics (in the Kantian sense) of historical knowledge."⁶ He also wondered if a reading of *Das Kapital* might aid one in explaining the economic crisis.⁷ Sadly, Marx's analysis provided neither a sufficient explanation for the crisis nor much of a boost to Aron's vague socialism.

What it *did* offer, however, was a bold and comprehensive philosophy of History, and if that highly intrigued Aron, it was because, long before many of his countrymen, Aron had sensed that history was again "on the move". He was also ahead of almost all of his compatriots in his reading of Marx's earlier texts which, fortunately, were published in 1932 in Berlin during Raymond Aron's sojourn there.⁸ This wide range of material would likely have indicated to Aron very early on three crucial components of the totality of Marx's thought: his philosophical-anthropological assumptions, his socialist teachings, and his economic analysis, corresponding to German dialectics, French socialist thought, and the English analysis of political economy.⁹

As banal as this division may be,¹⁰ it is sufficient to make the equally banal statement that Marx's thought encompasses a number of different disciplines (philosophy, sociology/history, economics) and – considering the frequency with which Marx hops between the various disciplines in his works – we should try to take it as a whole. This division has as its corollary three motifs that dominate Marx's thought: the romantic, the Faustian-Promethean, and

⁴ See, for example, Eagleton, 2011; Hobsbawm, 2011; Sperber, 2013.

⁵ Aron, 2002, 304.

⁶ Aron, 1989, 147.

⁷ Aron, 2010 [1983], 85-86.

⁸ Colquhoun, 1986, 162.

⁹ Kolakowski, 2005 [1976], 9.

¹⁰ Aron, 1967, 172.

the rationalist, determinist Enlightenment.¹¹ Marx's construct claimed to explain not only the present, but also the past and predict a better future, lending it a mystical allure that was inversely proportional to its scientific plausibility; indeed, the very ambiguity of some of his terms (materialism, ideology, social classes, dialectic) accounts for both the difficulty and charm of interpreting Marx,¹² and Aron would have plenty of time in the future to combat the various apparitions that were Marx's progeny.¹³

Introduction à la philosophie de l'histoire

All that lay ahead of him, however, and the Marxist stranglehold on France would have to wait until after the war. In the meantime, he contented himself with critiquing Marx directly, and his early articles bear witness to some of Aron's fundamental insights not only about Marx, but about the world itself, such as the lack of a *primum movens*, the reciprocal relation between the economic and political spheres, and the primacy of politics.¹⁴ These observations indicated that our understanding of the world and causality was perhaps more complex than many cared to admit. Consequently, they suggested that there were certain limits to historical objectivity. If this seemingly modest conclusion strikes us today as self-evident, it is because we have so thoroughly assimilated the lesson that was shocking and extreme to Aron's academic superiors, steeped as they were in the firmly rooted French positivism of the day, namely that we constantly renew our perspective on history as we go on living in history.¹⁵ This would be the topic of Aron's primary doctoral dissertation, *Introduction à la philosophie de l'histoire: Essai sur les limites de l'objectivité historique*.

¹¹ Kolakowski, 2005 [1976], 335-341. Cf. Aron's three motifs of Marxism in Aron, 1954 [1951], 116-118.

¹² Aron, 2002, 543-545.

¹³ See, for example, Aron's *L'opium des intellectuels, Marximes imaginaires : D'une sainte famille à l'autre*, and *Histoire et dialectique de la violence*.

¹⁴ Aron, 1931, 43-47; Aron, 1932, 647-648; Aron, 1934, 509-511; Aron, 1937, 16-47. These sources have been taken from Colen, 2013. All involuntary traces or paraphrases of this work have the permission of the author.

¹⁵ Judt, 1998, 140.

Written between October/November 1935 and Easter 1937,¹⁶ the *Introduction* was defended at the Sorbonne on 26 March 1938, two weeks after the *Anschluss*.¹⁷ While the work cannot be said to contain a formal refutation of Marxism – Aron is willing to concede as much himself¹⁸ – it raises enough questions about historical objectivity as to constitute a powerful criticism of the Marxist philosophy of History. A word must first be said about Aron's inability to refute formally Marxism in his *Introduction*. Once it has been admitted that Marxism is more of a philosophy than a science then it becomes obvious that Raymond Aron's analysis of causality has little to do with Marxism's foundation, which is a philosophy of man, and not the primacy of a particular cause.¹⁹ It is this observation that permits Aron later on to state that “authentic Marxism”, i.e. that which deems itself a philosophy, would be compelled to recognize causal complexity and the plurality of interpretations.²⁰ The problem begins when Marxism is interpreted as a science, an error for which its founder is not entirely blameless.²¹ The moment it comports itself as a science is the moment that it becomes accountable to the rules of causality, and it is this vulgar Marxism that Aron's *Introduction* has the capacity to critique.²² This is most readily apparent in the third part of the third section, on causal thought, when the *Introduction* discusses historical laws, causal systematization, and historical determinism.

Aron's first task is to establish the difference between a law and a cause, which are united in Comtean positivism. The distinction between the two

¹⁶ Aron, 2010 [1983], 163.

¹⁷ Aron, 1986 [1938], i.

¹⁸ Aron, 1986 [1938], 312.

¹⁹ Aron, 1986 [1938], 312. See also the affiliated endnote on pp. 495-496.

²⁰ Aron, 1986 [1938], 387-388.

²¹ See Aron, 2006 [1937], 133. Aron justifies his distinction between authentic and vulgar Marxism on the basis of the latter's proclivity to seek after the prestige of a positive science. He claims that this was an error Marx himself did not make, provided that one is willing to take seriously the texts of his youth in which he outlined his definitive philosophy. This argument is unconvincing as long as we count *The German Ideology* as one of Marx's youthful texts (and Aron himself seems to do so—see Aron, 1967, 145) wherein he makes it quite clear that the premises on which is based the materialist conception of history are real and “can thus be verified in a purely empirical way.” See Marx and Engels, 1998 [1932], 36.

²² Given that I will endeavour to explore primarily Aron's critique of vulgar Marxism, and the *Introduction* does not contain a detailed critique of Marx's economics, Marx's economics-heavy works, namely *Das Kapital*, have been excluded from this analysis. Furthermore, “Marxism” will henceforth refer to vulgar Marxism and its philosophy of History, while the name “Marx” will be reserved for the aforementioned “authentic Marxism”.

terms depends on the inevitability of a particular effect resulting from a particular cause, or a particular cause always existing prior to a particular effect. If B results from A as certainly as night follows day, then we are in the presence of a law. If generalization is impossible, then we have likely come across but a cause for a unique effect located within an historical whole where many unique factors are at work. One speaks, for example, of the causes, and not the laws, of suicide, for we recognize that suicide is an act located at the convergence of many particular factors in an historical whole. Similarly, there are not laws, but effects of devaluation which are contingent on the circumstances in which devaluation occurs.²³ Nonetheless, these examples aside, are there laws in history? Raymond Aron concludes negatively: it is not possible to discern laws in the historical totality. This is not to say that there are no partial laws, for example in linguistics; but the assertion that there are laws that apply to the historical totality, such as the inevitability of cycles, are as incontestable as they are meaningless if that's the extent of their explanatory power.²⁴ The best we can hope for then is a fragmentary determinism.

Causal Systematization

And yet Marxism pretended to be way beyond this point as it had already isolated the cause of primary importance — the economic order — and, by virtue of a sweeping philosophy of History, had demonstrated both how history could be explained in terms of this cause and how the contradictions inherent to the present economic order were doomed to inch closer to that extreme point at which the current bourgeois society would collapse under the weight of its own injustice, bringing about an order free of alienation and oppression and thus essentially different from all previous societies.

Such an exhaustive causal systematization begs three questions: does the primacy of a single cause even exist and, if so, can one discover it? Can one identify *all* of the causes of any given occurrence? Can one discern constant relations amongst typical causes? In brief, to what extent is causal thought amenable to systematization?²⁵ The second and third questions can be

²³ Aron, 1986 [1938], 294.

²⁴ Aron, 1986 [1938], 300-301.

²⁵ Aron, 1986 [1938], 307.

consigned to irrelevance in the Marxist schema because of the affirmation of the first question. What, then, constitutes the primacy of the economic order? Marx states clearly that

[I]n the social production of their existence, men inevitably enter into definite relations, which are independent of their will, namely relations of production appropriate to a given stage in the development of their material forces of production. The totality of these relations of production constitutes the economic structure of society, the real foundation, on which arises a legal and political superstructure and to which correspond definite forms of social consciousness. The mode of production of material life conditions the general process of social, political and intellectual life. It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness. At a certain stage of development, the material productive forces of society come into conflict with the existing relations of production or – this merely expresses the same thing in legal terms – with the property relations within the framework of which they have operated hitherto. From forms of development of the productive forces these relations turn into their fetters. Then begins an era of social revolution. The changes in the economic foundation lead sooner or later to the transformation of the whole immense superstructure.²⁶

To begin with, how are we to understand the forces and relations of production? Do they include political and legal institutions or are they merely technology and the economic order? If political and legal institutions are incorporated in the terms – which is perfectly acceptable, for everyone can initially define his terms however he likes and then is obliged to be consistent – then it can be said that the cause is inherent to the system, or that the system's very own contradictions are the cause.²⁷ Yet, if the forces and relations of production *include* political and legal institutions (and our consciousness of material forces), then one can do no better than arrive at the vapid conclusion that the forces and relations of production effect themselves. And then how is the superstructure distinct from this all-inclusive infrastructure?

But Karl Marx wouldn't deign to leave us with such a trite and uninspired finding. Besides, he makes eminently clear in the aforementioned quote that the interaction between the forces and relations of production, which constitutes the economic structure of society (infrastructure), are the foundation of the superstructure (political institutions, law, consciousness, etc.). With these two entities' separateness established, it remains to be seen

²⁶ Marx, 1977 [1859].

²⁷ Aron, 1986 [1938], 308.

how the infrastructure determines the superstructure. Raymond Aron outlines two different ways of going about this: the sociological and the historical methods.²⁸

The former approach would require one to demonstrate that a particular economic situation has a particular political regime, ideology, etc. as its result; similarly, one should be able to determine, on the basis of a particular political regime, ideology, etc., the particular economic situation that is its cause. But it's undeniable that many capitalist regimes have different political systems or constitutions which are also sometimes transformed without having been stimulated by the economic system. Can the multiple political regime changes in France since 1789 be explained by economics alone?²⁹

The latter approach involves tracing historical events back to economic antecedents. The trouble here is that there is no scientific reason why one should arrest his investigation once the desired economic cause has been found. Furthermore, causal regression is bound to discover an economic antecedent at some point. To suggest, then, that it is the *first* and/or *only* cause says more about the analyst's curiosity or scientific disingenuousness than about the phenomenon in question. Karl Marx himself was too intelligent to fall consistently into the same trap that some of his disciples did. The closer he got to the complexities of *histoire-se-faisant*, the more clearly the contradictions shone between the in-depth analysis required by his journalistic integrity and the sweeping generalizations demanded by his theory. *The Eighteenth Brumaire of Louis Bonaparte* is a testament to the difficult balancing act he tried to perform. On the one hand, the work is loaded with allusions to continual class conflict and the folly inherent in any effort on the part of the socialists to achieve parliamentary compromise – what Marx derisively referred to as "parliamentary cretinism".³⁰ On the other hand, he's compelled to admit that the Legitimists and Orleanists – whose conflict, Marx maintains, is provoked by the rivalry between landed property and industrial property, respectively – are able to come to terms with each other in a parliamentary republic, hence granting that a change of political regime can effect a reconciliation. But if the struggle between the two classes boils down

²⁸ Aron, 1986 [1938], 308.

²⁹ Aron, 2006 [1937], 131.

³⁰ Marx, 2012 [1852], Loc. 1433.

to socio-economic factors alone, then any sort of compromise should be impossible regardless of the political regime.³¹

What this example illustrates is that while one can *describe* a society well enough on the basis of the relations of production, this is insufficient to *explain* a society. "If the economy obeyed a purely autonomous law then prediction and explanation would be equally possible",³² in much the same way that the physicist points indifferently to the formula $F = ma$ both when asked with what force an object of given mass hit the ground, as well as when asked with what force an object of a different mass *will* hit the ground. It is therefore true to say that the economic and political spheres interact, but a more precise analysis of *how* exactly is beyond the scope of Aron's dissertation.³³

The relation between the infrastructure and the superstructure is manifested on the historical stage as the class struggle. Even if one were to be swept up into a state of euphoria by those declarative and audacious words – "The history of all hitherto existing society is the history of class struggles"³⁴ – one would still be tempted, in a moment of clarity, to inquire into the content behind them. How, for example, does one designate a class? There's a brief enumeration to be found in *The Class Struggles in France, 1848-1850*,³⁵ and a more comprehensive outline (dealing with the lack of a peasant class, properly speaking) in *The Eighteenth Brumaire*.³⁶ It goes without saying that a class is composed of many families living in similar conditions with similar interests; however, the peasants do *not* constitute a class because they lack class consciousness.³⁷ According to Karl Marx, they lack the consciousness requisite for a class because "each individual peasant family is practically self-sufficient, directly producing the majority of their own

³¹ Aron, 1967, 290-292.

³² "Si l'économie obéissait à une loi purement autonome, prévision et explication seraient également possibles." Aron, 1986 [1938], 309.

³³ Aron's Sorbonne trilogy on industrial society (*Dix-huit leçons sur la société industrielle*, *La lutte de classes*, *Démocratie et totalitarisme*) explores the complex interrelations between these different spheres.

³⁴ Marx and Engels, 2012 [1848], 73.

³⁵ Marx, 2012 [1895], 27.

³⁶ Marx, 2012 [1852], Loc. 2008-2018.

³⁷ Aron, 2002, 531.

consumption and thereby sustaining themselves more in interacting with nature than society.”³⁸

But can a class become self-conscious only through economic intercourse (*Verkehr*) in society? If one answers negatively, then it stands to reason that there are other, potentially non-materialistic, ways of creating a class, and this would undermine the Marxist assertion that the infrastructure determines the superstructure and, thereby, the course of history. However, if one answers affirmatively, then the relation between infrastructure and superstructure is left intact, but one is then obliged to jump to the level of history and ask whether it is indeed the case that class struggles characterize the history of all hitherto existing societies. Is the class the only sort of social grouping into which individuals can be organized? Are there other groupings or forms of collective consciousness that have the potential to override the purely materialistic factor connecting people, such as national prejudices or religion? Marxism’s high-powered focus on only class struggle suffers, on the level of historical explanation, from the same defect that plagues its conception of the relation between the infrastructure and the superstructure on the level of sociological explanation: reality is too complex to permit only a single cause or perspective.

This innate complexity also hinders our ability to list every cause for any given phenomenon, not to mention guarantee constant relations amongst the typical causes (the reader will recall that these two additional conditions – deemed irrelevant in the Marxist schema – are also necessary to foresee the future, in all of its specificity, beyond a reasonable doubt). On the one hand, one could always have recourse to “faith” in order to escape the narrow confines of causal reason’s cage, but only at the cost of demoting one’s science to philosophy. In this way at least the teleology would be preserved. On the other hand, once Providence has been sacrificed on the altar of science there is no longer any reason to believe that history has an end.³⁹ As for the existence of a *primum movens* in the world, there is no reason, scientific or based on the rules of causality, i.e. analysis and comparison, to believe that the historical totality has a first cause, much less an economic

³⁸ “Jede einzelne Bauernfamilie genügt beinahe sich selbst, produziert unmittelbar selbst den größten Teil ihres Konsums und gewinnt so ihr Lebensmaterial mehr im Austausche mit der Natur als im Verkehr mit der Gesellschaft.” Marx, 2012 [1852], Loc. 2008.

³⁹ Aron, 1986 [1938], 322.

one at that.⁴⁰ We, and not History, then, are responsible for our actions and our own destiny.⁴¹

Partial Determinism

This realization does not leave us to wade through a swamp of uncertainty. There are still partial laws and fragmentary determinism, punctuated by chance and individual acts, which could be called probabilism, and it walks the fine line between absolute regularities and total incoherence. As Raymond Aron noted:

(...) once we realize there are no necessary relations, we're brought back to the antithesis of the part and the whole, because if a causal connection can coincide with an observed succession only by losing all of its generality, it's because the constellations in which a regularity manifests itself are singular, and each constellation belongs to an historical totality which is both unique and unified. Therefore, probability would be the result, in this instance, of the contradiction between the *necessity of classification* and the *impossibility of isolation*.⁴²

Room is thereby made for some of Marx's predictions *qua* predictions, and not *qua* prophecies, to be proved correct, provided they are predictions which recognize the fundamental impenetrability of the logic of History and instead readjust themselves to account for only a part of reality. Would Marxism be proved or disproved if capitalism, burdened by its own internal contradictions, finally did come to a standstill, but the proletariat opted for a political regime other than communism?⁴³ Is the increasing dissolution of the family today a victory for Marxism?⁴⁴ Such questions cannot expect a scientific answer when Marxism itself has been subject to so many different interpretations and poses at one moment as a science and, at the next, as a religion.

⁴⁰ Aron, 1986 [1938], 316.

⁴¹ Cf. with views in Popper, 2011 [1945], 474-483.

⁴² "D'une autre manière, l'irréalité des relations nécessaires nous ramène à l'antithèse de la partie et du tout, car si le rapport de causalité ne coïncide avec la succession observée qu'en perdant toute généralité, c'est que les constellations dans lesquelles se manifeste une régularité sont singulières et que chaque constellation appartient à une totalité historique, à la fois unique et relativement unifiée. La probabilité résulterait donc, ici, de la contradiction entre la *nécessité du découpage* et *l'impossibilité de l'isolement*." Aron, 1986 [1938], 328.

⁴³ Aron, 1986 [1938], 309-310.

⁴⁴ Hobsbawm, 2011 [1998], 112.

Long-term extrapolations meant to paint a detailed picture of the future in broad strokes are the result of mistaking a simple tendency for a law.⁴⁵ And when the prophecy comes to naught, the “high priest” need only declare that this historical event was, in fact, not the moment ordained in Scripture and that we must continue to wait. All doomsayers proclaiming the end of civilization or the world operate on the same pusillanimous principle: they can’t be proved wrong. It is for this reason that Karl Popper – with whom Aron shared an elective affinity⁴⁶ – scorned the principle of verifiability, which could impart a respectable scientific veneer to any trash theory, and replaced it with the principle of falsifiability, which would display not only the modesty of the researcher as an individual, but also his noble willingness to test his own ideas at every turn.⁴⁷

Conclusion – Marxism as Existential Choice

By this point it should be quite obvious that the Marxist obsession with the relations between the infrastructure and the superstructure, forces of production and relations of production, class conflict, etc. is the result of a preference that precedes science and causal analysis. Whether the question is on what level one should analyze a phenomenon⁴⁸ or whether there are universal laws in history,⁴⁹ the curiosity and prejudices of the researcher are reflected at all times in his work.

In the case of Karl Marx there is a clear predilection for materialism over idealism.⁵⁰ While scientifically illustrating which of the two precedes the other is bound to be an exercise in futility, it’s patent enough that, early on, Marx was trying to distance himself from the imperious influence of Hegel and the

⁴⁵ For a discussion of this error see Popper, 2002 [1957], 116-119.

⁴⁶ Aron, 2010 [1983], 539, 936.

⁴⁷ “Here was an attitude utterly different from the dogmatic attitude of Marx, Freud, Adler, and even more so that of their followers. Einstein was looking for crucial experiments whose agreement with his predictions would by no means establish his theory; while a disagreement, as he was the first to stress, would show his theory to be untenable.” Popper, 2002 [1992], 38.

⁴⁸ Aron, 1986 [1938], 285.

⁴⁹ Aron, 1986 [1938], 306.

⁵⁰ It is worth noting that, in Aron’s view at least, Marx’s interpretation of history in 1848 did not strictly imply materialism. See Aron, 2002, 56.

Young Hegelians' idealism.⁵¹ In his mind, after Hegel, no one had made any significant theoretical contributions to German philosophy other than Feuerbach.⁵² And with Feuerbach one gets the impression that, for all of his materialism, he had still failed to overcome Hegel in at least one respect that was absolutely essential for Karl Marx: action.

The chief defect of all previous materialism (that of Feuerbach included) is that thing [*Gegenstand*], reality, sensuousness are conceived only in the form of the object, or of contemplation, but not as *sensuous human activity, practice*, not subjectively.⁵³

Marx's call to action and his desire to subject consciousness to material reality, not to mention a grave concern for the state of Germany, are already to be found in some of his earliest letters and works.⁵⁴

He began with a critique of religion, that opium of the people,⁵⁵ and progressed by way of a critique of politics and law to a critique of the economy. This last critique was the most important because it aimed at revealing the exploitative nature of the present socio-economic order, and so Marx's theoretical contributions post-1848 were very much concerned with analyzing the economy.⁵⁶ His fundamental contention in *Das Kapital* was that labourers were clearly being paid less than the value of what they were producing – this was the peculiarity of labour as a commodity.⁵⁷ The products themselves were crystallized labour and so their value was equal to the amount of labour required to produce them (theory of labour value), while the labourer's wages were equal to the amount required to sustain the labourer and his family (theory of wage value). Given that labourers were not being remunerated at the price of their goods, they must have been working partly

⁵¹ See his polemical works *The Holy Family* and *The German Ideology*. This does not mean that he effected a total break from Hegel's influence on all counts. After all, desperately trying to escape from under the wing of an individual's influence is also an indication of that very influence and, in a sense, an homage to the predecessor. Marx's notion of alienation derives from Hegel (see Aron, 1967, 176) and the absolute validity that Marxism is meant to represent is a literal interpretation of Hegelianism (see Aron, 1986 [1938], 393).

⁵² Marx, 1988 [1932], Loc. 89.

⁵³ Marx, 1998 [1938], Loc. 12550.

⁵⁴ See, for example, Marx, 1843a; Marx, 1843b; Marx, 2013 [1844].

⁵⁵ Marx, 2013 [1844], Loc. 1279.

⁵⁶ The key works are the *Grundrisse*, *A Contribution to the Critique of Political Economy*, and *Das Kapital*.

⁵⁷ Kolakowski, 2005 [1976], 211. The following summary of some of the salient concepts of *Das Kapital* draws heavily on Aron, 1967, 160-164.

for themselves and partly for someone else, namely the employer (theory of surplus value), who enjoyed the privileged position of being able to extract this additional effort from the workers because he owned the means of production. This critique allowed Marx to explain profit – the very essence of capitalism – and it also laid bare the perverted core of the capitalist system.

These critiques were necessary in order to penetrate the many illusions of our existence and get at reality. Philosophy itself was one of these illusions insofar as it remained contemplative – not rooted in practice – and thereby alienated man,⁵⁸ with this alienation permeating man's material (economic) existence as well. Bourgeois society was predicated on the market, composed of proprietors and wage-labourers whose only human connection to each other was money. In the older guild system, the more limited intercourse (*Verkehr*) between towns, coupled with the less advanced division of labour, meant that craftsmen could really involve themselves in their work which could assume a more artistic and personal human value.⁵⁹ Bourgeois society diminished this level of human involvement in their products and rendered social interaction and human life in general as nothing more than making a living, as opposed to allowing individuals to realize all of their aptitudes.⁶⁰ And because humans are defined by what and how they produce, individuals were, in effect, becoming dehumanized.⁶¹

The group most adversely affected by this phenomenon was, according to Marx, the proletariat. The proletariat is given a momentous role for the first time in *Introduction to a Critique of Hegel's Philosophy of Right*. "It is here that he expresses for the first time the idea of a specific historical mission of the proletariat, and the interpretation of revolution not as a violation of history but as a fulfillment of its innate tendency."⁶² Marx, ever the opponent of injustice,

⁵⁸ Lichtheim, 1964 [1961], 44.

⁵⁹ Marx and Engels, 1998 [1932], 74.

⁶⁰ Aron, 1967, 178.

⁶¹ Marx and Engels, 1998 [1932], 37.

⁶² Kolakowski, 2005 [1976], 105. See Marx, 2013 [1844], Loc. 1538: "A thorough Germany cannot make a revolution without making a revolution thoroughly. The emancipation of the German is the emancipation of man. The head of this emancipation is philosophy; its heart is the proletariat. Philosophy cannot be realized without the transcendence of the proletariat; the proletariat cannot transcend itself without the realization of philosophy."

"Das gründliche Deutschland kann nicht revolutionieren, ohne von Grund aus zu revolutionieren. Die Emancipation des Deutschen ist die Emancipation des Menschen. Der Kopf dieser Emancipation ist die Philosophie, ihr Herz das Proletariat. Die

must also have been appalled when, in 1850, he arrived in London to see first-hand the working conditions of the proletariat: workdays over 16 hours long, every day of the week, with men, women, and children as young as ten working and sleeping in the factory for a wage that could hardly sustain them – life was grueling work and nothing more.⁶³ Bourgeois society had simply prolonged the oppression of one class by another, and whether the disadvantaged class was called slaves, serfs, or wage-labourers did little to conceal their servility. What was perhaps most despicable, however, was that this order of things was supposed to be progressive and liberal.

Karl Marx surely thought that the capitalism of his day was an outrage, a crime against humanity that could be answered only by revolutionary action. Marxism is rather an existential attitude⁶⁴ where choice – the refusal of capitalism and the wish to destroy it – precedes theory.⁶⁵ Therefore Marx spent his life trying to dispel the many insidious illusions that had developed and were obscuring the true nature of reality. In so doing he created an enormously impressive and infectious *Weltanschauung*. A century later Raymond Aron – no less a defender of truth and justice than Marx – would spend his life fighting the illusions that Marx's own system had produced.

Philosophie kann sich nicht verwirklichen ohne die Aufhebung des Proletariats, das Proletariat kann sich nicht aufheben ohne die Verwirklichung der Philosophie.”

⁶³ Morrison, 2006 [1995], Loc. 1296-1304.

⁶⁴ Not to mention a detailed analysis of capitalism.

⁶⁵ Aron, 1986 [1938], 389.

References

- Aron, R., 1931, De Man, Au-delà du marxisme. *Libres Propos*, Jan, 43-47 (reed. and expanded in Aron, R., 1975, Au-delà du marxisme. *Contrepoint*, 16, 161-169).
-- 1932, F. Fried, La Fin du capitalisme. *Europe*, Dec, 647-648.
-- 1934, Comptes rendus de: A. Cornu, La jeunesse de Karl Marx, de l'hégélianisme au matérialisme historique (1818-1845); Moses Hess et la gauche hégélienne. *Recherches Philosophiques*, IV, 509-511.
-- 1937, Les Rapports de la politique et de l'économie dans la doctrine marxiste. In: Aron, R. et al., *Inventaires, II, L'économique et le politique*, Paris, Félix Alcan, 16-47 (Reed. Aron, R., 2006, Politique et économie dans la doctrine marxiste. In: Aron, R., *Les sociétés modernes*, Paris, Presses Universitaires de France, 121-138).
-- 1938, *Introduction à la philosophie de l'histoire: Essai sur les limites de l'objectivité historique*, Paris, Gallimard (Reed. 1986).
-- 1951, *Les Guerres en chaîne*, Paris, Gallimard (Eng. Trans. And modified as *The Century of Total War*, by E. Dickes and O. Griffiths, New York, Doubleday, 1954).
-- 1955, *L'opium des intellectuels*, Paris, Calmann-Lévy.
-- 1962, *Dix-huit leçons sur la société industrielle*, Paris, Gallimard.
-- 1964, *La lutte de classes. Nouvelles leçons sur les sociétés industrielles*, Paris, Gallimard.
-- 1965, *Démocratie et totalitarisme*, Paris, Gallimard.
-- 1967, *Les étapes de la pensée sociologique*, Paris, Gallimard.
-- 1969, *D'une sainte famille à l'autre. Essais sur les marxismes imaginaires*, Paris, Gallimard.
-- 1973, *Histoire et dialectique de la violence*, Paris, Gallimard.
-- 1979, De l'existence historique. *Cahiers de philosophie politique et juridique de l'Université de Caen*, 15, (1989), 147-162.
-- 1983, *Mémoires. 50 ans de réflexion politique*, Paris, Julliard (Reed. Aron, R., 2010, *Mémoires : Édition intégrale inédite*, Paris, Robert Laffont).
-- 2002, *Le Marxisme de Marx*, Paris, Éditions de Fallois.
- Colen, J., 2013, *Guide to the Introduction to the Philosophy of History*, Lisbon, Aster.
- Colquhoun, R., 1986, *Raymond Aron. Volume I: The Philosopher in History, 1905-1955*, London, SAGE Publications.
- Eagleton, T., 2011, *Why Marx Was Right*, London, Yale University Press.
- Gordon, D., 2011, In search of limits: Raymond Aron on 'secular religion' and communism. *Journal of Classical Sociology*, 11 (2), 139-154.
- Hobsbawm, E., 1998, On the *Communist Manifesto*. In: Hobsbawm, E., 2011, *How to Change the World. Marx and Marxism 1840-2011*, London, Little, Brown.
- Judt, T., 1992, *Past Imperfect: French Intellectuals, 1944-1956*, Oxford, University of California Press.
-- 1998, *The Burden of Responsibility. Blum, Camus, Aron, and the French Twentieth Century*, Chicago, University of Chicago Press.
- Kołakowski, L., 1976, *Glówne Nurty Marksizmu*, Paris, Instytut Literacki (Eng. Trans. *Main Currents of Marxism*, by P.S. Falla, New York, W.W. Norton & Company, 2005).

Lichtheim, G., 1961, *Marxism*, London, Routledge and Kegan Paul (Reed. 1964).

Marx, K., 1843a, Letter from Marx to Ruge, Cologne. *Deutsch-Französische Jahrbücher*, May. Accessed 5 November 2013:

http://www.marxists.org/archive/marx/works/1843/letters/43_05.htm.

-- 1843b, Letter from Marx to Ruge, Kreuznach. *Deutsch-Französische Jahrbücher*, September. Accessed 28 October 2013:

http://www.marxists.org/archive/marx/works/1843/letters/43_09.htm.

-- 1844, *Zur Kritik der Hegel'schen Rechtsphilosophie*. In: *Werke von Karl Marx*, Die Digitale Bibliothek, 2013, Kindle.

-- 1852, *Der achtzehnte Brumaire des Louis Bonaparte*, Jazzybee Verlag, 2012, Kindle.

-- 1857, *Grundrisse der Kritik der politischen Ökonomie*, Berlin, Dietz, 1974.

-- 1859, *A Contribution to the Critique of Political Economy*, Preface, Eng. Trans. S. W. Ryazanskaya, London, Progress Publishers, Moscow, 1977. Accessed 31 October 2013:

<http://www.marxists.org/archive/marx/works/1859/critique-pol-economy/preface.htm>.

-- 1872, *Das Kapital: Kritik der politischen Ökonomie*, Hamburg, Nikol Verlag, 2011.

-- 1895, *The Class Struggles in France, 1848-1850*, Aristeus Books, 2012, Kindle.

-- 1932, *Economic and Philosophic Manuscripts of 1844*, Eng. Trans. Martin Milligan, Amherst, Prometheus Books, 1988, Kindle.

-- 1938, *Theses on Feuerbach*. In: *The German Ideology*, Amherst, Prometheus Books, 1998, Kindle.

Marx, K. and Engels, F., 1845, *Die heilige Familie oder Kritik der kritischen Kritik*, Hamburg, Tredition Classics, 2011, Kindle.

-- 1848, *The Communist Manifesto*, New Haven, Yale University Press, 2012, Kindle.

-- 1932, *The German Ideology*. In: *The German Ideology*, Amherst, Prometheus Books, 1998, Kindle.

Morrison, K., 1995, *Marx, Durkheim, Weber: Formations of Modern Social Thought*, London, SAGE Publications (Reed. 2006), Kindle.

Popper, K., 1945, *The Open Society and Its Enemies*, New York, Routledge and Kegan Paul (Reed. 1952; 1957; 1962; 1966; Routledge Classics, 2011), Kindle.

-- 1957, *The Poverty of Historicism*, New York, Routledge and Kegan Paul (Reed. Routledge Classics, 2002).

-- 1992, *Unended Quest. An Intellectual Autobiography*, New York, Routledge (Reed. 2002), Kindle.

Sperber, J., 2013, *Karl Marx: A Nineteenth-Century Life*, New York, Liveright.

Logic and Computers. A Sketch of Their Symbiotic Relationships

Jordi Vallverdú

(Philosophy Department, Universitat Autònoma de Barcelona)

jordi.vallverdu@uab.cat

0. Introduction

It is a well established fact that Logic and Computer Sciences have a long and close relationship, which has benefited both sides of this symbiotic relationship. Therefore, this is a case of mutualist symbiosis, a relationship in which both fields benefit. Nevertheless, and according to the main studies about this topic, this relationship could looks like as a parasitist or commensalist one: that is, one field benefits but the other is harmed (in the first case), or one field benefits, while the other remains unaffected (in the second one).

Our departure hypothesis is: not only logic has influenced computer sciences development (a common topic on the historiographical approaches to computer sciences)¹, but computer sciences have also changed the logical thinking and practice. This is a mutualist symbiosis. This second idea, the impact of computer sciences into logical field, has been overlooked by most of the researchers who have studied the relationships between both fields. This symmetric symbiosis reaches from practical to very fundamental aspects of the relationship.

To be honest, we are not just talking about logic and computer sciences fields, but also of engineering and mathematics. From a conceptual and a practical point of view, there are crucial interconnections between all these

¹ Davis 1988, 2000.

disciplines. Logic is a parent discipline of computer science², as well as a basic constituent of computer science curricula. At the same time, logic has provided computer sciences with several tools, methods and theories. But more than a similarity there is also an equivalence relation between programs and logical formula³: a program is a sequence of symbols constructed according to formal syntactic rules and it has a meaning which is assigned by an interpretation of the elements of the language. These symbols are called *statements* or *commands* and the intended interpretation is an execution on a machine rather than an evaluation of a truth value. Its syntax is specified using formal systems but its semantics is usually informally specified. Then, a statement in a programming language can be considered a function that transforms the state of a computation. Going further⁴, we can affirm that there is no line between software and mathematics. Translation problems between them are trivial, because they are equivalent. Therefore, by the Church-Turing Thesis we can establish an identity between *Programs, Logics and Mathematics*. There is a common nature between these domains, with different historical backgrounds and main objectives, but at the end they speak the same language. Ben-Ari affirms that

a program is not very different from a logical formula. It is a sequence of symbols which can be parsed according to formal syntactical rules and it expresses a meaning, in this case a computation, according to the intended interpretation of the atomic elements and connectives of the language. In programming, these elements are called *statements* or *commands* because the intended interpretation is machine execution rather than just computation of a truth value⁵.

According to the *Handbook of Logic in Computer Science*⁶, there are several fields in which logic is present into computer science:

programming language *semantics*: to make sure that different implementations of a programming language yield the same results, programming languages need to have a formal semantics. Logic provides the tool to develop such semantics.

- type theory, linear logic, categorical theories
- Lambda calculus, Pi calculus

² Thomas, 2000.

³ Ben-Ari, 1993.

⁴ As Klemens, 2006, states.

⁵ Ben-Ari, 1993, 244.

⁶ Abramsky, Gabbay & Maibaum, 1993.

- specification logics (e.g. dynamic logic, Hoare logic, temporal logic...)
- Finite model theory, data base theory
- Term rewriting, unification, logic programming, functional programming
- Automated theorem proving

Program validation and verification (V&V): it uses temporal logic (finite state systems can naturally be represented as models for temporal logic; verifying if a system satisfies some desired property thus amounts to checking that the formula representing the specification is satisfied by the model representing the system).

- Process calculi and concurrency theory
- Modal logic, logics of knowledge.

Logic plays also an important role in very different areas of Computer Science, like computer architecture (logic gates), software engineering (specification and verification), programming languages (semantics, logic programming)⁷, databases (relational algebra and SQL), artificial intelligence (automatic theorem proving), algorithms (complexity and expressiveness), and theory of computation (general notions of computability).

This is the reason for which specialists of these disciplines have worked together or have established working relationships. This is the reason of the present paper: to show the relationships between computer and logic research domains and how they have influenced each other.

1. From Logic to Computers

1.1. A little bit of history

In the second half of the nineteenth century a new branch of logic took shape: mathematical logic. Its aim was to link logic with the ideas of arithmetic and algebra, in order to make logic accessible to the algebraic techniques of formula manipulation⁸. Deductive reasoning could be then reduced to algebraic formulisms. The author who made this possible was George Boole (1815-1864). By reducing Logic to Algebra, Boole made possible to introduce

⁷ At the same time, logic programming provides attractive models for inferential information flow (Bentham & Martínez, 2007: 54).

⁸ Thomas, 2000.

mathematics into logical thinking through the Boolean Algebra. At a certain point, Boole realized Leibniz's conception of an *algebra of thought*. With only numbers 0 and 1, Boole built up an entire algebra, turning logic into a fully symbolic practice which could be computed easily. So easy that its implementation into a computational binary framework made Boolean circuits possible. This late idea came from Claude Shannon, whose 1938 master's thesis in electrical engineering showed how to apply Boole's algebra of logic into electronic switching circuits. Although some of the first computers were not working on a binary approach, like ENIAC or Harvard's Marc I (still decimal), very soon all computers used Boolean algebra implemented into Boolean circuits.

Between 1890 and 1905, Ernst Schröder wrote his *Vorlesungen über die Algebra der Logik*, where developed the ideas of Boole and De Morgan, including at the same time ideas of C.S. Peirce showing that the algebraization of logic was possible and with a great impact for the new logic studies.

By independent ways, the mathematic Charles Babbage (1791-1871) and his co-researcher Augusta Ada Byron, Countess of Lovelace (1816-52) and a mastered young mathematician too, developed the Difference Engine, first, and Analytical Engine, second. To be precise, Ada only collaborated in some aspects of the second machine: she expanded considerable effort on developing the first programs for the planned machine and on documenting its design and logic. In fact, Ada developed a method for calculating a sequence of Bernoulli numbers with the Analytical Engine, which would have run correctly whether the Analytical Engine had ever been built. Both machines were far beyond the capabilities of the technology available at the time. Nevertheless, as a theoretical concept, the idea of the Analytical Engine and its logical design are of enormous significance. This is the first realization that, by mechanical means, it might be possible to *program* complicated algorithms. The Analytic Engine had, in principle, all of the important components (Memory, Processor and Input/Output protocol) that are present in modern-day computer systems. For this reason Babbage has a strong claim to be the inventor (even if not the first builder) of the modern computer. Designed for military purposes (calculating firing maritime ballistic tables), the Difference Engine was intended to evaluate polynomial functions, using a mathematical technique called the *Method of Differences*, on which Babbage had carried out important work. With later digital computers was possible to

apply propositional calculus to the two voltage levels, arbitrarily assigning the symbols 0 and 1 to these voltage levels.

With his *Begriffschrift* (1879), Gottlob Frege pioneered first attempts to put into symbols natural languages, which were used to make science. He tried to achieve the old Leibniz's dream: calculus as rational thinking (with the *characteristica universalis*), trying to develop a syntactic proof calculus that could be as good as mathematical proofs. Unfortunately, the system Frege eventually developed was shown to be inconsistent. It entails the existence of a concept R which holds of all and only those extensions that do not contain themselves. A contradiction known as "Russell's Paradox" follows.

Very often, Zermelo and Russell discovered the set theoretic paradoxes, which demolished the main logic proof building. Russell and Whitehead published between 1910 and 1913 their *Principia Mathematica*, in which they re-established the foundations of pure mathematics in logical terms⁹... something not so useful for practical purposes if we consider the fact that both authors required 379 pages to justify the truth of '1+1=2' (in the Volume I, §54.43 and completed in Volume II, §110.643).

The same bad news for idealistic approaches to mathematical entities were found to Hilbert's Program about the foundations of mathematics after Gödel's ideas on incompleteness. According to him, any logical system powerful enough to include natural numbers was also necessarily incomplete. Beyond this idea, Gödel also introduced the fundamental technique of arithmetization of syntax ('Gödel-numbering'), which led to the transformation of kinds of data useful for the purposes of computation¹⁰. At the same time, Gödel introduced the first rigorous characterization of computability, in his definition of the class of the primitive recursive functions.

Alfred Tarski created in 1929 the notion of independent semantics, a very important idea for the future of logics. The syntactic concepts comprising the notation for predicates and the rules governing inference forms needed to be supplemented by appropriate semantic ideas and Tarski was the guy who created the fundamentals of the semantics of First Order Logic (Robinson, 2000). With these conceptual tools, Tarski provided a rigorous mathematical concept of an interpretation.

With a deep insight for the work of the future, David Hilbert proposed stimulant problems which required from enthusiastic mathematicians

⁹ Flach, 2005.

¹⁰ Dawson, 2006.

generations for its solution. One of the problems, the *Entscheidungsproblem*, interested to a young and brilliant mathematician: Alan Turing. In 1936, Turing wrote a crucial paper, *On computable numbers, with an application to the Entscheidungsproblem*, in which he proposed the idea of an universal machine (the antecedent of the programmable processor). At the same time, Turing demonstrated that this problem was undecidable for first-order logic. The infallibility was off the agenda, because no fixed computation procedure by means of which every definite mathematical assertion could be decided (as being true or false). In the same year of 1936, Emil Post published a characterization of computability remarkably similar to that of Turing¹¹.

Nevertheless, four years before Turing's crucial paper, Alonzo Church introduced lambda calculus, as a new kind formulation of logic (basing the foundation of mathematics upon functions rather than sets) and a way to go away from the Russell paradox. Nevertheless, the Kleene-Rosser paradox showed that the lambda calculus was unable to avoid set-theoretic paradoxes. Today, lambda calculus is at the heart of functional programming and has had a big influence in compilations representation as well as in reasoning representation. The Lambda-definability made possible developing functional programming (like Lisp), also creating efficient compilers for functional languages. And related to the computer algebra, mathematical proofs by computers can be done more efficiently by lambda terms¹².

Also in 1936, Church realized that lambda calculus could be used to express every function that could be computed by a machine¹³. Perhaps Gödel had broken the gold dream of logical perfection but people like Church still worked on to developing more powerful logical systems. From 1936 to 1938, Turing had studied with Church at Princeton University. Some years later, the mathematician John von Neumann, developed his idea of a computer architecture after the ideas of Turing, leading to the so-called "von Neumann architecture", that has been implemented in most of actual computers¹⁴.

But not only Church established a strong relationship between logic and computability. In 1960 Curry first noticed what a bit later was called the *Curry-*

¹¹ van Benthem et al 2006.

¹² Barendregt, 1997.

¹³ Wadler, 2000.

¹⁴ There is also the Harvard architecture are several middle or combined options, like the Modified Harvard architecture. In fact, most modern computers instead implement a modified Harvard architecture.

Howard Isomorphism. The fundamental ideas of Curry-Howard are that proofs are programs, that formulas are types, that proof rules are type checking rules and that proof simplification is operational semantics. At the end, that the ideas and observations about logic were ideas and observations about programming languages.

Nevertheless, we must go back to the year 1954. In that date, Martin Davis carried out one of the first computational logical experiments: he programmed and runned the Presburger's Decision Procedure for the first order theory of integer addition¹⁵. He proved that the sum of two even numbers was itself an even number. A year later, in 1955, Evert Beth and Jaakko Hintikka (independently) described a version of the basic proof procedure, being a computationally powerful technique.

Dag Prawitz in 1960 made another important (re)discovery: the *resolution rule*, described something obscurely thirty years before by Herbrand in his Ph.D. Thesis and definitively coined in the actual form by John Alan Robinson in 1963. The process of its creation consisted on combining a known inference rule with unification. Ironically, the new mechanical efficiency brought also a cognitive opacity. According to Robinson, resolution was more machine-oriented than human-oriented¹⁶. Although resolution was to be a great contribution the paper in which it was described remained unpublished in a reviewer's desk for a year. Artificial intelligence and formal logic converged into a main research field, although the MIT view was being build (as we'll see in the next section).

This same critical process led to Larry Wos and George Robinson (at Argonne) to the development of a new rule, which they called *paramodulation*. This rule was implemented into several famous Argonne theorem provers, being perhaps one of the most famous William W. McCune's OTTER (see section 2.3). Theorem proving was one of the first field in which symbolic (as opposed to numeric) computation was automated.

All these results made possible a functional automated deduction technology, useful for several fields of academy as well as of industry.

¹⁵ Robinson, 2000, 8.

¹⁶ MacKenzie, 2004, 79.

1.2. The summer of the heuristic approach and other things

As Robinson (2000) remarks, two very important meetings for the history of computational logic took place in the 1950s:

Darmouth Conference (1956): this conference, championed by John McCarthy, gave birth to the field of AI. The crucial results of the common research of Herbert Simon (an economist) and Allen Newell (a mathematician) were presented there. They created a heuristic theorem-proving program, using the computer JOHNNIAC at RAND Corporation. With their computer program Logic Theorist, they solved automatically thirty-eight of the first fifty-two theorems in chapter 2 of the *Principia Mathematica*. Simon tell us that after informing Russell upon these results, he received an ironic answer: "if we'd told him this earlier, he and Whitehead could have saved ten years of their lives. He seemed amused and, I think, pleased"¹⁷. Logic Theorist can be considered like the first Expert System (ES).

Cornell Summer School in Logic (1957): Plenty of researchers attended this course; among them, Martin Davis, Hilary Putnam, Paul Gilmore and Herbert Gelernter. As Robinson (2000) recalls, Gelernter (a heuristic enthusiast from IBM) provoked Abraham Robinson to give a lecture using these methods and focusing on proof seeking. Gelernter's lecture influenced Gilmore, Davis and Putnam, researchers that would write their Herbrand-based proof procedure programs.

Without developing a long history about AI debates, I consider necessary to expose, at least briefly, some historical notes on its basic schools. There were two basic AI approaches¹⁸:

a) The MIT View: it considered AI as a heuristic, procedural, associative way of producing artificial-generated knowledge. Marvin Minsky and Seymour Papert were members of this approach. For these authors, formal logic was inadequate for the representation of knowledge required by any general approach to AI. They considered such a view as too static and rigid, preferring a procedural approach.

b) The Edinburgh-Stanford View: on the other hand, we have the logic view, championed by John McCarthy, who considered that AI knowledge could be mechanized because it could be axiomatized declaratively using

¹⁷ Stewart and Simon, 1994.

¹⁸ Robinson, 2000.

First Order Logic. They considered computational logic as the only way to achieve an Artificial Intelligence.

To be honest, both approaches were highly symbolic had more things in common than differences. In the middle of an AI's civil war, they were also called *neats* (logicians) and *scruffies* (proceduralists). It was only later that two real AI confronted approaches appeared, which can be summarized as *top down* and *bottom up* approaches:

i. *Top Down*: symbol system hypothesis (Douglas Lenat, Herbert Simon). The *top down* approach constitutes the classical model. It works with symbol systems, which represent entities in the world. Following to Brooks (1990, 4): "The symbol system hypothesis, states that intelligence operates on a system of symbols. The implicit idea is that perception and motor interfaces are sets of symbols on which the central intelligence system operates. Thus, the central system, or reasoning engine, operates in a domain independent way on the symbols". SHRDLU (Winograd), Cyc (Douglas Lenat) or expert systems are examples of it.

ii. *Bottom Up*: physical grounding hypothesis (situated activity, situated embodiment, connexionism). On the other side, the *bottom up* approach (championed by Rodney Brooks), is based on the physical grounding hypothesis. Here, the system is connected to the world via a set of sensors and the engine extracts all its knowledge from these physical sensors. Brooks talks about "intelligence without representation": complex intelligent systems will emerge as a result of complex interactive and independent machines.

In this sense, the MIT View and the Edinburgh-Stanford View both belonged to the *top down* approach. But let me continue with the conceptual analysis of computational logic.

1.3. About programming languages and several logics

During the 20th century logicians, mathematicians, philosophers and computer scientists studied a wide range of alternative formalisms designed for specific applications which do not appear to be easily handled by classical logic¹⁹: intuitionist, temporal, etc. Temporal logic, for example, is widely used

¹⁹ Galton, 1992.

for automated verification of semiconductor designs, as a specification language for design²⁰.

And as Wadler (2000) remarks, in 1969, W. A. Howard put together the results of Curry and Prawitz, and wrote down the correspondence between natural deduction and lambda calculus. Moreover, Wadler also states that the “Curry-Howard correspondence led logicians and computer scientists to develop a cornucopia of new logics based on the correspondence between proofs and programs.” This implied the common work between logicians and computer scientists.

Not all philosophical efforts towards better reasoning were developed under deductive modes of reasoning: the work of Mill or Peirce in non-deductive modes of reasoning, like induction or abduction, influenced also the field of Artificial Intelligence²¹. Traditional logic was not only flawed, it lacked soundness and completeness: *non-monotonic logics* appeared as the way to develop new powerful logics. But as Turner (1985) notes, while non-standard logics were investigated by philosophers and logicians since the 1930's at least, it was only in the late 1960's that McCarthy and others began to exploit those resources, in AI. And as Ben-Ari (1993) notes, the special demands of computer science generated interest in non-standard logical systems (like modal, temporal, many-valued, epistemic, intuitionistic, and other recent systems), leading to some highly applications of logic to software: the specification and verification of programs. For example, logic for action requires from certain previous, contextual and operational knowledge that must be used in order to achieve that action (Mishra, Aloimonos, Fermuller, 2009).

Although several authors worked on the relationships between the logics of knowledge and belief (as did Rudolf Carnap, Jerzy Los, Arthur Prior, Nicholas Rescher), it was G. H. von Wright who recognized that our discourse concerning knowledge and belief exhibits systematic features that admit of an axiomatic-deductive treatment²². He wrote a seminal work on epistemic logic (1951, *An Essay on Modal Logic*), whose ideas were extended by Jaakko Hintikka in *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (1962). Hintikka's book was a first attempt to combine efficiently knowledge and action (McCarthy & Hayes, 1969). This enabled the existence

²⁰ Halpern et al., 2001.

²¹ van Benthem, 2000.

²² Hendricks & Symons, 2009.

of a logic based on possible worlds (formalized by Kripke structures). Multi-agent systems can define a system as a non-static entity that changes over time. According to van Benthem & Martínez²³, Hintikka opened the possibility of working formally with a crucial idea: *information*. Information could then be considered as an information flow from states information to agents information (who have beliefs, knowledge or other informational situations). Perhaps it was not the perfect knight for logic, but better married than alone. Independently of Hintikka and belonging to a different research field, economy, Robert Aumann developed in the 1970's similar ideas. Aumann gave a mathematical formulation in a set-theoretical framework to the idea of common knowledge.

Later, in the 1980s and 1990s, epistemic logicians focused on the logical properties of systems containing groups of knowers and later still on the epistemic features of the so-called "multi-modal" contexts²⁴. It was in the 1980's that the notion of knowledge from epistemic logic deeply influenced computer science. TARK conferences (and later LOFT conferences) were an example of this new interdisciplinary approach. After considering the notions of inference, observation, introspection and self-correction agent (interactive) powers, logic was closer to the idea of behavioral equilibrium in groups of agents. This enabled a broad range of theories about it: logical dynamics (van Benthem), information dynamics in computer science (Abramsky), interactive epistemology or mathematical learning theory (Kelly)²⁵. Since then logical thinking may be considered as an interaction of its dynamical and social properties. As van Benthem (2007: 25) notes: "information is a pervasive aspect of reality, prior to cognitive action", that is, information is more than (formal) knowledge. *Substructural logics* arose (as nonclassical logics) in response to problems in foundations of mathematics and logic, theoretical computer science, mathematical linguistics, and category theory (Dosen and Schroder-Heister, 1993). Here are some of those logics: (1) lambek calculus (1950's), (2) linear logic (Girar 1980's), (3) fuzzy and multivariate logics, (4) relevance logics (Lukasiewicz), and (5) BCK logic.

²³ Benthem & Martínez, 2007.

²⁴ Hendricks & Symons, 2009.

²⁵ See van Ditmarsch, van der Hoek & Kooi, 2008 for an analysis on Dynamic Epistemic Logic.

As a resume of the explained data, we can conclude that:

Author	Development	Formalization of
Gentzen	Natural deduction	Proofs
Church	Lambda-calculus	Programs

As noted by Wadler (2000), lambda calculus, both typed and untyped, inspired several programming languages such as: LISP, 1960: J. McCarthy; Iswim, 1966: Peter Landin; Scheme (Lisp dialect), 1975: Guy Steele (one of the three future programmers of JAVA) and Gerald Sussman; ML ('metalinguage'), 1979: Milner, Gordon, Wadsworth, later 'Standard ML', the base of primordial LCF theorem prover, which later also inspired HOL and Isabelle theorem provers; Miranda, 1986: David Turner; Haskell, 1987: Hudak, Peyton Jones, Wadler et al, and, finally, O'Caml, 1996: Xavier Leroy.

Keith Clark, Alain Colmerauer, Pat Hayes, Robert Kowalski, Alan Robinson, Philippe Roussel, etc. deserve a lot of credit for promoting the concept of logic programming and helping to build the logic programming community. Currently, most programming is not done in logic programming but in imperative programming language (like Pascal, C, etc.). Other languages, like Prolog, for *Programming in Logic*, were developed from a syntactic-arithmetic perspective, rather than semantic, in order to program not with a set of instructions but directly with formal arithmetic logic as a total solution²⁶. According to Ben-Ari²⁷ Prolog was the first logic programming language and extensive implementation efforts transformed this language into a practical tool for software development. The history of Prolog is, perhaps, the history of the birth of logic programming. All started when Alain Colmerauer was working in 1963 on parsing and syntactic analysis. As Wadler (2000) remarks, at the end of this decade, and working on automatic translation, he tried to make automatic deduction from texts instead of just parsing them. Colmerauer studied the resolution principle and contacted Bob Kowalski, who worked at Edinburgh. Kowalski and D. Kuehner had recently devised a refinement of resolution (SL-resolution), which permitted, as Wadler (2000) also states, linear deductions with great efficiency and made possible the Edinburg Structure-Sharing Linear Resolution Theorem Prover. The meeting and sharing of ideas between Colmerauer and Kowalski created the

²⁶ Kowalski, 1982.

²⁷ Ben-Ari, 181.

language Prolog. Then the Marseille-Edinburg Prolog implementation started, offering an immediately applicable, available and attractive way to do logic programming. Prolog is expressive enough to execute non-procedural programs, and yet also contains enough ‘compromises’ with the real world to allow the execution of many programs efficiently²⁸.

In 1970 Robinson (2000) suggested that the new field should be called *computational logic*, and in December 1971 Bernard Meltzer convinced his university to allow him to rename his research department as “Department of Computational Logic”. The Horn clause version of resolution, as Robinson (2000: 12) declares, “could be used to represent knowledge declaratively in a form that (...) could then be run procedurally on the computer. Knowledge was both declarative and procedural at once.” This led to sequential logical programming, an approach that had also LISP as the crucial reference. By its competence and elegance, PROLOG was used as the first Kernel Language (KL0) of the Japanese Fifth Generation Project (FGP) of Computing Systems (1979-1981). The FGP dominated the mainstream in the 1980’s on computational logic (Robinson, 2000). With the increase of complex tasks assigned to a single CPU (for example in multiprogramming operating systems and real-time systems), was developed the concurrent computation and the concurrent logic programming. In that case, the multiple tasks can be executed sharing a single CPU rather than being executed in true parallelism on multiple CPUs. But the problem with parallel and concurrent computation is the difficulty of constructing and verifying algorithms that can benefit from parallelism, and of expressing these algorithms conveniently in a programming language. Nevertheless, logic programs have a natural interpretation as concurrent computations because do not express algorithms procedurally but instead express them declaratively where the formulas has no inherent ordering. Concurrent logic programming languages specify concurrent procedural interpretations for Horn clause programs, just as Prolog specifics a sequential procedural interpretation for the clauses. As Ben-Ari, 1993 remarks, an example of this case of concurrent logic programming language is GHC (Guarded Horn Clauses). A GHC program consists of a set of *guarded clauses*. A guarded clause is like an ordinary clause except that the literals in the body may be preceded by a sequence of literals called *guards*. They are separated from the body literals by a vertical bar rather than by a comma:

²⁸ Ben Ari, 1993, 8.

A :- G₁,...,G_k | B₁,...,B_n.

There is another kind of logic, the Constraint logic programming (CLP) which combines the flexibility and ease of declarative programming in logic with the power of search techniques that were developed during research on AI. According to van Roy & Seif, there are almost 30 useful programming paradigms²⁹. Each paradigm supports a set of concepts that makes it the best for a certain kind of problem. One programming language can support different paradigms (defined as a ‘set of programming concepts’). Perhaps we can classify them under two main programming paradigms, declarative and imperative, as I resume it in this box:

Programming Paradigms	Declarative		Imperative
	<i>Functional</i>	<i>Logical</i>	<i>Imperative/Procedural</i>
	Evaluation of mathematical functions avoiding state and mutable data	Use of mathematical logic for computer programming	Statements that change a program state.
Languages	APL, FP, Lisp, Erlang, Haskell, ML, F#, Scheme	PROLOG	FORTRAN, ALGOL, COBOL, Java, C, C++

As it is usual in scientific and technical opposite controversies, both approaches have interesting benefits and annoying side problems. One of the most famous approaches was exemplified by Edsger Dijkstra in this paper “Go To Statement Considered Harmful”³⁰. Dijkstra argued against the abusive use of imperative constructs and considered the command ‘GO TO’ as the archetypical kind of such an abuse IN his own words: “For a number of years I have been familiar with the observation that the quality of programmers is a decreasing function of the density of go to statements in the programs they produce. More recently I discovered why the use of the go to statement has such disastrous effects, and I became convinced that the go to statement

²⁹ Van Roy & Seif, 2004. Van Roy also offers a very useful visual chart of programming paradigms; see here <http://www.info.ucl.ac.be/~pvr/paradigmsDIAGRAMeng.pdf>

³⁰ Letter to *Communications of the ACM* (CACM), vol. 11 no. 3, March 1968, 147-148.

should be abolished from all "higher level" programming languages (...) The unbridled use of the *go to* statement has an immediate consequence that it becomes terribly hard to find a meaningful set of coordinates in which to describe the process progress. (...) The *go to* statement as it stands is just too primitive; it is too much an invitation to make a mess of one's program". Structured programming, according to Dijkstra, could not solve these problematic statements. At the same time, there was an attempt to develop a new managerial approach to production programming³¹. Perhaps imperative languages are not the perfect choice but it is undoubtable that they are dominating several computing fields. As a middle point position, Bob Kowalski proposed the equation ALGORITHM = LOGIC + CONTROL. A new paradigm was in the agenda, with a basic necessary characteristic: to maintain a strict separation of the declarative from the imperative aspects. From this perspective, algorithms consist of a problem description (logic) along with a strategy to carry out computations of these descriptions (control).

First-order logic has also been used at the heart of a lot of database systems, like its syntactic variants 'structured query language' (SQL) or 'query-by-example' (QBE). First-order logic has also been implemented into database using *relational algebra*, and at the same time making possible to scale correctly the large databases³².

The last of our covered topics in this section this section is information. On the early 1980's, Jon Barwise and John Perry took seriously the idea of study the concept of 'information' from a logical perspective. They set up 'situation semantics' and founded, in 1983 at Stanford University, the Center for the Study of Language and Information (CSLI). It was a place of great interaction among philosophers, mathematicians and computer scientists. On the other hand, not only the American researchers were involved into this new analysis, European colleagues were also working on natural language semantics (and the informational turn), as Peter van Emde did, who was a pioneer on the study of the parallels between natural and programming languages. By 1986 it was created the Institute for Language, Logic & Information (ILLI, but originally known in Dutch as ITLI, the Instituut voor Taal, Logica en Informatie), later renamed as ILLC (1991, Institute for Logic, Language and Computation). In 1990, the European Association for Logic, Language and Information (FoLLI) created the annual ESSLLI summer schools, with a great

³¹ MacKenzie, 2001: 39.

³² Halpern et al, 2001.

orientation toward mixed research fields. Alexandru Baltag, focusing on information aspects, and a researcher at Oxford University, explains in his website³³ that Logic is currently evolving into a new interdisciplinary field, devoted to studying 'qualitative informational interactions'; i.e. its sources, flows, gatherings, as well as processing, combining and upgrading (or more generally the 'dynamics') of qualitative Information. In sum, logic naturally underlies most of theoretical computer science, but it also has the potential of acting as a unifying force for many academic disciplines around the concept of Qualitative Informational Interaction.

1.4. New Journals and the Institutionalization of the Field

In 1984 the first issue of The Journal of Logic Programming appeared (Robinson, 2000). Several publications on the field appeared across the next years, like: Journal of Logic and Computation (1990)³⁴, Journal of Functional and Logic Programming (1995)³⁵, Journal of Multiple-Valued Logic and Soft Computing (1996)³⁶, or Transactions on Computational Logic (TOCL, 2000)³⁷. The increasingly growth of journals in the field denotes a strengthening of this research area and the existence, for the first time, of a critical mass of specialists on the field. A huge literature was also being written. For example, Michael Poe *et al* compiled in 1984 a comprehensive bibliography on Prolog programming with more than 700 entries. In a few short years, the concepts, methods, and topics of computer scientists changed and were shaped to attend the requirements of professionalization and institutionalization.

1.5. Programming robots

Computational logic has also provided solutions to engineering developments on robotics. Flux high-level programming language, for example, enables robots with an internal model of their environment. With an implementation of fluent calculus, this language provides a solution to the

³³ <http://www.comlab.ox.ac.uk/people/Alexandru.Baltag/> accessed in March 4th, 2014.

³⁴ <http://logcom.oxfordjournals.org/>, accessed in March 4th, 2014.

³⁵ <http://scholar.lib.vt.edu/ejournals/JFLP/>, accessed in March 4th, 2014.

³⁶ <http://www.oldcitypublishing.com/MVLSC/MVLSC.html>, accessed in March 4th, 2014.

³⁷ <http://tocl.acm.org/>, accessed in March 4th, 2014.

classic *frame problem*, using the concept of state update axioms. This allows us to address several aspects of reasoning about actions like ramifications, qualifications, nondeterministic actions, concurrent actions, continuous change or noise inputs.

2. From Computers to Logic

In Thomason it is explained³⁸ that the AI authors have made real contributions to philosophical logic, and that they also have drawn heavily on past work in logic and philosophy. Computer science is not a passive receiver of logic, on the contrary: it is a direct arrow onto logic development. John McCarthy challenged philosophers to contribute to AI by surveying a number of problems of interest to both fields. This was not only a possible way of improving AI, but it was also an opportunity to obtain a new and significant philosophical advancement. From the perspective of AI, there were, said McCarthy, several philosophical problems, like the importance of common sense, the roles of logic, non-monotonic reasoning, free will, knowledge and belief, reasoning about context, and some more general epistemological concerns.

Van Benthem³⁹ states that computer sciences are changing academia and that at the same time they have a cultural role, providing a steady stream of new notions and information to the academia itself (brain sciences, logic, philosophy, etc.). For example, Turing machines, by allowing for the computability of mathematical notation, lead to deep results as the undecidability of natural questions like the halting problem. On the other hand, computer sciences provided a great number of ideas that were later decoupled from its initial practical setting, such as: automata theory, complexity theory, semantics of programs, type theory and linear logic, process theory, database theory, AI, and so on.

Once computers were made, they proved to be useful for practical purposes like the mechanization of theorem proving. Considering proofs as programs, Simon and Newell created⁴⁰ (and called their program) the Logic Theory Machine, which was also programmed by J.C. Shaw, a scientist at

³⁸ Thomason, 1989.

³⁹ Floridi, 2008.

⁴⁰ MacKenzie, 2001: 66.

RAND Corporation and Carnegie Technical Institute. To Newell, Simon and Shaw's surprise, Logic Theorist (the other name for Simon and Newell's program) produced a shorter, more elegant proof of theorem 2.85 of Whitehead and Russell's *Principia Mathematica* – indeed, the axioms and theorems to be proven were taken from the *Principia*. Nevertheless, The Journal of Symbolic Logic would not publish the description of a proof co-authored by a computer program. Logic Theorist represented a cutting-edge use of computers because of its reliance on heuristics, and because it manipulated not numbers but information, represented in symbolic form.

Soon it was clear that, as MacKenzie remarked (2001: 72), the automation of mathematical proof would require not just propositional logic, but also predicate logic. In this sense, the initial contributions of logicians to automated theorem proving focused on automating 'decision procedures', that is, algorithms.

In summer of 1958, MacKenzie continues, Hao Wang was able to design programs which had proven all the propositional logic theorems from the *Principia*, using less than 30 seconds of CPU. This drove Wang to defend the superiority of the algorithmic approach over the heuristic one. Wang was skeptic about Simon & Newell's heuristic approach, considering their research as an amateur and unprofessional work. Hilary Putnam and Martin Davis spent the summer of 1958 studying algorithms for propositional calculus and developed the so-called Davis-Putnam procedure for checking the satisfiability of formulae in propositional calculus.

In early July 1959, Robert S. Ledley and Lee B. Lusted wrote for *Science* a paper entitled "Reasoning Foundations of Medical Diagnosis"⁴¹. At the conclusions section they wrote: "The mathematical techniques that we have discussed and the associated use of computers are intended to be an aid to the physician. This method in no way implies that a computer can take over the physician's duties. Quite the reverse; it implies that the physician's task may become more complicated. The physician may have to learn more; in addition to the knowledge he presently needs, he may also have to know the methods and considerations under consideration in this Today you can change 'physician' for 'logician' and the value of their words is still meaningful. Computer science would not exist without logic, but once it was created, it changed the ways of doing logic. This is a reflexive process. And there are

⁴¹ *Science*, vol. 130, num. 3366, 9-21.

several forms of analyzing such symbiotic relationship; one of those ways is to think about the influence of computers upon logic.

In 1962, John Alan Robinson found a new and, as MacKenzie (2001: 77) wrote, “powerful rule of deductive inference which became paradigmatic in automatic proving techniques: resolution”. Resolution is the deductive system most often used in mechanizations of logic. And in the 1960’s, Dijkstra and Hoare worked on the effects of structural sequential programs running on single computers, which afterwards lead to dynamic logic (developed by Salwick, Pratt, and others). Morevoer, as van Benthem (2007) remarks, “the study or never-ending computation over infinite data streams, which cannot be constructed, only observed”, opened a new fundamental theory: co-algebra (Aczel and Meldler, 1989), a field that already changed mathematical proof methods in analysis and set theory.

In the 1980’s the analysis of distributed systems and parallel computation on many computers open the way for Milner’s process algebra, which became an excellent general theory of communication processes in physics and biology, among others.

Despite engineers’ preference for variable-free frameworks over first order logic, such inclination is quite silent: computer sciences still have a pragmatic influence over logics. For example, computer scientists offered logicians new challenges and perspectives for modeling and constructing discrete systems (Thomas, 2000). The interest in the specification of complex systems has allowed the merge of formula based frameworks (like temporal logic, VDM, Z) with diagram based formalisms (like SDL, UML, Statecharts).

2.1. Educational purposes

Among a broad list of computer programs for learning of (formal) logic, several applications like Tarski’s World, Hyperproof, Logic Daemon, Jape 3.2., CPT I or MyC have been developed. The first steps towards a visual approach to logic through computer tools was made by J. Barwise and J. Etchemendy. The later is involved nowadays in The Openproof Project⁴² at Stanford’s Center for the Study of Language and Information. The center is mainly concerned with the application of software to problems in logic, and since the early 1980’s they have been developing innovative and

⁴² See: <http://ggweb.stanford.edu/openproof/> accessed on March, 3rd 2014.

commercially successful applications in logic education. *Hyperproof*, one of the leading programs in this new era of teaching logic is a *heterogeneous* logic program; the term *heterogeneous* arises from the formal integration of the diagrammatic and sentential representations. Barwise and Etchemendy designed Hyperproof to support reasoning in which information from differing modalities (sentential and graphical) is combined or transferred from one modality to another. Oberlander, Stenning and Cox⁴³ argued that the virtue of Hyperproof, lays not in visualization, but in its multimodality. I disagree completely: the choice of *seeing* in completely 3D visual environment allows for a different kind of cognition, better adapted to human cognitive capacities⁴⁴. Hyperproof has a better cognitive fitness and introduces students into the new research paradigm of e-Science (Vallverdú, 2009). Visual thinking has been rehabilitated in epistemology of mathematics and logic due to computer sciences (Mancosu, Jørgensen & Andur, 2005).

The educational logic platforms have reached even the cellular phones commercial domain. The program Logic 101 is designed for constructing derivations in introductory logic courses. It checks several parameters, such as the syntax of symbolic sentences or whether each line can be derived from previous lines, and also indicates derivable lines by colour, and displays derivation rules for each derivable line. The completed derivations can even be sent by e-mail. Logic 101 runs under both iPhone and iPod. There are also two other logic apps there: Syllogism and Logic 100.

We can also find plenty of Java applets (Tilomino 2, xLogicCircuits, Truth Table Constructor, DC Proof, and so on) freely hosted and distributed among important education institutions all around the world. These technological machines and artifacts make possible not only to study logics with contemporary tools but also to think differently about logics. That the Internet has changed deeply the ways by which we make science and establish human relationships is something very clearly known. Computational logic has made significant contributions to this, especially to the very-large-scale integration (VLSI). VLSI is the process of creating integrated circuits by combining thousands of transistor-based circuits into a single chip⁴⁵. Testing

⁴³ J. Oberlander, K. Stenning and R. Cox (1996), The Multimodal Moral, online document. See: <ftp://ftp.cogsci.ed.ac.uk/pub/graphics/itallc96.ps>, accessed on April 14th 2014.

⁴⁴ Vallverdú, 2010.

⁴⁵ See: http://en.wikipedia.org/wiki/Very-large-scale_integration, accessed on April 14th 2014.

and verifying modern hardware and software are really complex tasks and computational logic has contributed to make this process smoother. At the same time, the Internet has modified the social ways of communicating and debating computational logic. Websites, middleware, multimedia tools have provided a new thinking space for the logic community.

2.2. From theory of computing to logics

One of the seminal influences between computer science and logic was the Curry-Howard correspondence, which is the simple observation that two at-the-time-seemingly-unrelated families of formalisms, the proof systems on one side and the models of computation on the other were in fact structurally the same kind of objects⁴⁶. If it is true that the works of Curry⁴⁷ and Howard⁴⁸ belonged to the field of logic, its consequences were analyzed by experts belonging to the computer science domain. N. G. de Bruijn (used the lambda notation for representing proofs of the theorem checker *Automath*, and represented propositions as "categories" of their proofs), who was very likely unaware of Howard's work stated the correspondence independently, but also Stephen Kleene, Joachim Lambek, among others, made significant contributions to this debate.

Another field in which an initial philosophical development has led to a computer science innovation is that of *game semantics*. This is an approach to formal semantics that grounds the concepts of truth or validity on game-theoretic concepts. And as it is well known, Paul Lorenzen was the first to introduce, in the late 1950's, a game semantics for logic. By mid-1990's, game semantics had changed several related fields, such as theoretical computer sciences, computational linguistics, AI or the way we understand programming languages, especially through the efforts from van Benthem who took a deep interest on the relation between logic and games. For van Benthem⁴⁹, game logics describe general games through powers of players for forcing outcomes. Specifically, they encode an algebra of sequential game operations such as choice, dual and composition. In this sense, logic games

⁴⁶ See: http://en.wikipedia.org/wiki/Curry-Howard_correspondence, accessed on April 14th 2014

⁴⁷ Curry, 1934, 1958.

⁴⁸ Howard, 1969.

⁴⁹ van Benthem, 2003.

are special games for specific purposes such as proof or semantic evaluation for first-order or modal languages. This author has made some crucial to computer logic as well as some critical thinking about the future of logic, arguing that there is something like a triangle of theory, reality, and new design, and also an interaction between all these different perspectives. Phenomena like reasoning or information flow suggest a natural Triangle of perspectives: (somewhat normative) *theory*, *empirical reality*, but also *virtual reality*, the construction of *new systems and new forms of behavior* by the interplay of the former two. Accordingly, theoretical logic, empirical psychology, and constructionist computer science form a natural Triangle of disciplines, each approaching the topics on his agenda with a different thrust. There is thus a fusion of fields more than a movement of influences from one field to another⁵⁰.

The P = NP problem, a relationship between the complexity classes P and NP, as Ivancevic & Ivancevic (2007: 124) described it, “is an unsolved question in theoretical computer science”; it is considered “to be the most important problem in the field”. Logic has played a crucial role in the development of the theory of NP-completeness by formalizing the concept of combinatorial explosion. At the same time, this intensive research led to a new set of relativizing proofs, natural proofs and algebrizing proofs⁵¹.

2.3. Automation of sound reasoning

Beyond educational uses of computers into logical arena, we can find fields in which the introduction of computers has changed the way of doing and understanding logic. One of them is *proof automation*. Despite the fact that most of computer programs related to proofs have been designed for proof verification (Coq, HOL, etc.), some of them are automated reasoning programs. OTTER, as Raynor (1999:217) wrote “is a fourth-generation Argonne National Laboratory deduction system whose ancestors (dating from the early 1960s) include the TP series, NIUTP, AURA, and ITP”. Otter must have a heuristic or strategy in order to avoid the many millions of possible deducible conclusions, but is doesn’t imply that it must be ‘person-oriented

⁵⁰ van Benthem, 2006.

⁵¹ Garey & Johnson, 1979.

reasoning⁵². This automatic canonicalization and simplification (by demodulation or subsumption) can lead us to a neat proof that would not be so obvious to a human investigator⁵³. Does it imply that this automation creates black boxes or unintuitive proofs? No, as Robinson said (2000:15), the Argonne group undertook “a large-scale pursuit of computing real mathematical proofs”. Machines can do mathematics as well as humans. The Argonne program found a proof of the Robbins conjecture. Indeed, as Robinson continues, in approximately 20 hours of search on a SPARC 2 workstation, the program EQP (designed by McCune), found a proof in which it could be demonstrated that every Robbins algebra was a Boolean algebra.

2.4. Verification systems

Computer sciences also need logic. The work on verification and validation (V&V) of huge and complex systems requires logical tools in order to make things easier and more effective. As soon as these tools appeared, its use on proof verification allowed for a complete view on automatic reasoning and automated V&V. This led to the actual IEEE Std 610.12-1990 V&V standard. Proofs are now faster, more reliable and cheaper. But this came with a price as one can see by a short list of the historical V&V mistakes: Ariane 5 rocket, London Ambulance Dispatching System, Therac-25 or the Nike disaster. Debugging was considered a necessary task in the beginnings of systematic computer programming. Computers models and their results should be reliable. This process could be done with a statistical approach, although, as Dijkstra said in 1969: “program testing can be used to show the presence of bugs, but never to show their absence!”⁵⁴. On that same year, Tony Hoare wrote the article “An Axiomatic Basis for Computer Programming”, which led to the consolidation of the formal verification process. At that point, two thirds of the cost of software projects were dedicated to locating and remove programming bugs.

⁵² Wos & Fitelson, 2005.

⁵³ Dana Scott, quoted by Wos & Fitelson, 2005, 717.

⁵⁴ McKenzie 2001, 45.

2.5. From logic and computer sciences to other fields

Finally, we have the social software. It is a new (since 1996) and, as Pacuit (2005: 10) remarks, interdisciplinary research program that borrows mathematical tools and techniques from game theory and computer science in order to analyze and design social procedures. Rohit Jivanlal Parikh (Pacuit's PhD advisor), who was influential in the development of the field, is leading these researches. The goals of research in this field are modelling social situations, developing theories of correctness, and designing social procedures⁵⁵.

3. Other related issues: Patents

I would like to finish my analysis with a brief overview about one key problem for any researcher dedicated to computer sciences: patents. Klemens (2006) has made an extended study of the main problems of patenting mathematics or programs (for him, equivalent domains). I am not against (computer logic) patents, but we should consider its limits. Can we patent numbers or mathematical operations? If not (and obviously we cannot), why do we allow patented algorithms?

Let me consider though an example shown to me by Thomas Hales⁵⁶, the Ståalmarck's method⁵⁷. Ståalmarck's method is a proof search algorithm for finding proofs of propositional tautologies in a proof system called *the dilemma proof system*. The search procedure is based on a notion of *proof depth* corresponding to the degree of nestings of assumptions in proofs. Ståalmarck's algorithm has been successfully used in industrial applications since 1989, for example in verification of railway interlockings and of aircraft control systems. It has a Swedish Patent (467 076), where Ståalmarck's

⁵⁵ For further readings on social software, see Parikh, 2002. See also: [http://en.wikipedia.org/wiki/Social_software_\(social_procedure\)](http://en.wikipedia.org/wiki/Social_software_(social_procedure)) , accessed on april 14th 2014.

⁵⁶ Private electronic correspondence January, 28th 2009.

⁵⁷ G. Ståalmarck, "A system for determining propositional logic theorems by applying values and rules to triplets that are generated from a formula". Swedish Patent No. 467 076 (approved 1992) U.S. Patent No 5 276 897 (approved 1994), European Patent No 0403 454 (approved 1995).

algorithm is a patented technique for tautology-checking which has been used successfully for industrial-scale problems, implemented as a HOL derived rule.

Conclusions

From several evidences we must conclude that the relation between logic and computer sciences is mutual and symbiotic. Both fields have influenced each other and as a result of this symbiotic process, their ideas and performances were improved. Moreover, such mixture of techniques and ideas has produced a spontaneous and continuous interchange between both communities of researchers. In sum, this opened up new thinking spaces through which boundaries between are blurred under a deep layer of common interests, tools and ideas.

Acknowledgements

Financial support for this research was received from the Spanish Government's DGICYT research project: FFI2011-23238, "Innovation in scientific practice: cognitive approaches and their philosophical consequences". I thank to Amnon Eden, Thomas Hales and Johan van Benthem for their support, suggestions, answers or commentaries to my queries in some step of my long and slow research. I thank to Peter Skuce for his patience and linguistic support. Finally, I really appreciate the deep, hard and fine task of two patient anonymous reviewers, who suggested very good ideas to improve the quality and coherence of this paper. After this process, I need to clarify that all errors, mistakes, etc. in the article are exclusively my own. This is the true way of making science.

References

- Abramsky, S., Gabbay, D.M., and Maibaum T.S.E., 1993, *Handbook of Logic in Computer Science*, Vol 2, London, Oxford University Press.
- Aczel, P. and Mendler N.P., 1989, A Final Coalgebra Theorem. *Category Theory and Computer Science*, 357-365.
- Barendregt, H., 1997, The Impact of the Lambda Calculus in Logic and Computer Science. *The Bulletin of Symbolic Logic*, 3 (2), 181-215.
- Ben-Ari, M., 1993, *Mathematical Logic for Computer Science*, UK, Prentice Hall.
- Bringsjord, S. and Ferrucci, D.A., 1998), Logic and Artificial Intelligence: Divorced, Still Married, Separated ...?, *Minds and Machines*, 8 (2), 273-30.
- Brooks, R., 1990, Elephants Don't Play Chess. *Robotics and Autonomous Systems*, 6 (1-2): 3-15.
- Curry, H., 1934, Functionality in Combinatory Logic. In: *Proceedings of the National Academy of Sciences*, 20, 584-590.
- Curry, H. B. and Feys, R., 1958, *Combinatory Logic Vol. I*, Amsterdam, North-Holland.
- Davis, M., 1988, Influences of mathematical logic on computer science. In: *The Universal Turing Machine: A Half-Century Survey*, USA: OUP.
- 2000, *The Universal Computer: The Road from Leibniz to Turing*, USA: W. W. Norton and Company.
- Dawson, J.W., 2006, Gödel and the Origins of Computer Science. In: *CiE 2006*, LNCS 3988, 133-136.
- Ditmarsch van, H., van der Hoek & W, Kooi, 2008, *Dynamic Epistemic Logic*, London, Springer.
- Dosen, K. and Schroder-Heister. P., 1993, *Substructural Logics*, USA, Clarendon Press.
- Floridi, L., 2008, *Philosophy of Computing and Information: 5 Questions*, UK: Automatic Press/VIP.
- Flach, P.A., 2005, Modern Logic and Its Role in the Study of Knowledge. In: *A Companion to Philosophical Logic*, London, Blackwell, 680-693.
- Galton, A., 1992, Logic as a Formal Method, *The Computer Journal*, 35 (5), 431-440.
- Garey, M.R. and Johnson, D.S., 1979, *Computers and Intractability: A Guide to the NP-Completeness*, New York: W.H. Freeman.
- Gentzen, G., 1934, Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, 39, 176-210, 405-31.
- Halpern, J.Y. et al., 2001, On the Unusual Effectiveness of Logic in Computer Science. *The Bulletin of Synthetic Logic*, 7 (2), 213-236.
- Harrison, J., 2009, *Handbook of Practical Logic and Automated Reasoning*, UK, CUP.

Hendricks, V. and Symons, J., 2009, Epistemic Logic. In: *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, Edward N. Zalta (ed.). Accessed on March 3rd, 2014: <http://plato.stanford.edu/archives/spr2009/entries/logic-epistemic/>

Howard, W., 1969, The formulae-as-types notion of construction. In: Seldin, J. et al., *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, Academic Press, 479-490.

Ivancevic, V.G. and Ivancevic, T.T, 2007, *Computational Mind: A Complex Dynamics Perspective*, Berlin, Springer.

Klemens, B., 2006, *MA+H You Can't Use. Patents, Copyright and Software*, Washington, D.C., Brookings Institution Press.

Kowalski, R., 1982, Logic as a computer language. In: *Logic programming*, London, Academic Press, 3-18.

Mackenzie, D. 2004, *Mechanizing Proof: Computing, Risk, and Trust*, Cambridge (MA), The MIT Press.

Mancosu, P., Jørgensen, K.F. and Andur, S., 2005, *Visualization, Explanation and Reasoning Styles in Mathematics*, USA, Springer.

McCarthy, J. and Hayes, P.J., 1969, Some Philosophical Problems from the standpoint of the Artificial Intelligence. Accessed on March 4th, 2014:
<http://www-formal.stanford.edu/imc>

Mishra, A., Yiannis Aloimonos, Y, & Fermuller, C. (2009), Active Segmentation for Robotics, *IROS 2009*. Downloadable at:
http://www.umiacs.umd.edu/~mishraka/downloads/iros2009_activeSeg.pdf, Accessed on April 12th 2014.

Pacuit, E., 2005, *Topics in social software. Information in strategic situations*. PhD thesis, New York.

Parikh, R., 2002, Social Software. *Synthese*, 132, 187-211.

Poe, M.D. , Nasr, R., Potter, J. and Slinn, J., 1984, A KWIC (Key Word in Context). Bibliography on Prolog and Logic Programming. *Journal of Logic Programming*, 1 (1), 81-142.

Post, E., 1936, Finite combinatory proceses. Formulation I. *Journal of Symbolic Logic* 1, 103-105.

Raynor, J., 1999, *The International Dictionary of Artificial Intelligence*, Chicago, Glenlake Publishing Company, Ltd.

Robinson, J.A., 2000, Computational Logic: Memories of the Past and Challenges for the Future. In: *CL 2000, LNAI 1861*, Berlin, Springer Verlag, 1-24.

Stewart, D, 1994, Herbert A. Simon: Thinking Machines, *OMNI Magazine*, Interview June 1994. Accessed on March 4th, 2014:
<http://www.astralglia.com/webportfolio/omnimoment/archives/interviews/simon.html>

Thomas, W., 2000, Logic for Computer Science: The Engineering Challenge. Lecture given at the Dagstuhl Anniversary Conference, Saarbrücken.

- Thomason, R., 1988, Philosophical logic and artificial intelligence. *Journal of Philosophical Logic*, 17 (4), 321-327.
- 1989, *Philosophical Logic and Artificial Intelligence*, Dordrecht, Kluwer Academic Publishers.
- Turner, R., 1985, *Logics for AI*, Chichester, Ellis Horwood.
- Vallverdú, J., 2009, Computational Epistemology and e-Science. A New Way of Thinking. *Minds and Machines*, 19 (4), 557-567.
- 2010, Seeing for Knowing. The Thomas Effect and Computational Science. In: *Thinking Machines and the Philosophy of Computer Science: Concepts and Principles*, USA, IGI Global Group, 140-160.
- van Benthem, J. 2000, Reasoning in reverse. In: *Abduction and Induction: Essays on their Relation and Integration*, Dordrecht, Kluwer Academic, ix-xi.
- 2003, Logic Games are Complete for Game Logics. *Studia Logica*, 75, 183-203.
- 2006, Where is Logic Going, and Should It?. *Topoi*, 25 (1-2), 117-122.
- van Benthem, J., et al, 2006, *The Age of Alternative Logics. Assessing Philosophy of Logic and Mathematics Today*, Series: Logic, Epistemology, and the Unity of Science , Dordrecht, Springer.
- van Benthem, J., 2007, An Interview on the Philosophy of Information, accessed on March 4th 2014 <http://www.illc.uva.nl/Research/Publications/Reports/PP-2008-01.text.pdf>
- van Benthem, J. and Martínez, M., 2007, The stories of logic and information. In: *Philosophy of information*, Amsterdam, North Holland.
- van Roy, P and Seif, H., 2004, *Concepts, Techniques and Models of Computer Programming*, Cambridge (MA), MIT Press.
- Wadler, P., 2000, Proofs are Programs: 19th Century Logic and 21st Century Computing, *Dr. Dobb's Journal* (published as "Old ideas form the basis of advancements in functional programming"), downloadable from: <http://homepages.inf.ed.ac.uk/wadler/papers/frege/frege.pdf>, accessed on April 12th 2014. See also: <http://homepages.inf.ed.ac.uk/wadler/topics/history.html>, ibid.
- Wos L. and Fitelson, B., 2005, The Automation of Sound Reasoning and Successful Proof Finding. In: *Companion to Philosophical Logic*, USA, Blackwell, 709-723.

O problema em torno do conceito de “raça”: será que a “raça” é uma *ilusão*?*

Joana Inês Pontes
(Instituto de Filosofia, Universidade do Porto)
joanainespontes@gmail.com

(...) o estudo da raça, tal como a teoria dos números, é algo que leva os homens à loucura.¹

Ao longo da última década, o conceito de “raça” tem estado no centro de diversos debates políticos, científicos e filosóficos, tendo surgido uma grande afluência de trabalhos em metafísica da “raça” e algumas abordagens semânticas. Embora seja amplamente aceite que as “raças” não partilham de essências que demarcam as suas características físicas e intelectuais, as opiniões sobre aquilo que o termo “raça” significa divergem, dando origem a movimentos distintos: de que falamos quando falamos em “raça”? A “raça” é (ou não) real? Pode a “raça” ser utilizada num contexto científico? As recentes investigações dos teóricos da “raça” dividem-se em três respostas gerais que diferem, fundamentalmente, no carácter metafísico da “raça”: i) *eliminativismo racial*: visão segundo a qual as “raças” não existem - falar de “raças” não é melhor do que falar sobre bruxas ou unicórnios²; ii) *naturalismo racial*: as “raças” podem existir como grupos populacionais com relevância biológica, baseados em características naturais (genéticas), mas sem qualquer relação com as teorias populares da “raça”³; iii) *construtivismo*

* Trabalho realizado sob o financiamento da FCT do projecto “O (des)acordo científico em torno do conceito de ‘raça’”.

¹ Banton & Haerdwood, 1975, 8.

² Appiah, 1992, 1996, 1998; Zack, 2002, 2003.

³ Andreasen, 2000, 2005; Kitcher, 1999, 2007.

racial: a “raça” existe, é real, mas é uma construção social – um grupo cultural, *folk cultural concept*⁴.

A escolha entre uma destas aproximações à “raça” pode parecer, num sentido depreciativo, apenas semântica, mas o debate sobre aquilo que se entende por “raça” joga um papel determinante na “avaliação complexa de uma série de práticas e considerações normativas”⁵. A eliminação ou conservação do conceito de “raça” no vocabulário corrente (senso comum) e científico está na base de várias tomadas de decisão, tais como o abandono da variável “raça” na produção de fármacos em benefício de certos grupos genéticos e/ou da sua utilização na realização de estudos de grupos de risco, entre outros estudos populacionais muitas vezes denominados de étnicos.⁶

Partindo da consideração da tese de Anthony Appiah (1992) de que *não há tal coisa chamada “raça”* (i.e. da desconstrução biológica do conceito de “raça” vulgarizada em filosofia e nas ciências humanas em geral) e colocando ênfase no problema de saber se as “raças” podem ou não ser pensadas como tipos naturais (TN), este artigo pretende mostrar que a “raça” existe, é real, possui implicações aquando da sua utilização e que Appiah não está correcto ao concluir que a “raça” (concebida como conceito de TN) é uma ilusão: i) há um desacordo semântico e metafísico entre *eliminativistas*, *construtivistas* e *naturalistas* sobre aquilo que os termos ou conceitos raciais se referem, determinando o debate actual sobre o lugar da “raça”; ii) os *eliminativistas* erram ao negar a possibilidade de se desenvolver um conceito adequado de “raça”: nem sempre o que as pessoas pensam sobre a “raça” no senso comum (SC) corresponde ao seu significado – o facto de as “raças” não serem categorias biológicas, isso não faz delas coisas “ilusórias”; iii) não

⁴ Glasgow, 2009; Haslanger, 2008; Mills, 1999.

⁵ Haslanger, 2008.

⁶ As questões relativas à definição e validade do uso da variável “raça” e as suas limitações na captura de similaridades biológicas ou genéticas têm sido objecto de reflexão crítica de vários investigadores na saúde pública; ver Laguardia J. (2004), O uso da variável “raça” na pesquisa em saúde. *Physis*, 14(2): 197-234; Aspinall PJ. (1998), Describing the “white” ethnic group and its composition in medical research. *Soc Sci Med*, 47(11):1797-1808; Escarce, J. J. (2005) How Does Race Matter, Anyway? *HSR: Health Services Research*, 40. Thomas, S. B. (2001) The Color Line: Race Matters in the Elimination of Health Disparities. *American Journal of Public Health*, 91, 1046-1048. Frank, R. (2007) What to make of it? The (Re)emergence of a biological conceptualization of race in health disparities research. *Social Science & Medicine*, 64.

faz sentido continuarmos presos na falsa divisão radical entre cultura e biologia (cultura/ciência): a “raça” traduz o carácter biosocial da investigação biomédica, apresentando-se como uma realidade biológica e, simultaneamente uma construção social – *biosocial concept*; iii) as diferenças entre as “raças” não são diferenças essenciais, mas *superficiais*, i. e. visíveis a “olho nu”⁷ e suficientemente relevantes para a sua classificação em classes/grupos e utilização em contexto comum e científico.

Desacordo metafísico

A ideia de raça só foi desenvolvida no último século, por isso, mesmo as pessoas que juram a sua existência, não concordam com a sua natureza.⁸

Deve o conceito de “raça”⁹ ser banido do nosso discurso e práticas (*eliminativistas*) ou deve ser conservado, pelo menos no presente (*anti-eliminativistas ou preservacionistas*)? Saber se a “raça” deve ou não permanecer no nosso discurso implica, antes de mais, que o termo “raça” possua um significado específico, sendo que aquilo que o conceito de “raça” significa faz parte do próprio problema, tornando-se difícil objectivar.

⁷ Hardimon, 2003, 442.

⁸ Marks, 2008, 21.

⁹ Importa esclarecer que o conceito de “raça” aparece ao longo deste artigo entre aspas pelo simples facto de não existir um consenso universal sobre a sua natureza e definição – o conceito de “raça” é ambíguo e o seu uso arbitrário. Em contexto anglo-saxónico, particularmente, no universo americano, o conceito de “raça” nunca é utilizado entre aspas, a legislação americana comporta a divisão racial nos seus estatutos – ver em OMB, Standards for the Classification of Federal Data on Race and Ethnicity http://www.whitehouse.gov/omb/fedreg_race-ethnicity – apesar de não existir consenso em relação à definição. Em Portugal, considere-se Europa, a regra geral quanto ao emprego e manipulação do conceito de “raça” segue o consenso da UNESCO (1992) de que “raça” é um termo conflituoso e violento, sem base biológica que deve ser permanentemente substituído pelo termo etnia. O dicionário de língua Portuguesa contém o termo, sendo que as definições apresentadas dividem-se sempre entre origem, descendência e diferenças étnicas. No discurso comum não parece existir uma directiva, lei ou declaração que obrigue à eliminação radical do uso do conceito “raça”, embora existam investigadores oriundos das ciências sociais que argumentam a favor da eliminação do conceito e do seu uso exclusivo como categoria social, ver Manuela Cunha, 2000, A Natureza da “raça”, *Sociedade e Cultura* 2, 13: 191-203.

O actual debate sobre a ontologia da “raça”¹⁰ aparece sumariamente em resposta à tese segundo a qual a “raça” é um *mito* ou uma *ilusão*¹¹. Esta proposta pode ser classificada de *anti-realista*, por oposição ao grupo de *realistas* que concebem a “raça” como *alguma coisa de real* existente no mundo (mesmo que não concordem quanto à sua natureza). A resposta *eliminativista* de que a “raça” (concebida como entidade biológica) não é real e deve ser eliminada do contexto de debate não é consensual e levanta conflitos, em particular, com a complexa teoria popular da “raça” - que partilha da ideia de que as “raças” são entidades reais. Como veremos adiante (fig.1), não existe uma única maneira de dizer que as “raças” são reais, nem uma única forma de as conceber como categorias biológicas. Na verdade, existem teóricos que estão de acordo quanto à natureza da “raça”, mas diferem quanto à sua utilização e/ou importância em determinados contextos. Em contraste com o *eliminativismo*, está a actual perspectiva do *naturalismo racial* segundo a qual, os membros de uma determinada “raça” possuem algum tipo de propriedades naturais (ou talvez até mesmo genéticas) que fazem deles aquilo que eles são e a partir das quais nós os concebemos. As pessoas estão ligadas entre si por laços de ancestralidade que as conectam a certos grupos e locais geográficos específicos, influindo no desenvolvimento de certos traços comuns. Apesar de os *naturalistas raciais* não partilharem de uma visão unívoca sobre o conceito biológico de “raça” mais apropriado (uns concebem a “raça” como *clado*, outros como *população* ou *ecótipo* e alguns como *geneticamente relevante*), todos aceitam que, independentemente daquilo que podemos dizer acerca dela, falar de “raças” é falar de uma categoria biológica real, cuja existência não depende de uma qualquer construção social. A formulação pode ser apresentada da seguinte forma: as “raças” existem, embora “de modo geral, não correspondam às categorias populares da raça”¹².

Segundo a ideia de que falar de “raças” é falar de algo bem real, o *construtivismo racial* (outra forma de realismo) – inspirado em grande medida por Du Bois – defende que a “raça” é uma “construção social, ainda que com consequências práticas”¹³: a “raça” é real como uma coisa de tipo social (TS)

¹⁰ A ontologia da “raça” tem sido recentemente desenvolvida por vários autores ligados à filosofia da linguagem; ver Peter Osborne (2002). *Philosophies of Race and Ethnicity*. Continuum International Publishing Group.

¹¹ Appiah, 1992, 1996, 1999; Zack, 2002.

¹² Pigliucci & Kaplan, 2002, 1169.

¹³ Du Bois, 1897.

– e não de tipo natural. Do mesmo modo que médicos, professores, políticos e advogados são reais porque são grupos de pessoas socialmente construídos, também os tipos raciais de pessoas são reais, mas socialmente construídos. Neste sentido, os grupos raciais são grupos reais que foram criados pelas nossas práticas sociais - não por um qualquer processo biológico - encontrando-se *isentos* de análise científica, pelo menos ao nível da sua definição. A ideia do *construtivismo social* não é uma novidade, mas no caso do estudo da “raça” pode tornar-se incómoda e radical, gerando múltiplas perspectivas sobre aquilo que podemos entender por “raça” (identidade, pertença, nacionalismo, pátria, etc.). Esta perspectiva acaba por colocar todos os teóricos numa situação de aceitação, na medida em que não existe filósofo ou cientista que resista à ideia de que “raça”, “laboratório” ou “gene” são conceitos socialmente construídos e que o contexto no qual são definidos e aplicados possuem efeitos directos naquilo que apreendemos de cada um deles. As perspectivas descritas podem ser representadas no seguinte quadro:

		ANTI-REALISMO	REALISMO	
		Eliminativismo	Construtivismo	Naturalismo
A raça é uma categoria natural?	Não		Não	Sim
A raça é real?	Não		Sim	Sim

Fig. 1 – Desacordo metafísico¹⁴

O debate entre *eliminativistas*, *construtivistas* e *naturalistas* apresenta-se, primeiramente, como um debate ontológico acerca da “raça” que vai, provavelmente, permanecer na arena durante muito tempo. A não existência de um estudo científico que determine uma divisão radical (extrema) entre os diferentes grupos raciais (genes raciais ou marcadores genéticos), permite que as perspectivas no interior destes grupos se multipliquem nas várias

¹⁴ Esta tabela foi elaborada a partir da do esquema simplificado intitulada *Fontes de Desacordo* que pode ser consultada em Haslanger, S., 2008, *A Social Constructionist Analysis of Race*, 60 (in *Revisiting Race in a Genomic Age*, edited by B. Koenig, S. S. Lee and S. S. Richardson: Rutgers University Press.).

áreas em que a questão racial é abordada (sobretudo, na sociologia, biomedicina, antropologia, filosofia e psicologia). Alguns cientistas preferem alojar-se na ideia de que a ciência irá um dia desvendar a verdade oculta sobre o ser humano e o debate em torno da *natureza* da "raça" terminará – *the truth will out device* (TWO). Enquanto outros cientistas, os quais denominamos de cépticos¹⁵, preferem deixar-se absorver pelo consenso obtido pelo *master argument* de que similitude no interior de grupos é maior do que entre diferentes grupos¹⁶, mantendo-se abertos na procura de outras explicações para a definição e uso da "raça". Perante a multiplicidade de discursos epistemológicos, alguns filósofos acabam por abandonar o terreno ontológico, solidificando a investigação em filosofia da "raça"¹⁷.

O debate sobre o conceito de "raça" enfrenta, assim, não apenas um desacordo semântico, mas um desacordo metafísico: a questão aqui é saber se a "raça" é (ou não) um tipo natural. Quando falamos de "raça" estamos a propor falar de: i) tipos científicos reais, tal como ouro ou água; ii) de tipos ilusórios, tal como fazemos quando falamos de bruxas ou pégasos; iii) ou de um tipo de elemento do mundo socialmente construído (como os universitários, os médicos) que não podem constituir-se como objecto de estudo científico?

¹⁵ Estes cépticos são uma espécie de *quasi-racialist* – termo usado por Taylor (2004) e utilizado aqui para identificar indivíduos que não aceitam o *eliminativismo* porém, aceitam a possibilidade de atribuir outro uso ou significado à "raça" atendendo ao contexto de manipulação.

¹⁶ A expressão *master argument* é comumente conhecida nas controvérsias raciais como referência ao "poderoso" argumento de Lewontin contra a existência de "raças" do ponto de vista biológico. Lewontin (1972) declara que as classificações raciais estão isentas de qualquer "significado genético ou taxonómico" baseando-se na teoria de que as diferenças registadas no interior de grupos raciais é superior a 85%, pelo que as diferenças entre grupos raciais são mínimas, segundo o autor insuficientes para falarmos de "raças"; ver Richard Lewontin, 1972, *The Apportionment of Human Diversity. Evolutionary Biology*, 6: pp. 391-398. Apesar de existirem respostas científicas a desmentir esta publicação, são muitos os biólogos, sobretudo, filósofos que fundamentam a tese da não existência de "raças" biológicas neste estudo estatístico.

¹⁷ Grande parte dos trabalhos explicitamente dedicados à filosofia da "raça" debruçam-se na *critical race theory*, desenvolvendo perspectivas muito interessantes do ponto de vista psicológico e cognitivo, refugiando em teorias da linguagem e na exploração do *reconstrucionismo* como uma solução (*anti-realista*) para a substituição radical do discurso racial (minorias, discriminação, *racial profiling*, etc.). Ver Albert Atkin, 2012, *The Philosophy of Race*, Acumen; Joshua Glasgow, 2009, *Theory of Race*, New York: Routledge.

A proposta *eliminativista racial*: “raça” como *ilusão*

There are no races: there is nothing in the world tha can do all we ask 'race' to do for us.¹⁸

Baseada no consenso de que não existem essências raciais, a teoria do *eliminativismo racial* sustenta que a “raça” não se refere a coisas existentes, sendo que a única coisa no mundo à qual o termo/conceito se poderia referir - “raças” biológicas – não existe. O argumento pressupõe que o conceito biológico de “raça” se compromete com algum tipo de essencialismo e uma vez que o mesmo não existe, a “raça” não é real. O argumento pode ser exposto da seguinte forma:

(E1) Se não há base biológica adequada para as nossas categorias raciais, então a “raça” é uma ilusão.

(E2) Não há base biológica adequada para as nossas categorias raciais.

(E3) Logo, a “raça” é uma *ilusão*¹⁹.

Segundo Appiah, a eliminação da “raça” não implica que não podemos falar sobre “raça”, significa apenas que a “raça” (sobre a qual debatemos e construímos ideias) é um conceito vazio. Será que os falantes que usam o termo “raça” sabem que estão a falar de um conceito oco? Como se explica que as pessoas falem com tanta certeza da “raça” e/ou criem discursos raciais? Segundo Appiah, a tese de que a “raça” é uma ficção remete-nos para o problema de saber se as pessoas comuns sabem de que é que falam quando empregam o conceito de “raça”.²⁰ Para resolver este problema, Appiah (inserido num discurso linguístico) sugere a exploração de dois modelos filosóficos tradicionais sobre a referência. Ao primeiro modelo chama teoria “ideativa” (*ideational*); ao segundo, teoria *referencial*. Appiah argumenta que, segundo estas tradições, ambos os termos raciais falham na tarefa de referir.

A teoria “ideativa” que Appiah descreve é muito semelhante ao que é sustentado pelas teorias *descritivistas* (Frege, Russell) – responsáveis pelo

¹⁸ Appiah, 1992, 57.

¹⁹ Argumento semelhante apresentado por Joshua Glasgow, 2009, 112-14.

²⁰ Embora as “raças” - concebidas como tipos biológicos – não sejam reais, tal não significa que não “existam” nas nossas cabeças; nós temos crenças sobre a “raça”; falamos sobre a sua estrutura e relações, tal como falamos da estrutura dos extraterrestres ou da cor dos unicórnios. Mas, não é por acreditarmos que as “raças” existem que elas se tornam efectivamente reais.

modelo explicativo internalista: i) Um termo ou conceito para um TN está a associado a uma descrição definida ou conjunto de descrições (proposições) às quais os falantes associam o termo ou conceito; ii) O termo/conceito só pode referir a uma única coisa que deve satisfazer os componentes dessa descrição; iii) se não há uma única coisa que satisfaça a descrição, o termo não se refere. Durante muito tempo, as teorias *descritivistas* foram importantes na discussão filosófica sobre a "raça", sobretudo, na investigação dos conceitos raciais. Para justificar a sua utilização, Appiah argumenta que as pessoas que utilizam o conceito de "raça" possuem um entendimento (ou noção) *comum* do termo que lhes permite a comunicação entre si:

"As pessoas que sabem o que conceito de "raça" significa, aprendem certas regras de utilização do conceito: por isso, embora as pessoas possuam diferentes crenças acerca das raças, elas partilham crenças especiais – às quais chamo de *crenças de critério* – que definem o conceito."²¹

A "raça" é, neste sentido, alguma coisa que satisfaz um *bom número de crenças de critério*, que foram adquiridas num determinado contexto, por um grupo de falantes: alguma coisa é uma "raça" apenas se as crenças de critério acerca dela são verdadeiras. Mas no meio das crenças de critério – responsáveis pela definição do conceito – podem existir algumas que não são sustentadas por todas as pessoas que utilizam o termo "raça", por exemplo, para os antropologistas físicos, a noção de "raça" classifica as pessoas num ranking de espécie, mas para os antropologistas genéticos a "raça" é utilizada para classificar a espécie humana em agrupamentos genéticos. Esta ausência de critérios unívocos torna, segundo Appiah, muito difícil argumentar contra a existência de "raças", razão pela qual parte para exploração daquilo que as pessoas pensam em relação às "raças": *como é que as "raças" devem ser para que sejam verdadeiras para todos* - se é que as "raças" podem ser alguma coisa. Em que é que as pessoas que acreditam na existência da "raça" estão a pensar quando empregam o termo? Quanto à ideia que as pessoas possuem sobre o significado da "raça", Appiah é claro e objectivo: as descrições associadas ao conceito ordinário de "raça" estão comprometidas com o *racialismo*²² – quando as pessoas falam em "raça" referem-se, geralmente, a diferenças "naturais". Assumindo que não existem tais *essências raciais*, então, para os *eliminativistas* não há "raças": as

²¹ Appiah, 1996, 52.

²² Appiah, 1996, 54.

pessoas enganam-se quando pensam que estão a falar de “raças” como se fossem entidades reais; na verdade não estão a falar de qualquer coisa de real. Esta perspectiva equivale a dizer que “raça” como TN com essências não existe, logo, não é real. Um dos erros de Appiah é privilegiar a noção ordinária de “raça” e associar a impossibilidade de desenvolvimento de um conceito científico de “raça” com a sua inexistência: nem sempre aquilo que as pessoas pensam no SC sobre os TN (mamíferos, vidro, árvores) corresponde ao seu significado científico²³. Existe uma grande dificuldade em estabelecer na linguagem do SC um rigor científico e/ou biológico aquando da utilização de conceitos de *teor científico*.

Nas últimas décadas, a preponderância da teoria descriptivista foi derrubada devido ao trabalho desenvolvido por Saul Kripke (1972) e Hilary Putman (1975). Ambos forneceram um conjunto de argumentos poderosos contra o *descriptivismo* mediante a introdução da *teoria causal da referência* – a qual Appiah intitula de teoria *referencial*. Na abordagem semântica dos termos para TN, Putman e Kripke argumentam que: i) o sentido dos termos para TN não é dado pelas suas descrições (comuns aos falantes de uma língua), mas pela sua propriedade essencial (*microscópica*); ii) os termos de TN são *designadores rígidos*: têm como referente o mesmo objecto (ou objectos) em todos os mundos possíveis em que exista um designador rígido referente (ex. Água = H₂O; Ouro = Au); os termos para TN referem-se rigidamente, “(...) o significado não pode ser identificado com a extensão.”²⁴; iii) a referência é aprendida, inicialmente, por actos de batismo e fundamentação múltipla que chegam até ao falante mediante uma cadeia causal externa. Mais importante para os actuais propósitos é que enquanto as teorias descriptivistas referem mediante uma relação de *satisfação* entre o referente e os elementos da descrição, a teoria causal refere em virtude de um vínculo causal entre o uso original do termo para identificar um *tipo* e as suas utilizações posteriores. Quando introduzido, um termo escolhe a

²³ Andreason, 1998; Dupré, 1981.

²⁴ Putman, 1973, 245. O exemplo mais famoso é, talvez, o argumento da terra gémea de Putman em que a água é entendida como uma substância com uma estrutura química diferente na terra (H₂O) e na terra gémea (XYZ). Apesar de satisfazer a descrição associada com água, Putman argumenta que XYZ não é água porque possui uma estrutura química diferente. Neste sentido, dois falantes podem associar a mesma *intensão*, mas terem diferentes extensões, “Mais uma vez vemos que o estado psicológico de um falante não determina a extensão (ou o significado) no mundo.”(Putman, 1975, 226)

propriedade ou propriedades que oferecem a melhor explicação causal das características centrais dos usos dessa palavra.

Aplicada ao caso da "raça", estas teorias sugerem a possibilidade de os termos raciais se referirem a alguma outra coisa que não essências biológicas, mesmo que as pessoas tenham uma vez acreditado que as "raças" eram caracterizadas por essências – ou ainda acreditem. Se assim for, não devemos concluir do consenso de que não há essências raciais que a "raça" não existe, mas que as "raças" existem, ainda que as pessoas tenham estado enganadas sobre aquilo que as "raças" são. Se a teoria causal da referência está correcta, devemos perguntar se existem candidatos para servir como referente para os termos ou conceitos raciais. A ideia é simples, para aceder à ideia de "raça" é necessário encontrar o referente, i. e. encontrar o objecto no mundo que está na origem da corrente causal que possui a expressão "raça": se há alguma coisa no mundo que melhor explica o debate sobre a raça, então *isso será aquilo a que a 'raça' se refere*²⁵. Appiah considera duas possibilidades e, posteriormente, rejeita-as: i) "raça" como *população*: ancestralidade comum (onde a genealogia ocupa um lugar primordial); ii) "raça" como *grupo*: padrão de características dos maiores continentes humanos – incluindo a cor da pele, tipo de cabelo e morfologia corporal.

Consideremos o referente população, termo central nos estudos de variação genética, utilizado frequentemente como eufemismo de "raça" em estudos genéticos actuais²⁶. População é, tal como a "raça", um conceito complexo em biologia (e filosofia) – existem muitos debates interessantes em torno da forma como se devem traçar os limites das comunidades de potenciais indivíduos que se cruzam entre si. Porém, Appiah não se mostra preocupado com o termo "raça" para determinar populações no domínio da genética populacional, a sua dúvida não está na existência de "raças"

²⁵ Appiah, 1996, 40.

²⁶ Face à pressão exercida pelo abandono da "raça" como categoria biológica (Montagu, 1941; Livingstone, 1962; Lewontin, 1972; UNESCO, 1950; AAPA, 1996), os investigadores biomédicos começaram a utilizar *eufemismos genéticos*, tais como população, grupo populacional, grupos continentais, geografia, etnia, grupo étnico, ancestralidade (entre outros) com o objetivo de evitar as críticas ao uso do conceito "raça" para inferir variações genéticas e acentuar a objectividade científica; ver Johnny Williams, 2011, They Say It's in the Genes: Decoding Racial Ideology in Genomics. *Journal of Contemporary Ethnography*, 40 (5): 550-581; Dani Gonzalez, Linda Suleiman and Gabriel Ivey (2010) Is There a Role for Race in Science and Medicine. 96 (9), 12-18.

biológicas *aplicadas* a determinados organismos, mas da sua extensão à espécie humana:

De facto, em muitas plantas e animais existem populações locais que são reprodutivamente isoladas de outras, diferentes na forma de agrupamento e interesse biológico, e mesmo assim, capazes de cruzamento quando colocadas artificialmente juntas; e os biólogos, tanto antes como depois de Darwin, poderiam ter-lhes chamado “raças”. Isso não acontece nos seres humanos.²⁷

A dificuldade de Appiah está na aceitação da noção de *reprodução isolada* e da sua extensão às populações humanas tendo por base o *pensamento tipológico* (essencialista). Durante muito tempo, as teorias biológicas foram dominadas pela tradição filosófica *essencialista*, segundo a qual as espécies biológicas (por ex. as plantas) eram compostas por um conjunto de características essenciais²⁸ que não eram partilhadas com outras espécies. Os essencialistas referiam-se a tais espécies como TN com essências (*kind essentialism*).²⁹ Uma vez definidas como TN, as “raças” (entendidas como subespécies), eram igualmente concebidas como TN com essências, sendo que algumas das principais taxonomias raciais foram criadas por biólogos essencialistas, tais como Linnaeus e Buffon, os quais criaram taxonomias de espécies e, posteriormente, “raças” segundo as suas essências específicas. Apesar de o *essencialismo* permanecer como uma visão muito comum entre os filósofos, são pouquíssimos os biólogos que resistem essencialistas. Após a revolução darwiniana, espécies e populações deixaram de ser vistas como tipos, classes *essencialmente* definidas passando a ser concebidas como *biopopulações* compostas por indivíduos geneticamente únicos³⁰. Em consequência, a procura de uma propriedade essencial ou traço distintivo partilhado unicamente por um único grupo de indivíduos, tornou-se, em biologia contemporânea, teoricamente supérflua. As teorias biológicas são, hoje, dominadas por uma perspectiva anti-essencialista de classificação de espécies, caracterizada como pensamento

²⁷ Appiah, 1996, 73.

²⁸ Trata-se do elemento básico (primário) do ser de uma coisa, a sua natureza; aquilo sem o qual não poderia ser o que é.

²⁹ A natureza cria diferenças entre os indivíduos e tais diferenças são “reais” ou “naturais” (por oposição a construídas), pois são a natureza essencial de uma categoria natural. Quando elementos de uma classe possuem tais essências, chamamos-lhes TP. Esta teoria será explorada na próxima secção.

³⁰ Mayr, 2002, 91.

populacional (*population thinking*)³¹. O *tipo* passou a ser considerado uma abstracção e a variação tornou-se real: as populações começam a ser caracterizadas em virtude das características comuns a toda a população, que se reproduzem livremente na natureza separadamente de outros agrupamentos, mediante a *reprodução isolada* (ou ausência de cruzamento)³². Esta *reprodução isolada* é biologicamente importante porque indica a barreira do fluxo de genes e sustenta o potencial para a diferenciação biológica entre populações. Se as populações raciais são isoladas desta forma, podem ser candidatos para os referentes de termos ou conceitos raciais, “os naturalistas identificam tais populações como subespécies ou ‘raças’”³³. Porém, este isolamento não significa, para um *populacionista*, que todos os membros daquela subespécie ou “raça” são detentores exclusivos de todas as características imaginárias e actuais atribuídas à “raça” à qual pertencem. Um dos problemas de Appiah está na incompreensão de que essas diferenças não são essenciais, também existem dentro de todas as “raças” humanas:

Uma vez na posse da figura moderna genética, é possível vermos que cada pessoa é o produto de um enorme número de características, que interagem entre si e com o ambiente e não há nada na teoria da evolução que nos garanta que um grupo que partilha certas características irá partilhá-las todas e ainda outras mais somente entre si. Características em diferentes cromossomas são, segundo os mendelianos, independentemente distribuídas.³⁴

Appiah rejeita a *reprodução humana isolada* em grupos ou populações distintas, mostrando-se céptico relativamente aos grupos contemporâneos que têm o requisito de reprodução isolada. Tanto Appiah (1996) como Zack (2002) pensam que da ausência de *reprodução isolada* entre os principais grupos raciais (tal como são identificados no discurso comum) não se segue a formação de grupos raciais no sentido de populações biológicas, pois aceitar a *reprodução isolada* em populações significa sublinhar a existência

³¹ Os *populacionistas* (ao contrário dos *tipologistas*) desenvolvem uma forma radical de olhar para a natureza: “o mundo vivo não consiste em variáveis essenciais, mas em populações altamente variáveis” (Mayr, 2002, 92).

³² Na formulação clássica de Mayr, “espécies são grupos de actual ou potencial cruzamento de populações naturais as quais são potencialmente isoladas de outros grupos semelhantes.” (Mayr, 1999, 120).

³³ Kitcher, 2007, 296.

³⁴ Appiah, 1996, 68.

de características que são partilhadas, “somente entre si” (*pensamento tipológico*) Mas, dizer que não há diferenças biológicas inatas/fixas que separam categoricamente indivíduos não é o mesmo que dizer que nenhuma diferença biológica existe ao longo desses grupos:

Aqueles que subscrevem este parecer [“não existem raças humanas”] são, obviamente, ignorantes em biologia moderna. As raças não são algo especificamente humano; as “raças” ocorrem numa grande percentagem de espécies de animais. Pode ler-se em qualquer livro sobre a evolução que quando raças geográficas de animais são isolados de outras raças da sua espécie, podem tornar-se, no devido tempo, novas espécies.³⁵

Apesar da controvérsia entre cientistas relativamente ao uso da “raça” como variável em estudos científicos, sobretudo no acordo em relação ao que se entende por “raça” (ancestralidade, parentesco, linhagem, etc.) no que respeita a variabilidade genética, existe consenso de que grupos progressivamente isolados tendem a desenvolver e partilhar características diferenciadoras de outros grupos igualmente isolados (a controvérsia gira em torno da importância quantitativa e qualitativa dessas diferenças para as nomearmos “raças”)³⁶:

Se raças biológicas são definidas pela singularidade, em seguida, há um número muito grande de raças biológicas que podem ser definidas, contrariando o conceito clássico de raça biológica da antropologia física. Além disso, os nossos resultados mostram que os seres humanos podem ser precisamente classificados em origem geográfica usando craniometria mesmo que haja sobreposição entre os grupos.³⁷

Geneticistas populacionais e antropologistas físicos demonstram que um entendimento tipológico das diferenças biológicas é insustentável perante resultados que descrevem os padrões de variação genética no interior e entre grupos de populações humanas. As “raças” não são entendidas como

³⁵ Mayr, 2002: 89.

³⁶ Contemporaneamente, são muitos os artigos científicos que abordam as diferenças genéticas em populações humanas, acentuando a possibilidade de tratar grupos humanos como grupos populacionais. Porém, esta situação não é consensual. O debate gira em torno de saber se essas diferenças entre grupos são suficientemente relevantes para as utilizarmos como variável em estudos biomédicos, ver Silvia Bittencourt and Sandra Caponi, 2009, *O Uso da Categoria "raça" na Biomedicina. Cadernos de Pesquisa Interdisciplinar em Ciências Humanas*, 10(97): pp. 290-311; Josué Iaguardia, 2004, *O Uso da Variável “Raça” na Pesquisa em Saúde. PHYSIS: Rev. Saúde Coletiva*, 14(2): pp. 197-234.

³⁷ Ousley, 2009, 68.

categorias biológicas, no sentido de estarem divididas categoricamente na natureza (inatas/fixas). No entanto, nenhum geneticista deixa de subscrever que a variação entre seres humanos permanece. As diferenças podem ser superficiais (*superficial kinds*) mas, não deixam de estar presentes na natureza e de possuir informação genética acerca das origens dessas populações. Appiah e Zack demonstram, assim, ignorância ao sustentar a sua argumentação na ideia filosófica arcaica de que: i) as "raças" são constituídas por indivíduos que partilham a mesma essência; e que; ii) cada "raça" é continuamente descontinuada de todas as outras³⁸. Refugiando-se no *ontological statement* de que não há base biológica para a "raça" e servindo-se do *master argument* segundo o qual as diferenças no interior dos grupos são maiores do que as diferenças entre grupos, Appiah ignora a base epistemológica da crença científica continuada na "raça", crença que não faz sentido segundo o modelo de "raças" como categorias taxionómicas discretas mas, pode ter significado em estudos probabilísticos noutros contextos empíricos (estudos geográficos, ancestralidade, epidemiologia). Na actual investigação da biologia humana, a validação operacional do conceito biológico de "raça" como uma medida aproximada da distância genética em populações humanas tem vindo a revelar-se funcional na construção informacional.³⁹ Quando o termo "raça" é utilizado em aspectos de origens geográficas e identidade racial pode revelar-se útil na previsão das diferenças genéticas entre os grupos.

Appiah (ao contrário de Zack) reconhece que há populações que estão e têm estado durante algum espaço de tempo reprodutivamente isoladas. Porém, não é plausível dar o salto lógico (em terminologia de Appiah, semântico) para a consideração que subgrupos reconhecidos socialmente podem ser considerados tais populações: "(...) existem raças humanas, porque existem populações humanas, no sentido genético, mas não há um grande grupo social na América a que podemos chamar raça."⁴⁰ Appiah oferece aquilo a que podemos chamar de argumento de *incompatibilidade (extencional)*⁴¹. Um argumento de *incompatibilidade* sustenta que a

³⁸ Zack, 2002, 63.

³⁹ Baker, 1967, 21.

⁴⁰ Appiah, 1996, 73.

⁴¹ Appiah pensa que estes argumentos de *incompatibilidade* mostram que o referente proposto viola aquilo a que chama "condição de adequação", a qual requer que alguma coisa que é pensada como sendo verdadeira acerca do que é denotado por um termo, tem de estar, pelo menos, aproximada da verdade do candidato a

verdadeira extensão de um termo ou conceito x seria drasticamente diferente do que é acreditado sobre a extensão de x. Se nós permitíssemos que o termo “raça” determinasse populações biológicas, daí não se seguiria que grupos considerados no senso comum como “raças” (tal como “brancos” ou “pretos”) seriam “raças”. Em contrapartida, outros grupos que não são primeiramente pensados como “raças” (Appiah sugere *Amish*⁴²) poderiam contar como “raças”. Dado que tal incompatibilidade iria frustrar as intenções comuns que regem a utilização dos termos raciais, um argumento de incompatibilidade poderia apoiar o abandono ou uso de tais termos. Existe uma incompatibilidade explícita entre o candidato a referencial (conceito biológico) e a visão do senso comum da “raça” (diferenças raciais essenciais).

Quanto ao candidato referencial *grupo*, Appiah reforça a ideia de que não existindo uma completa substituição de genes entre populações (os estudos genéticos geralmente referem-se apenas a algumas características específicas), há uma partilha consideravelmente maior de similitudes do que diferenças. Logo, duas pessoas pertencentes a uma população podem diferir tanto geneticamente como um indivíduo escolhido à sorte em qualquer parte do mundo. Acrescentando o facto de que a diferença genética nem sempre corresponde à diferença morfológica, tornar-se-ia, segundo o autor, muito difícil a identificação das pessoas pela sua identidade racial, necessária à constituição do grupo:

(...) um grande número de pessoas [nos EUA] não encaixa nesta categoria, porque são produtos de misturas (por vezes, de longa data) entre pessoas que apenas grosseiramente se encaixam nesse padrão, mesmo que as distinções sociais a que chamamos de “raciais” nos Estados Unidos cubram, por contraste, quase toda a gente.⁴³

A utilização de tal noção biológica ficaria presa em substratos genéticos, fisiológicos tais como cor da pele ou morfologia, os quais possuem uma relação muito ténue com as características que são consideradas importantes para a vida moral e social. A reivindicação de Appiah é que enquanto o uso comum do termo implica uma importância social e

referencial. (Appiah, 1996, 40).

⁴² Grupo religioso (cristão) que procura viver isolado da sociedade moderna e dos seus confortos (automóveis, telefones, etc.); são conhecidos pelos seus costumes conservadores e por recriarem o modo de vida rural do séc. XVII.

⁴³ Appiah, 1996, 74.

psicológica nos grupos a que esses termos são aplicados, não há razão para esperar que o conceito biológico de população ou o conceito de grupo baseado na observação de um padrão dominante de características possuam este género de importância. Appiah não está inconsciente que grupos socialmente reconhecidos como "raças" podem ser distinguidos ao nível genético. A ignorância do reconhecimento da dimensão biológica da "raça" não decorre de uma consideração inadequada ou incompleta do consenso científico mas, de uma decisão deliberada para negar a importância das práticas científicas no desacordo com a conclusão empírica preferida.⁴⁴ Appiah acaba por utilizar essas diferenças como base para acreditar na realidade ontológica previamente estabelecida como semanticamente e filosoficamente injustificada.

O acordo ontológico de que as "raças" não são entidades reais não está estabelecido em biologia. Esta vontade filosófica de negar *a priori* a possibilidade de debater as possibilidades do conceito de "raça" em ciência, em prol da igualdade e da extinção do racismo, sem a realização de um estudo aprofundado sobre aquilo que vai sendo determinado como "raça" em genética, evidencia o seu distanciamento da presente prática científica. Independentemente do acordo ontológico instituído em filosofia em relação à inexistência de "raças" biológicas, "raça" e biologia permanecem ligadas, a diferenciação biológica é tomada como garantida⁴⁵ e o conceito de população parece ser o melhor candidato a referencial:

Os dados de muitas fontes indicam que os seres humanos são geneticamente homogéneos e que a variação genética tende a ser compartilhada amplamente entre as populações. As variações genéticas geograficamente estruturadas, como esperado a partir do isolamento parcial das populações humanas durante a maior parte da sua história. Uma vez que os conceitos tradicionais

⁴⁴ Fullerton, 2007, 250.

⁴⁵ Ver Michael Campbell and Sarah Tishoff, 2010, *The Evolution of Human Genetic and Phenotypic Variation in Africa*. *Current biology : CB*, 20(4): R166-R173; Armand Leroi, 2005, *A Family Tree in Every Gene*. *NY Times*, A22-23; Neil, Risch, Esteban Burchard, Elad Zib, et. al, 2002, *Categorization of humans in biomedical research: genes, race and disease*. *Genome Biology*, 3 (7). Stephan, Ousley, Richard Jantz, Sonna Freid, 2009, *Understanding Race and Human Variation: Why Forensic Anthropologists are Good at Identifying Race*. *American Journal of Physical Anthropology*, 139: 68-76; Duana Fullwiley, 2007, *Race and Genetics: Attempts to Define the Relationship*. *BioSocieties*, 2: 221-237.

de raça estão, por sua vez, relacionados com a geografia, é inexato afirmar que a raça é "biologicamente sem sentido".⁴⁶

A objecção do *eliminitivismo racial* baseia-se, assim, na ideia de que existe uma incompatibilidade ou desfasamento entre o que os falantes comuns esperam dos conceitos raciais e aquilo que eles recebem deles⁴⁷ – desacordo *extensional* entre o uso científico e o uso no SC dos termos de TN (neste caso é dísparo). Appiah conclui que das tradições apresentadas (e dos candidatos a referente), *you can't get much of a race concept*⁴⁸. Ambos podem ser candidatos do referencial da noção de sentido, mas nenhum deles corresponde aos grupos sociais a que chamamos "raças". Esta ausência de correspondência leva Appiah a rejeitar a possibilidade de construção de uma noção de "raça", eliminando-a do discurso comum. Simultaneamente, esta distância de correspondência possibilita que o conceito de "raça" continue a ser explorado em contexto científico. Será que o desacordo *extensional* (entre ciência e CS) e a associação equívoca do conceito de "raça" no SC ao *racialismo* são motivos suficientes para desistirmos do conceito de "raça"? E, se não, pode o conceito de "raça" ter pertinência biológica sem estar comprometido com o essencialismo? Os passos seguintes vão procurar mostrar que podemos ser realistas em relação à "raça", sem que para isso tenhamos de ser essencialistas ou *racialistas*.

O mito das essências raciais

Provavelmente nós tendemos a pensar as raças como essencialmente diferentes apenas porque queremos desculpar ou justificar a dominação de uma raça por outra.⁴⁹

Em primeiro lugar, perguntar acerca da natureza das "raça" é, apesar das críticas dos geneticistas, fazer uma pergunta filosófica e biológica. Quando falamos de "raças" estamos a falar de tipos naturais ou de indivíduos? Será

⁴⁶ Jorde, 2004: S29.

⁴⁷ Glasgow, 2003, 461. Glasgow discute o desacordo *extensional* entre o uso dos temos para TN em ciência e no SC, alertando para o facto de o conceito biológico de "raça" não encaixar no uso ordinário do termo: "aquilo que no meio científico se concebe como sendo *raça* não se adequa ao número de classificações reconhecidas no senso comum" (Glasgow, 2003, 467).

⁴⁸ Appiah, 1999, 276.

⁴⁹ Hacking, 2005, 103.

que as “raças” se referem a uma categoria na natureza ou apenas a um mecanismo útil para organizar o nosso conhecimento biológico? Para oferecer uma resposta completa será necessário não apenas recorrer a informação empírica e a teorias biológicas, como a elementos metafísicos e epistemológicos. Em segundo, perguntar pela natureza da “raça” significa abordar dois problemas conceptuais (similares ao problema da natureza das espécies): i) estatuto ontológico da “raça”: tal como no problema das espécies, existem filósofos e biólogos que defendem a existência de “raças” como TN, refugiando-se no *kind essentialism* – encontrado no trabalho filosófico de Kripke (1972) e Putman (1975), cujas origens remontam a Aristóteles e Locke; e no trabalho biológico de Linnaeus e Buffon; enquanto outros filósofos e praticamente todos os biólogos rejeitam a presença de essências, refugiando-se na teoria da estrutura populacional segundo a qual a “raça” é uma entidade genealógica, uma linhagem de populações – Mayr (1999), Gannet (2002) e o filósofo Kitcher (1999) são alguns representantes; ii) o modelo de classificação taxionómica pluralista: quer a “raça” como a espécie foram ao longo da sua história definidas de diversas formas e incluídas em múltiplas classificações; existem filósofos e biólogos que defendem a existência de apenas uma classificação correcta do mundo orgânico (monistas); em oposição, os pluralistas consideram que o mundo é pluralista, aceitando mais do que uma classificação – Kitcher (1984), Ereshefky (1992) e Dupré (1993) defendem que o pluralismo nas espécies deve ser adoptado por razões ontológicas; Kitcher, juntamente com Andreasen (2000, 2004) estendem esse pluralismo ao conceito de “raça”.

Historicamente, a “raça” (como TN esteve associada à crença de que as raças possuem algum tipo de *essências raciais*, aquilo a que Appiah chama de *racialismo*⁵⁰: crença que os membros de uma particular “raça” partilham

⁵⁰ Importa distinguir *racialismo* de *racismo*. *Racialismo* é o nome de uma teoria ou família de teorias acerca de espécies humanas e as suas divisões; está ligado à ideia filosófica do essencialismo através da qual as coisas que se parecem semelhantes entre si são acreditadas como partilhando propriedades semelhantes (da mesma natureza). O *racismo* é, por contraste, uma atitude afectiva e voluntária exercida sobre indivíduos e grupos, pressupondo normalmente o reconhecimento da desigualdade: i) existem diferenças essenciais (fixas) entre grupos humanos (a que chamamos “raças”); ii) tais grupos estão hierarquicamente classificados segundo a natureza (do superior ao inferior); iii) os grupos superiores podem exercer domínio sobre os inferiores. Ser racista não implica ser *racialista*, do mesmo modo que um *racialista* não está comprometido com a doutrina do racismo; para uma abordagem detalhada ver Moore, J. H. (2008) *Encyclopedia of Race and Racism*. Farmington Hills, MI: Thomson Gale, 183-184.

certas tendências e características imutáveis comuns entre si (hereditárias, físicas, psicológicas, intelectuais e culturais) que não partilham com os membros de outras “raças”:

Podemos dividir os seres humanos num pequeno número de grupos, chamados de "raças", de tal forma que os membros desses grupos compartilhavam certas características fundamentais, hereditárias, físicas, morais, intelectuais e culturais que não compartilhavam com os membros de qualquer outra raça.⁵¹

Durante muito tempo a “raça” (tal como a espécie) foi acreditada como sendo algum tipo de essência *bio-comportamental*, possuindo um conjunto de propriedades essenciais (naturais e, talvez até genéticas). O sistema de transmissão racial apoiava-se na “crença falsa de que existiam propriedades genéticas, um gene da ‘raça’ ou grupo de genes”⁵² que individuavam as “raças”, explicando um grande número de alegadas diferenças raciais no comportamento, inteligência e morfologia dos indivíduos. A *crença falsa* de que as “raças” eram uma *subespécie*⁵³ demarcada pela natureza estava associada à tradicional visão filosófica que concebe as espécies biológicas como TN com essências (*tipological thinking*)⁵⁴:

um tipo natural é uma classe de entidades que partilham um tipo específico de essência; tais essências capturam a estrutura fundamental do mundo; ou usando a frase de Platão, *esculpem a natureza nas suas articulações*.⁵⁵

Dizer que uma “raça” é um TN, significa dizer que é composta por propriedades naturais ou “reais”, i. e. que pode constituir-se como uma categoria ou classificação taxonómica em que determinados objectos podem ser agrupados com base em características “reais” (propriedades essenciais) comuns aos membros de um certo tipo natural – e não uma consequência das marcas (ou atribuições) através das quais os homens as agrupam em classes. Podemos encontrar TN no vocabulário comum (ex. cavalo, abóbora)

⁵¹ Appiah, 1996, 81.

⁵² Root, 2000, S634.

⁵³Uma *subespécie* é: 1) uma subdivisão de espécies taxonomicamente reconhecida; 2) definidas geograficamente e/ou ecologicamente com características distintas; ver Mulligan & Stansfield King, 2006.

⁵⁴O problema da definição de “raça” encontra-se em estreita relação com o problema da definição das espécies.

⁵⁵ Ereshefsky, 2009, 256.

e em quase todas as disciplinas científicas, nomeadamente, em biologia e química (ex. H₂O, Au, *Drosophila melanogaster*).

As propriedades naturais estão, neste sentido, demarcadas pela natureza, mas o seu reconhecimento e consequente classificação são um acto do homem. Deste modo, as propriedades podem até ser “reais”, mas as classificações taxonómicas são sempre “artificiais” (construídas), por isso, ao longo da história biológica das “raças”, somos confrontados com um pluralismo de sistemas de taxonomia “raciais”⁵⁶. Na prática, o critério para a classificação das pessoas em “raças” procedia como se existissem propriedades genéticas singulares. A classificação como branco, asiático ou preto baseava-se genericamente em três crenças, sendo que as duas primeiras são falsas e apenas a última é verdadeira: i) as pessoas de diferentes “raças” possuem diferentes *genes da raça* que são partilhados apenas (exclusivamente) entre os membros de uma mesma “raça”; ii) as propriedades genéticas da “raça” explicam as predisposições comportamentais, culturais e intelectuais dos indivíduos e grupos raciais; iii) os genes são transmitidos hierarquicamente (geneticamente) ao longo de gerações entre pais e filhos biológicos. De facto, a transmissão genética só é possível entre indivíduos que partilham laços de sangue mas, a razão da sua relevância encontra-se nas duas primeiras (falsas) – crença que existem

⁵⁶ Linnaeus, por exemplo, no seu *Systema Naturae* (1735) divide (pela primeira vez) o ser humano em quatro “raças”: 1) americanos vermelhos; 2) europeus brancos; 3) asiáticos negros; 4) negros pretos; (1749); Buffon (1749) na famosa *História Natural* introduziu o termo *subespécie* (posteriormente associado a Linnaeus) e, embora reconhecesse (ao contrário de Linnaeus) que as diferenças eram *transitórias*, encarando as categorias como meras *conveniências*, apresentou uma lista não permanente de diferentes *tipos* de homem (Molnar, 2006, 6): 1) Polar; 2) Tártaros (mongol); 3) Asiáticos do Sul; 4) Europeus; 5) Etiópes; 6) Americanos; Blumenbach (1776) introduz o termo caucasiano e reconhece cinco variedades principais na humanidade: 1) caucasiano (*white race*); 2) Mongol (*yellow race*); 3) Malário (*brown race*); 4) Etiópe (*black race*); 5) American (*red race*); Meiners (1795) com base em estudos geográficos determina a existência de apenas duas “raças”: 1) bem-parecidos (celtas, sarmatianos, asiáticos); 2) feios (negros). Voltaire reconhece sete, Kant apenas quatro (branco; negro; mongol e Hindu). Até meados do séc. XX a genética mendeliana tornou a “raça” concebida como essência ainda mais sustentável, podendo o número de “raças” crescer de duas (Virey), até onze (Pickering), vinte e duas (Morton) ou mesmo sessenta (Crawford). Uma das principais críticas ao conceito biológico de “raça” é a aparente inexistência de critério (unanimidade): não há um número de “raças” preciso em ciência ou uma definição (universal) que satisfaça os grupos raciais comuns; ver Glasgow, J. M. (2003) *On the new biology of race*. *Journal of Philosophy*, 100, 456-474; ver Laguardia (2005), Raça, evolução humana e as (in)certezas da genética. *Antropo*, 9, 13-27.

essências que delimitam os grupos raciais e as suas habilidades físicas e intelectuais. O erro desta visão está em pensar que as “raças” como TN possuem “essências genéticas” (*race-as-essence*). A procura de um traço essencial partilhado somente por um grupo e concebido como propriedade essencial que joga um papel fundamental na ocorrência de outras propriedades nos membros desse tipo tornou-se uma tarefa impraticável e “biologicamente desnecessária”⁵⁷. Para explicar a variação não é necessário recorrer a essências (singularidade de um traço numa determinada espécie)⁵⁸. Os recentes evolucionistas citam a frequência de genes dentro das populações e as forças evolutivas que afectam as frequências que ocorrem na próxima geração. A variação elevada é explicada recorrendo à citação das frequências de gene relevantes numa determinada geração, mais a ocorrência de tais forças evolutivas, como a selecção ou adaptação.

Hoje, contrariamente ao que muitas pessoas pensam no SC, existe um consenso entre os naturalistas de que as “raças” (mesmo concebidas como TN) não partilham de tais essências e que as diferenças entre os indivíduos (físicas, intelectuais e psicológicas) estão “menos na natureza das coisas do que estavam antigamente”⁵⁹. O pensamento tipológico é substituído, tal como no estudo das espécies, pelo *pensamento populacional*. Devido à facilidade de viajar e contactar com outros povos, os grupos raciais têm vindo a reproduzir-se entre si, misturando grande parte do seu material genético, fazendo com que as tradicionais distinções raciais (Branco, Negro e Asiático) se tornem menos visíveis, podendo mesmo vir a extinguir-se (num futuro longínquo). Há cada vez mais pessoas com uma ascendência “racial” múltipla, devido ao cruzamento entre pais provenientes de diferentes culturas e grupos raciais. A mistura racial parece ser o futuro e a elevada variação genética no interior dos grupos (*ingroups*) é, talvez, uma das melhores provas⁶⁰. Porém, da similitude e mistura existente entre os vários grupos

⁵⁷ Ereshefsky, 2009.

⁵⁸ O estudo da “raça” pode ser dividido, segundo Marks (2008), em *dois paradigmas*: 1) até 1930 (genocídio): a “raça” concebida como sub-espécie com propriedades essenciais; 2) após o genocídio: a ideia de “grupos” parece ser insustentável e a “raça” é entregue aos antropologistas (factos sociais vs factos naturais), o termo permanece o mesmo (“raça”), mas a concepção subjacente é radicalmente reformulada” (Marks, 2008, 24), passando a ser estudada em termos de similitudes no *pool genético* (código de similitudes genéticas – ancestralidade).

⁵⁹ Hacking, 2005, 104.

⁶⁰ O anúncio oficial da compilação do projecto do genoma humano em 2001, confirmou que as sequências genéticas humanas são 99,9% idênticas, sendo que

populacionais humanos, não parece seguir-se a afirmação de Appiah segundo a qual a semelhança genética entre dois indivíduos pertencente a um mesmo grupo populacional é igual a um indivíduo escolhido à sorte numa população em qualquer parte do mundo. Segundo Risch et. al. (2002) esta afirmação é incorrecta, pois não possui validade empírica, “dois caucasianos são mais similares entre si geneticamente que um caucasiano e um asiático”⁶¹, caso contrário, seria impossível criar agrupamentos discretos de humanos que acabam por corresponder às principais “raças”:

Na generalidade, os indivíduos de diferentes populações possuem, em média, maior dissemelhança do que outros pertencentes ao seu grupo populacional (a variação *within* é maior do que a diferença *between* (...) Populações humanas são muito similares, mas podem ser frequentemente distinguidas.⁶²

Portanto, do consenso que “não há genes para a raça”⁶³, nem marcadores genéticos de diferenças físicas e comportamentais, não se segue (como é declarado pelos *eliminativistas*) que não existem diferenças significativas entre as “raças”. Apesar da diferença entre os grupos ser *superficial* (por oposição à essencial), é significativa, podendo revelar-se útil em investigação ancestral e biomédica. Como veremos adiante, “abandonar o realismo forte acerca dos tipos não significa desistir completamente do realismo”⁶⁴, é possível a “raça” ser real, sem que para isso esteja a ser comprometida com as essências. Em estudos biomédicos modernos, a “raça” não é assumida como uma entidade essencial ou demarcadora fixa de traços genéticos, mas como regulador de hereditariedade útil para o prognóstico e associação de grupos de risco ao desenvolvimento de doenças genéticas e no aprofundamento das origens ancestrais:

apenas 3% a 10% (desses 0.1% de variação) está associado com a ancestralidade geográfica: “as diferenças entre os maiores grupos constituem-se em 3% a 5%”, (ROSENBERG et al., 2002); para mais informações ver Satel, S. “Race and Medicine Can Mix Without Prejudice: How the Story of BiDil Illuminates the Future of Medicine”, in *Medical Progress Today* (December, 2004).

⁶¹ Risch et. al., 2002, 5.

⁶² Bamshad, 2003, 78.

⁶³ Dupré, 2008.

⁶⁴ Kitcher, 2007, 299.

⁶⁵ Na última publicação sobre a “raça”, Kitcher (2007) reafirma a sua teoria da “raça” como população e declara que o argumento *eliminativista* parte de uma premissa errada acerca dos TN (com essências). Kitcher não está seguro das essências nos TN e considera possível defender um tipo de realismo que não negue a possibilidade de classificações múltiplas, ao qual chama de “realismo pluralista”, semelhante ao *realismo promiscuo* desenvolvido por Dupré (1981).

A estrutura das populações humanas é relevante em vários contextos epidemiológicos. Como resultado da variação de frequências de ambos os factores de risco, genéticos e não genéticos, as taxas de doenças desses fenótipos como resposta adversa ao medicamento varia entre populações. Além disso, as informações sobre a população de origem do paciente podem fornecer aos profissionais de saúde, informações sobre risco quando causas diretas da doença são desconhecidas.⁶⁶

Uma proposta realista do conceito de “raça”

As aproximações *eliminativistas* falharam em reconhecer formas mais sutis em que as divisões em raças poderiam ter significado biológico.⁶⁷

Se as “raças” não existem como é que se explicam as diferenças fenótypicas entre membros de diferentes grupos humanos? Como se explica que é possível agrupar as populações mediante a frequência de alelos? Existe maior variação no interior dos grupos, do que entre grupos, será isso suficiente para não pensarmos mais na diferença entre grupos? Não será a própria variabilidade genética a prova da existência de diferença racial? Na sua tese, Appiah parece confundir racismo com racialismo, sacrificando a legitimidade histórica e biológica do conceito de “raça”. Nesta proposta realista procurar-se-á mostrar como é possível conceber a “raça” como uma realidade biológica sem que para isso tenhamos de cair num essencialismo ou naquilo a que Appiah chama de *racialismo*. Para os defensores do *naturalismo racial* a visão naturalista da “raça” não está comprometida com o *racialismo*: a “raça” possui legitimidade histórica, biológica e social uma vez que “nós dividimo-nos onde a natureza não o faz”⁶⁸. A perspectiva realista sobre a “raça” não é consensual, sendo amplamente discutida no meio científico e filosófico devido ao pluralismo de significações pelas quais é conduzido.

Para os defensores do naturalismo racial a “raça” não se encontra inscrita na natureza, porém, isso não significa que a “raça” não seja visível aos nossos olhos⁶⁹. Segundo Kitcher e Andreasen – principais mentores da perspectiva do *naturalismo racial* – a “raça” pode ser concebida como uma

⁶⁶ Rosenberg, Pritchard, Weber, et. al., 2002, 2384.

⁶⁷ Kitcher, 1999, 87.

⁶⁸ Root, 2000, S630.

⁶⁹ Hardimon, 2003, 441.

população biológica caracterizada por um certo nível de reprodução isolada. A imagem que Andreasen e Kitcher desenvolvem da "raça" é relativamente satisfatória de um ponto de vista científico, mas não é eficaz para compreendermos aquilo que no SC se pensa ou estabelece como sendo "raça". Conceber a "raça" como população biológica implica aceitar a existência de algum tipo de reprodução isolada durante a evolução do ser humano, a qual pode ou não perpetuar-se hoje – ideia rejeitada por Zack e Appiah. Considerando as polémicas envolvidas em torno da possibilidade de desenvolvimento de um conceito biológico de "raça", sobretudo, em torno da fomentação de diferenças sociais e políticas, convém esclarecer que as perspectivas apresentadas não constituem um argumento favorável à desigualdade entre os indivíduos, distinguindo-se do racismo⁷⁰, quer pelos conceitos envolvidos como pela natureza das investigações.

Em contraste com a negação da existência de "raças", Pigliucci e Kaplan (2003) num artigo intitulado *On the Concept of Biological Race and Its Applicability to Humans*, sugerem a aplicação do termo *ecotype* (ecótipo) para designar "raça". Os ecótipos designam a existência de um grupo de indivíduos ("raça" no interior de uma espécie) geneticamente adaptados a um certo ambiente: *populações locais* adaptadas a determinados ambientes ou regiões, que expressam diferentes características. Os ecótipos podem ser diferenciados com base em algumas diferenças genéticas, dado que existe uma conexão entre diferenciação genética e adaptação ecológica, mas tal não implica necessariamente uma distinção filogenética⁷¹. Segundo estes autores, a maioria dos argumentos favoráveis à eliminação da "raça" ou à conclusão da sua inexistência baseiam-se numa interpretação errada dos ecótipos. A evolução humana é marcada pelo *fluxo génico*⁷² (*extensive gene flow*) mas desta transferência e mistura de genes entre os diferentes ecótipos não se pode simplesmente concluir que não existem "raças" biológicas significativas nos humanos, como alguns reivindicam. Embora a variação

⁷⁰ Ver nota 11 sobre a definição de racismo.

⁷¹ Pigliucci & Kaplan, 2003: 1163.

⁷² O fluxo de génico é também conhecido por migração de genes (transferências de genes de uma população para outra) para dentro ou para fora de uma população que pode ser responsável por uma mudança acentuada nas frequências alélicas (a proporção de membros que exercem uma variante específica de um gene). A imigração pode também resultar na adição de novas variantes genéticas do *pool* genético estabelecido de uma determinada espécie ou população.

genética no ser humano seja relativamente pequena⁷³, as diferenças genéticas adaptativas podem ser mantidas entre populações através da selecção natural mesmo quando há fluxo genético significativo entre as populações⁷⁴, razão pela qual faz sentido aplicar o termo ecótipo ao ser humano. Para Kaplan e Pigliucci há uma evidência que as “raças” humanas biológicas existem, mesmo que tal descrição não corresponda ao número de “raças” populares partilhadas no SC – *desacordo extensional* reclamado por Glasgow. Falar de “raças” é tão natural como falar de genes ou células, a única diferença está na precisão da sua diferenciação e no uso que delas pretendemos fazer. Apesar de parecer demasiado óvia para alguns, a proposta de Kaplan e Pigliucci não é isolada, mas também não é amplamente partilhada pelos círculos académicos. Na maioria das vezes não é devidamente esclarecido que a redução da variabilidade nos humanos e da sua evolução não pretende oferecer mais do que uma via para salvaguardar a existência única e rica dos seres humanos. A ideia contrária de uma sobrevalorização de um grupo ou população está associada a movimentos específicos que envolvem teor diverso.

Uma perspectiva semelhante sobre o conceito de “raça” é desenvolvida por Kitcher (1999), Andreasen (2000) e, posteriormente, por Hardimon (2003). Esta abordagem irá centrar-se particularmente em Kitcher e Andreasen com o objectivo de mostrar que a noção de “raça” como população biológica não parece diferir abissalmente do conceito ordinário de “raça” (COR) – tal como é reclamado por Glasgow – embora não exista uma correspondência satisfatória com o número de “raças” reclamadas pelo SC. Apoiando-se na noção de populações biológicas, Andreasen defende a tese de que nós podemos entender as populações humanas como divisões em subespécies de linhas cladísticas, desenvolvendo aquilo a que se chama de *conceito cladístico de raça* (CCR). Para compreendermos e aceitarmos o CCR é necessário comprometermo-nos com o método de classificação cladística⁷⁵. Um cladista organiza as populações biológicas numa árvore

⁷³ As recentes publicações sobre a variação genética defendem a inexistência de variação significativa no interior da espécie humana, razão pela qual não há pertinência biológica na sua divisão em populações.

⁷⁴ Pigliucci & Kaplan, 2003: 1164.

⁷⁵ O cladismo é uma das maiores escolas de taxonomia biológica do séc. XX fundada por Willi Henning (1966). O cladismo é um método de classificação que tenta desvendar as relações genealógicas entre as duas espécies, salientando apenas as características avançadas que compartilham. A representação gráfica dos resultados de tal análise é o cladograma; ver em King, Robert C., Mulligan, Pamela K., and

filogenética, indicando apenas o ancestral comum e as relações descendentes. Tradicionalmente, as árvores filogenéticas são construídas com pontos de ramificação, representando eventos de especiação. Na maioria das vezes, as ramificações envolvem a divisão de uma população em duas populações filhas, exibindo isolamento reprodutivo:

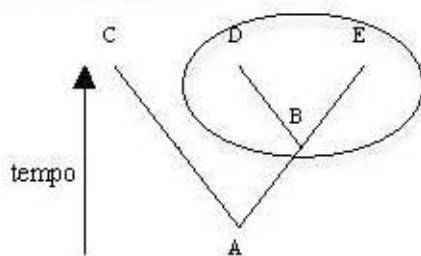


Figura 2. Cladograma

Os cladistas identificam na árvore espécies com segmentos monofiléticos (ver na figura 2: BD, BE). Andreasen estende a tradicional taxonomia cladística ao nível da subespécie, argumentando que um cladograma pode ser construído seguindo as relações geneológicas das populações humanas. Cada ramo da árvore filogenética representa as relações evolutivas entre a reprodução isolada das populações humanas. Para Andreasen, as “raças” podem ser definidas como grupos monofiléticos de tais populações⁷⁶. Isto implica que cada população tem de estar reprodutivamente isolada numa porção significativa da história evolutiva antes de poder ser designada uma “raça” cladística. Andreasen acredita que as “raças” humanas existiram no passado com ramificações claramente isoladas, mas que as mesmas dificilmente podem ser identificadas com as actuais populações nos E.U.A. devido à mistura racial das populações inicialmente isoladas conduzindo inevitavelmente à dissolução gradual da distinção racial: “Eu sustento que as raças existiram uma vez, mas podem estar a caminho de se tornarem extintas.”⁷⁷

Stansfield, William D. (2006), *A Dictionary of Genetics* (7th ed.: Oxford University Press), 86.

⁷⁶ Andreasen, 2007:95

⁷⁷ Andreasen, 2005:96

Kitcher, tal como Andreasen, argumenta que as “raças” podem ser definidas, em parte, por populações reprodutivamente isoladas, porém não requer a existência de grupos monofiléticos, revelando-se mais optimista quanto à existência actual de “raças”. Ao contrário da maioria dos investigadores raciais, Kitcher não está interessado nos vários conceitos de “raça” desenvolvidos ao longo dos tempos, mas no modo como os “biólogos empregam a noção de “raça” e como podemos considerar a nossa espécie de uma forma semelhante”⁷⁸. Enquanto para Andreasen as populações raciais podem existir como *clados*⁷⁹ ou segmentos monofiléticos de uma árvore filogenética que reconstrói a história evolutiva da humanidade (ou seja, a genealogia é suficiente para definir “raças”).

Kitcher utiliza a noção de “raça” pura⁸⁰ (ou população fundadora) para argumentar que a divisão racial só possui interesse biológico se satisfizer duas condições: i) os membros das raças puras possuírem um fenótipo distinto ou propriedades genéticas distintas; ii) as populações “misturadas” são relativamente pequenas, pelo menos, ao longo das gerações entre a origem e o presente⁸¹.

Kitcher defende um conceito de “raça” *minimalista*: a “raça” não é nada mais do que uma “linhagem pura em que a endogamia⁸² pode ter sido inicialmente resultado de um isolamento geográfico que eventualmente deu origem a diferenças no fenótipo e a alguma interferência no cruzamento livre.⁸³ A noção de “raça” que Kitcher sugere não é morfológica, mas baseada nos padrões de reprodução: a possibilidade da divisão racial é específica para um conjunto de população presente numa certa região geográfica específica⁸⁴. Kitcher comprehende as “raças” como um certo tipo

⁷⁸ Kitcher, 1999, 95.

⁷⁹ Em estudos evolutivos, um *taxon* ou outro grupo constituído por uma única espécie e os seus descendentes; um conjunto de espécies representando um ramo distinto em uma árvore filogenética. Graficamente um clado inclui a espécie representada pelo nó e todos os ramos que brotam da mesma; ver em King, Robert C. , Mulligan, Pamela K., and Stansfield, William D. (2006), *A Dictionary of Genetics* (7th ed.: Oxford University Press), 86.

⁸⁰ Esta noção de “raça” pura não possui nenhum tipo de conotação negativa, refere-se à ideia de “linhagens completamente inatas”, grupos de indivíduos com características “originais”, i.e. que não foram misturadas.

⁸¹ Kitcher, 1999, 95.

⁸² Cruzamento entre plantas ou animais relativamente próximos.

⁸³ Kitcher, 2007, 296.

⁸⁴ kitcher, 1999, 99.

de linhagens populacionais reprodutivamente isoladas⁸⁵, sendo possível, numa determinada região geográfica existirem dois grupos racialmente divididos, mesmo que noutros sítios não estejam: “quando as raças se desenvolvem juntamente, as diferenças no acasalamento na intra-racialidade e na inter-racialidade são suficientes para sustentar traços diferentes que marcam as raças”⁸⁶. Para Kitcher, o conceito de “raça” humana pode ter um significado biológico, uma vez que existem diferenças na frequência dos genes que podem ser preservados devido à baixa probabilidade de reprodução inter-racial: “quando as linhagens entram em contacto, as taxas de casamento inter-linhagem permanecem significativamente inferiores às taxas do casamento intra-linhagem”.⁸⁷⁸⁸ Eis um esquema exemplificativo da teoria de Kitcher:

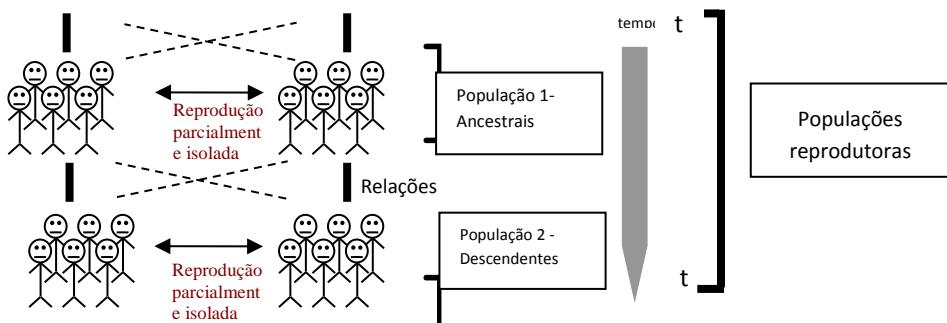


Figura 3. Populações no tempo

Kitcher opõe-se, assim, ao *eliminativismo* na medida em que considera que a divisão racial compreende o padrão da reprodução, “podemos tomar a raça como sendo remota, procurando o tempo da sua origem, o período geográfico de separação ou podemos supor que o processo de formação da

⁸⁵ Reprodução isolada - zonas híbridas (*hybrid zones*): espaço onde duas espécies se encontram e produzem espécies híbridas; são estáveis, porque os genes de uma espécie não fluem para as outras.

⁸⁶ Kitcher, 1999, 97.

⁸⁷ Kitcher, 2007, 297.

⁸⁸ Kitcher recorre a um vasto conjunto de dados relativos à percentagem de casamentos interraciais nos E.U.A. verificando que a taxa dos casamentos interraciais é muito inferior à taxa dos casamentos intra-raciais.

raça começou no tempo da reconstrução.»⁸⁹. Kitcher e Andreasen acreditam que os grupos populacionais que propõem e discutem podem desviar-se de forma importante da visão do SC, tal como acontece com muitos outros conceitos científicos (mamífero; vidro, etc.)⁹⁰ – desacordo extensional. Provavelmente, nenhum deles acredita que tais grupos populacionais possuam uma explicação tão importante como o SC lhes atribui. Aquilo que realmente importa a estes *naturalistas raciais* é saber se tais grupos populacionais podem ser apropriadamente chamados de “raças”, ou seja, se existem diferenças biológicas relevantes entre as populações que tornem legítima a divisão racial.

O conceito de raça encontra-se intrinsecamente relacionado com a etnicidade, não sendo possível afirmar que a “raça” não tem qualquer relação com a “cultura” ou a sociedade humana: a “raça” não está inscrita na natureza, “nós dividimo-nos onde a natureza não o faz”⁹¹ e não o fazemos à sorte, mas com objectivos. Independentemente das “raças” serem ou não uma realidade biológica, as pessoas pertencem a diferentes “raças” na medida em que nós reconhecemos algumas características físicas diferentes como sendo relevantes para contar como um grupo “racial”, em detrimento de outras⁹²⁹³. As diferentes características existentes entre grupos de populações são, tal como afirmam os elminativistas, *superficiais*⁹⁴ na medida em que se opõem às essenciais, mas reflectem distinções genéticas razão pela qual são também suficientemente “salientes”⁹⁵ para os seres humanos as captarem a “olho nu” e estabelecerem divisões. Esta visibilidade das características é importante, pois a “raça” é “essencialmente manifesta”⁹⁶.

A ideia de “raça” é, assim, a ideia de um grupo com um fenótipo distinto – estrutura facial, formato do nariz, cor da pele, textura do cabelo, forma corporal – sendo o que muitas vezes concebemos como COR:

⁸⁹ Kitcher, 1999, 102.

⁹⁰ No uso de termos para tipos biológicos, em nenhuma circunstância foi exigido à ciência que alterasse ou reajustasse as suas definições para que encaixassem no uso do SC; ver (Andreasen, 2005).

⁹¹ Root, 2000, S630.

⁹² Hardimon, 2003, 442.

⁹³ O conceito de “raça” tem legitimidade biológica, mas isso não significa que os conceitos actuais de raça se baseiam primeiro em considerações biológicas.

⁹⁴ Hacking, 2005, 102.

⁹⁵ Kitcher, 2007, 303.

⁹⁶ Hardimon, 2005, 443.

Toda a gente concorda que os seres humanos diferem entre as dimensões mencionadas. Estas são as diferenças que os *eliminativistas* se esforçam para não negar. E toda a gente concorda que estas diferenças são raciais – ou, em qualquer caso, que são o tipo de diferenças a que chamamos racial.⁹⁷

Ao sustentar que não existem “raças”, alguns eliminativistas também pretendem negar a existência de diferenças raciais entre os grupos de populações que são genuinamente raciais – *no race, no racial difference*. Mas parece possível chamar a certas propriedades de “raciais” sem que para isso tenha de se supor que a “raça” é real – as bruxas possuem poderes ocultos, independentemente da sua existência; os unicórnios possuem um chifre em espiral independentemente de serem ou não reais. Aliás, para filósofos como Appiah (não há “raças”) e Zack não parece difícil falar de classes de pessoas que partilham certas características físicas observáveis – as mais notáveis como a cor da pele – e que nós (humanos) somos hábeis a reconhecê-las. No entanto, a distância entre aceitar a existência de talas características em certos grupos e a *ideia* de que os membros dos grupos que as exibem são raciais (as “raças” existem) é tênue. Talvez esta seja a principal razão pela qual os *eliminativistas* mostram relutância em reconhecer que os seres humanos possuem características que podem ser apropriadamente chamadas de raciais.

Se é verdade que existe um mecanismo que permite o parcial isolamento de grupos humanos (reprodução isolada) e se nós somos capazes de observar e reconhecer características fenótipas diferentes entre certos grupos humanos, então a “raça” é real, podendo ser definida como uma população mendeliana (geográfica) que difere na distribuição da frequência dos genes no *pool genético*. Estas diferenças aplicam-se apenas ao nível da população em termos de relativa frequência de características entre populações. Posteriormente, as etiquetas que lhes colocamos e o uso que lhes destinamos estão relacionados com a nossa consciência histórica colectiva de “raça”, com a taxonomia biológica em uso e com a utilidade (fim) que essa mesma divisão nos pode trazer para o melhoramento da vida do homem. Em relação a esta segunda e muito importante tarefa de delineamento e aplicação do conceito muito pouco foi ainda feito, uma vez que este debate só pode ser equacionado quando se ultrapassa o consenso estipulado e partilhado pela maioria dos teóricos oriundos da filosofia, sociologia e biologia de que a “raça” não existe, situação que conduz este

⁹⁷ Hardimon, 2003, 444.

trabalho para um domínio puramente social e antropológico, deixando que as relações entre “raça” e biologia sejam utilizadas de forma livre.

Conclusão: o que podemos esperar do debate sobre conceito de “raça”?

Races do exist, and to say it again, race matters. The questions, rather, must be about the nature of race and then about what, if anything, we can or should do to improve things.⁹⁸

Demasiado centrados na possibilidade da “raça” como entidade biológica, os *eliminativistas raciais*: i) não fornecem uma resposta satisfatória sobre aquilo as “raças” são; ii) julgam que ao eliminar a “raça” toda a carga política e social conflituosa que esta acarreta desaparece; iii) impossibilitam o estudo de um dos conceitos mais violentos da história humana, assim como a possibilidade de descodificar o ser humano nas suas profundezas. Retomando a ideia inicial de que falar de “raças” é falar de algo real, sendo possível desenvolver uma conversação não “oca” entre indivíduos que partilhem o uso do conceito, afigura-se necessário perguntar aos realistas o queremos que a “raça” seja para nós, que realidade racial desejamos. Considerando que as nossas categorias de “raça” são mais fruto da nossa escolha do que da nossa natureza, mas que “não são escolhidas da mesma forma que podemos escolher uma *password* ou um lugar num autocarro vazio”,⁹⁹ importa questionar o que esperamos que a “raça” faça por nós. A sua eliminação não promove consequências imediatas no esquecimento do racismo ou na supressão das diferenças raciais, tal como os críticos raciais gostam de advogar, mas o seu uso e manutenção no discurso podem trazer resultados reais (e directos) para os indivíduos. Afinal, a rejeição da base biológica da “raça” foi unicamente a rejeição de um modelo de diferenciação racial e não a repudião da existência de distinções biológicas (genéticas) entre grupos de indivíduos socialmente reconhecidos como formando diferentes “raças”.¹⁰⁰ Constatando que a investigação genética populacional é hoje uma das áreas de maior entusiasmo na investigação biológica e as

⁹⁸ Ruse, 2006, 185.

⁹⁹ Root, 2000.

¹⁰⁰ Fullerton, 2007, 250.

diferenças estatísticas parecem estar presentes em todos os artigos, os filósofos e outros teóricos deveriam envolver-se criticamente nas práticas científicas e clínicas contribuindo com diferentes formas de reflexão da investigação empírica, procurando circunscrever a informação biológica racial no seu domínio de aplicação. Em vez de se dedicarem a procurar formas alternativas de velhos argumentos, na sua maioria fechados em construções biológicas arcaicas de tipos naturais fixos e imutáveis, desajustados às novidades da literatura genética. Que vantagens podem ter as categorias raciais? Poderão ajudar-nos a diminuir os danos e a obter um conhecimento mais vasto da variabilidade do ser humano? Por que razão tem vindo a ser realizada uma desconstrução massiva do conceito biológico de "raça"? Como é que a teoria da negação das "raças" se tornou comumente aceite? A resposta a estas e outras perguntas da mesma índole deverá ser o objectivo central do actual *conflito* (sobretudo, entre os profissionais da biomedicina) sobre a "raça" repensar a sua solidez e utilização. Ignorar a existência de uma base biológica ou genética na "raça" é ignorar o contínuo interesse científico, assim como as implicações sociais e biomédicas inerentes a este tipo de investigação. Os filósofos e outros críticos deviam dedicar-se ao esclarecimento do conceito, na averiguação do seu uso no contexto científico actual, sobretudo, na sua reconstrução numa ciência cada vez mais tecnológica.

Referências Bibliográficas

- Andreasen, R. O., 2000, Race: Biological Reality or Social Construct. *Philosophy of Science*, 67, 653-66.
- 2005, The Meaning of Race: Folk Conceptions and the New Biology of Race. *The Journal of Philosophy, Inc.*, 102 (2), 94-106.
- 2004. The cladistic race concept: a defense. *Biology and Philosophy*, 19, 425-442.
- Appiah, A. K. 1992, The Uncompleted Argument: Du Bois and the Illusion of Race. In: "Race", *Writing, and Difference*, University of Chicago Press Journals.
- 1996, Race, Culture, Identity: Misunderstood Connections. In: *Colour Conscious: The Political Morality of Race*, New Jersey, Princeton University Press, 30-74.
- 1999, Why There Are No Races. In: *Racism - Key Concepts in Critical Theory*. New York, Humanity Books, 267-77.
- Bamshad, M. J. O., Steve E., 2003, Does Race Exist?. *Scientific American*, 289, 78-85.
- Banton, M.; Haerdwood, J., 1975, *The Race Concept*, n. p., Praeger.
- Banton, M., 2010, *A Ideia de Raça*, Lisboa, Edições 70.
- Baker, P. T., 1967, The biological race concept as a research tool. *American Journal of Physical Anthropology*, 27 (1), 21-25.
- Bernasconi, R. and Lott, T. L., 2000, *The Idea of Race*, Hackett Publishing Company, Inc..
- Du Bois, W. E. B., 1897, The Conservation of Races. *The American Negro Academy Occasional Papers*, 2.
- Bhopal, R., 2003, Glossary of terms relating to ethnicity and race: for reflection and debate. *J. Epidemiol Community Health*, 58, 441-445.
- Dupré, J., 1981, Natural Kinds and Biological Taxa. *The Philosophical Review*, 90 (1): 66-90.
- 2008, What Genes Are and Why There Are No Genes For Race. In: *Revisiting Race in a Genomic Age*, Rutgers University Press, 39-55.
- Ereshefsky, M., 2009, Species, Taxonomy, and Systematics. In: *Philosophy of Biology*, Oxford, Wiley-Blackwell, 256-271.
- Fullerton, S. M., 2007, On the Absence of Biology in Philosophical Considerations of Race. In: Sullivan, S. and Tuana, N. (Eds.), *Race and Epistemologies of Ignorance*. New York, State University of New York Press.
- Glasgow, J. M., 2003, On the new biology of race. *Journal of Philosophy*, 100 (9), 456-474.
- 2006, A Third Way in the Race Debate. *The Journal of Political Philosophy*, 14 (2), 163-185.
- 2009. *Theory of Race*, New York, Routledge.
- Hacking, I., 2005, Why race still matters. *Dædalus Winter*, 102-16.

- Hardimon, Michael O., 2003, The Ordinary Concept of Race. *The Journal of Philosophy*, 100 (9), 437-55.
- Haslanger, S., 2008, A Social Constructionist Analysis of Race. In: *Revisiting Race in a Genomic Age*, Rutgers University Press, 56-69.
- Jones, R. A., 2005, Race and Revisability. *Journal of Black Studies*, 35 (5), 612-32.
- Jorde, B. L. W., Stephen P., 2004, Genetic Variation, Classification and Race. *Nature Genetics*, 36 (11), S28-S33.
- Kaplan, J.; Pigliucci, M., 2002, On the Concept of Biological Race and Its Applicability to Humans. In: *Philosophy of Science Assoc. 18th Biennial Mtg* (PSA 2002: Contributed Papers; Milwaukee, WI).
- King, R. C., Mulligan P. K.; Stansfield, W. D. (Eds.), 2006, *A Dictionary of Genetics*, 7th ed., Oxford University Press.
- Kitcher, P., 1999, Race, Ethnicity, Biology and Culture. In: *Racism - Key Concepts in Critical Theory*, New York, Humanity Books, 87-116.
- 2007, Does Race Have a Future?. *Philosophy & Public Affairs*, 35 (4), 293-317.
- Marks, J., 2008, Past, Present, and Future. In: *Revisiting Race in a genomic age*, Rutgers University Press, 21-38.
- Mayr, E., 1999. *Systematics and the Origin of Species*, n. p., Harvard University Press.
- 2002, *What Evolution Is*, Phoenix 350.
- 2002, The Biology of Race and the Concept of Equality. *Daedalus*, 231 (1), 89-98.
- Mills, C., 1999, The Racial Polity. In: *Racism and Philosophy*, Campbell University Press, 13-31.
- Montagu, A., 1997, *Man's Most Dangerous Myth*, Walnut Creek, Altamira Press.
- Mosley, A., 1999, Negritude, Nationalism and Nativism: Racists or Racialists? In: Leonard Harris (ed.), *Racism - Key Concepts in Critical Theory*, Humanity Books, 74-86.
- Ousley, S., Jantz R. and Freid D., 2009, Understanding Race and Human Variation: Why Forensic Anthropologists are Good at Identifying Race. *American Journal of Physical Anthropology*, 139, 68-76.
- Outlaw, L. T. J., 1996, Against The Grain Of Modernity - The Politics of Different and the Conservation of Race. In: *On Race and Philosophy*. New York, Routledge, 134-157.
- Pigliucci, M., Kaplan, Jonathan, 2003, On The Concept of Race and Its Applicability to Humans. *Philosophy of Science*, 70, 1161-1172.
- Reid, Jasper, 2002, Natural Kind Essentialism. *Australasian Journal of Philosophy*, 80 (1), 62-74.
- Risch, N., Burchard E., Ziv E., et al., 2002, Categorization of humans in biomedical research: genes, race and disease. *Genome Biology*, 3 (7).

- Root, M., 2000, How We Divide the World. *Philosophy of Science*, 67 (3), 628-639.
- Rosenberg, Noah A., et al., 2002, Genetic Structure of Human Populations. *Science*, 298 (5602), 2381-85.
- Ruse, Michael, 1996, *Darwin and Its Discontents*, New York, Cambridge University Press.
- Taylor, P. C., 2004, *Race - A Philosophical Introduction*, Cambridge, Polity Press.
- Zack, Naomi, 2002, *Philosophy on Science and Race*, Routledge.
-- 2003, Race and Philosophy of Meaning. In: Bernard Boxill (ed.), *Race and Racism*, Oxford University Press, 43-57.

O problema mente-corpo na primeira década do século XXI: visita guiada a pontos-chave da paisagem fisicista

António Zilhão

(Universidade de Lisboa - Departamento de Filosofia)

AntonioZilhao@letras.ulisboa.pt

1. Uma evolução surpreendente

Quem, como eu, tenha seguido com grande interesse o debate em torno do chamado ‘problema mente-corpo’ nos anos 80 e 90 do século passado, considerando todos os argumentos com um espírito aberto, mas apostando fortemente no lado materialista da disputa, não pode deixar de sentir-se surpreendido quando contempla o modo como esta discussão evoluiu ao longo da primeira década do século XXI.

A causa maior desta surpresa é, sem dúvida, a verificação de que o principal púlpito materialista foi, nestes últimos anos, reocupado pela teoria da identidade tipo-tipo. Como suponho que vos lembrareis, esta teoria era, até há bem pouco tempo, sistematicamente retratada, sobretudo no interior do próprio campo materialista, como uma forma ingénua, ultrapassada e pouco sofisticada de fisicismo. Outro factor indutor de surpresa é a constatação de que a teoria da identidade tipo-tipo efectuou essa reocupação vestindo a pele de um recém-chegado altamente sofisticado. Finalmente, uma terceira causa de surpresa é a descoberta de que, entre os principais proponentes deste remoçado ponto de vista teórico, se encontram filósofos que se lhe opuseram fortemente na segunda metade do século XX. Entre estes avulta o nome de Ned Block. Não creio ser necessário lembrar aqui que Block ganhou as suas credenciais teóricas no debate em Filosofia da Mente como defensor convicto do funcionalismo, o qual ele alegava ser o

único modo de combinar adequadamente a salvaguarda de intuições materialistas de base com a necessidade de preservar alguma autonomia para a Psicologia no campo teórico.

Que aconteceu? Para tentar alcançar alguma compreensão a este respeito, procurei encontrar uma resposta clara às duas perguntas seguintes:

Surgiram argumentos positivos realmente novos capazes de lançar uma luz inesperada sobre a teoria da identidade tipo-tipo, luz essa que, por sua vez, teria o condão de realçar aspectos anteriormente insuspeitados da dita teoria?

Encontraram-se realmente novas formas de responder convincentemente às objecções contra os seus amplamente conhecidos pontos fracos?

Por estranho que isso possa parecer-vos, a conclusão da investigação que empreendi, e cujo sumário aqui vos apresento, foi a de que a resposta a qualquer uma destas duas perguntas é negativa.

Pensareis, provavelmente, que estou a deixar-me levar pelo exagero retórico. Mas não é esse o caso. Para tentar convencer-vos disso mesmo, peço então a vossa indulgência e sugiro que me sigais naquilo que pode ser descrito como uma ‘visita guiada pela paisagem fisicista contemporânea’.

2. De volta à casa de partida

Como prelúdio às perplexidades conceptuais com que me deparei nesta investigação, deixai-me mencionar o supracitado exemplo dos pontos de vista mais recentes de Ned Block. No dealbar deste novo milénio, Block apareceu a defender (juntamente com Stalnaker) a posição de acordo com a qual os estados sensoriais deveriam ser identificados com tipos de estados neurais. Mais especificamente, ele argumentou que os diferentes tipos de experiência consciente deveriam ser identificados com tipos de padrão de actividade nas células piramidais do cérebro, no seguimento dos trabalhos recentes de Crick e Koch em Neurofisiologia.

Como justifica Block esta sua nova perspectiva? Uma análise atenta ao seu trabalho mostra que ele o faz recorrendo a um argumento que exibe a estrutura de uma inferência para a melhor explicação. Devo confessar que achei isto estranho. Com efeito, interroguei-me, como pode uma inferência para a melhor explicação propor-se resolver o problema metafísico tradicionalmente associado à identificação de propriedades mentais com um

tipo particular de propriedades físicas? Neste passo, deixai-me lembrar-vos que, a despeito da sua alegada ingenuidade, os primeiros teóricos da identidade tipo-tipo, como, por exemplo, J. J. Smart, sempre fizeram questão de sublinhar que nenhum argumento empírico poderia determinar a nossa escolha por ou contra qualquer uma das teorias filosóficas rivais que se propunham responder ao problema mente-corpo. Em coerência com este diagnóstico, as principais considerações introduzidas por Smart para sustentar o seu próprio ponto de vista no debate foram considerações de simplicidade ontológica, e não inferências para a melhor explicação.

Deixai-me contextualizar um pouco o problema, de modo a que seja possível dar-lhe uma forma mais precisa. Dentro do campo materialista, era costume pensar-se que havia um argumento, de contornos claramente anti-materialistas, ao qual não era possível escapar. Estou a falar do argumento modal que Kripke desenvolveu no seu famoso livro *Naming and Necessity*. A forma que a sabedoria materialista do século XX descobriu de responder ao desafio posto pelo argumento de Kripke foi a de negar que a adesão ao materialismo tivesse que implicar a visão estritamente reducionista da mente e do mental, típica da teoria da identidade tipo-tipo, e que constituía o alvo primeiro do seu argumento. Dizia-se desta que se tratava apenas de uma primeira tentativa de procurar uma perspectiva materialista consistente da mente e do mental e que, enquanto primeira tentativa algo rude, ela era, de facto, vulnerável às acusações de Kripke. Todavia – dizia-se também – havia surgido, entretanto, outra abordagem materialista da mente e do mental – a abordagem funcionalista – que era mais sofisticada que a teoria da identidade tipo-tipo, e que interpretava de um modo diferente desta a relação que supostamente se verificaria entre estados mentais e estados físicos. Mais em particular, o funcionalismo caracterizava-se por defender que os termos para estados mentais deveriam ser essencialmente interpretados como descrições de papéis causais ou, em alternativa, descrições de estados funcionais de ordem superior; ora, tanto num caso como outro, os estados mentais referidos por tais termos seriam, pelo menos em princípio, instanciáveis numa variada gama de substratos físicos. E esta interpretação do problema mente-corpo, sendo ainda materialista, mantinha-se imune ao ataque de Kripke.

Com o benefício da retrospecção, sabemos hoje que o funcionalismo tinha o seu lote próprio de problemas, bem difíceis de resolver; um deles, e não o menor deles, era, precisamente, o de descobrir como integrar o aspecto

qualitativo das experiências subjectivas conscientes na definição funcional dos estados mentais a elas associados. Como decerto sabeis, este problema veio a revelar-se pouco menos que intratável. Uma consequência desta intratabilidade foi o surgimento da ameaça de que o aspecto propriamente sentiente da nossa vida mental iria ter que ficar de fora do quadro teórico-explicativo. Esta ameaça adquiriu uma faceta particularmente vívida nas chamadas ‘objecções dos zumbis’ propostas por Chalmers. Por meio das mesmas, pretendia-se sublinhar, de um modo intuitivamente apelativo, o facto de que o ponto de vista funcionalista não tinha quaisquer meios conceptuais à sua disposição em termos dos quais pudesse diferenciar entre nós e a nossa vida mental, por um lado, e criaturas zombiescas e a sua vida mental, por outro lado, as quais seriam, por hipótese, funcionalmente semelhantes a nós, mas destituídas de qualquer vida consciente. As teorias funcionalistas da mente e do mental eram assim acusadas de serem inúteis, tanto de um ponto de vista explicativo, uma vez que seriam cegas ao aspecto mais notório que a vida mental teria para nós, humanos, como de um ponto de vista metafísico, uma vez que se revelariam incapazes de integrar no âmbito da perspectiva materialista um importante segmento da realidade.

Não tenho quaisquer dúvidas de que era necessário e urgente fazer algo para encontrar um modo satisfatório de sair deste beco teórico. Foi, todavia, com grande surpresa que me apercebi que alguns dos investigadores de maior nomeada nesta área estavam a responder ao problema retirando simplesmente na direcção das velhas teorias da identidade tipo-tipo. Creio que a minha surpresa se deixa facilmente justificar quando, após um momento de reflexão, constatamos que, se a solução que se propõe para o problema é essa retirada, então isso significa que estamos, simplesmente, de volta ao ponto de partida inicial!

Só por si, o sentimento de circularidade associado a esta retirada é já suficientemente desagradável. Mas, mais importante do que esse ou quaisquer outros sentimentos, é o facto de que, se somos trazidos de volta ao ponto onde começámos, então temos que voltar a enfrentar os problemas que nos fizeram de lá sair. E o mais proeminente de entre eles é, precisamente, o desafio de Kripke às versões do materialismo baseadas na teoria da identidade tipo-tipo. Parece-me claro, todavia, que não se pode responder ao desafio de Kripke recorrendo a argumentos estruturados como inferências para a melhor explicação. Com efeito, o que está em jogo no argumento de Kripke é a própria possibilidade de que as teorias da

identidade tipo-tipo possam ter sentido. E é claro que uma inferência para a melhor explicação tem que permanecer sem sentido se a teoria que estamos a tentar mostrar ser a melhor forma de explicar os factos for, ela própria, destituída de sentido.

3. O primeiro obstáculo

Para que o ponto que mencionei fique completamente claro, creio que necessito de introduzir mais algum detalhe. Assim sendo, deixai-me passar em revista, com alguma minúcia, tanto o argumento de Kripke como o seu alvo materialista.

De acordo com os proponentes originais das teorias da identidade tipo-tipo nos anos 50 do século passado – Feigl, Place e Smart – a identidade tipo-tipo que supostamente se verificaria entre estados mentais e estados neurofisiológicos deveria ser concebida à imagem e semelhança das identidades teóricas já descobertas pela ciência, tais como, por exemplo, as identidades ‘água=H₂O’ ou ‘calor=energia cinética molecular média’. Deste modo, a dor, por exemplo, seria identificada teoricamente com o disparo das chamadas fibras-C no corpo do sofredor.

A substância do argumento anti-materialista de Kripke consiste na alegação de que uma tal analogia é espúria. O seu argumento é, em traços gerais, o seguinte. Apesar de as identidades teóricas descobertas pela ciência, como as mencionadas acima, serem *a posteriori*, elas têm que ser necessárias, caso sejam verdadeiras. Esta contenção deixa-se derivar, por sua vez, da contenção de que é suposto que os termos que ocupam ambos os lados destas identidades sejam, de acordo com o vocabulário de Kripke, designadores rígidos. Ora, se dois designadores rígidos designam a mesma coisa, então a asserção de identidade que os relaciona afirma a identidade dessa coisa consigo própria. E uma tal asserção de identidade é, obviamente, necessária. Portanto, se a água é, de facto, H₂O, então é-o necessariamente; e se o calor é, de facto, energia cinética molecular média, então é-o necessariamente. Assim sendo, se a analogia entre dor, calor e água é para valer, então a dor tem que ser necessariamente idêntica ao disparo das fibras-C. Todavia, alega Kripke, não faz qualquer sentido sustentar que a alegada correlação que se verifica existir entre dores e disparo das fibras-C seja de carácter necessário. Se existe realmente uma

correlação entre a experiência da dor e o disparo das fibras-C do sofredor, essa correlação é contingente, como é o caso, por exemplo, da correlação causal; mas uma tal correlação contingente não configura um caso de identidade. Portanto, ao correlacionar a dor com o disparo das fibras-C, não se alcançou qualquer identificação teórica, e não se produziu qualquer redução das propriedades mentais a propriedades físicas.

Para que não sobrem quaisquer dúvidas acerca da natureza deste argumento, creio ser necessário clarificar um pouco melhor a razão pela qual a correlação da dor com o disparo das fibras-C, caso seja verdadeira, não pode, segundo Kripke, ser necessária. Basicamente, o que Kripke alega a este respeito é que nós temos intuições poderosas de acordo com as quais as dores poderiam ter estado associadas em nós a muitos outros mecanismos neurofisiológicos em vez de terem ficado associadas ao disparo de fibras-C e que o disparo das fibras-C poderia ter estado associado em nós à experiência de estados mentais completamente diferentes do de sentir dor em vez de ter ficado associado à experiência deste estado. Estas intuições mostram-nos, então, que a correlação actual entre a experiência das dores e o disparo das fibras-C, caso seja verdadeira, é uma correlação perfeitamente contingente, originada pelo modo particular como os organismos vivos complexos evoluíram no planeta Terra. Ora, e a despeito do que parecem ser inclinações em sentido contrário, Kripke alega que tais intuições não ocorrem nos casos em que se descobrem genuínas identidades científicas.

Creio que é necessário introduzir aqui ainda mais uma clarificação. Na realidade, e contra a alegação de Kripke, não parece ser de todo óbvio que intuições deste género não se encontrem disponíveis nos casos de identificação teórica da água com H₂O ou do calor com a energia cinética molecular média. Com efeito, e para pegar só num destes dois exemplos, parece ser claramente o caso que nós temos a intuição de que, num outro mundo possível, aquilo a que, neste mundo, chamamos ‘calor’ poderia ter estado correlacionado com algum outro fenômeno físico que não o aumento da energia cinética média das moléculas do meio envolvente, e que algum outro estado, a que, neste mundo, damos outro nome, poderia estar, nesse outro mundo possível, correlacionado com esse aumento. Esta intuição de contingência parece ser obviamente sensata. Por que é, então, que ela não conta como evidência a favor da tese de que as identidades teóricas da

ciência seriam igualmente contingentes? Onde reside a diferença, se alguma existe, entre os dois tipos de intuição?

Deixai-me recordar-vos a resposta de Kripke a esta questão. Ela é basicamente a seguinte. No caso das identidades teóricas reais, alcançadas pela ciência, a descoberta da relação de identidade que vigora entre os referentes dos seus termos é mediada por estados cognitivos do sujeito, os quais, apesar de apenas contingentemente associados com eles, são distintos tanto do elemento designador como do elemento designado indicados no lado esquerdo dessas identidades teóricas. Assim, a descoberta, neste mundo, da identidade que se verifica existir entre o calor e a energia cinética molecular média é mediada pelas nossas sensações de calor. E a descoberta da identidade que se verifica existir entre a água e a substância química H₂O é igualmente mediada, neste mundo, pelas nossas experiências de perceber um líquido incolor, insípido e inodoro em rios, lagos e tanques, de bebê-lo, tomar banho nele, etc. A contingência que a intuição mencionada no parágrafo anterior revela não é, então, uma contingência que se encontraria associada a uma relação de identidade que vigoraria entre uma substância ou fenômeno natural e ele próprio. A contingência que ela revela é, antes, a contingência da conexão que verificamos existir entre um termo (e.g., ‘calor’) e o mediador semântico, de carácter cognitivo, por meio do qual a referência do mesmo é fixada. Acontece, porém, que, no caso da alegada identidade que se verifica existir entre a dor e o disparo das fibras-C do sofredor, tais intermediários cognitivos não se encontram disponíveis.

Ao tornar-nos conscientes desta distinção, Kripke leva-nos então a tomar consciência de que, quando dizemos, para exprimir o que parece ser uma intuição válida, que, num outro mundo possível, o calor poderia ser algo de diferente da energia cinética molecular média, esse é um modo inapropriado de nos exprimirmos. Na realidade, o modo apropriado de exprimir a intuição válida que subjaz a essa expressão deficiente seria o de dizer que, nesse mundo possível, a sensação que nós, neste mundo, estamos habituados a associar com o calor, ou com a energia cinética molecular média, e que exprimimos por meio da expressão ‘está calor’, não se encontraria, nesse outro mundo, correlacionada com o calor ou, o que é o mesmo, com a energia cinética molecular média.

Na realidade, as coisas podem ficar ainda mais confusas se supuermos que os habitantes desse outro mundo possível usariam a palavra ‘calor’ para referir o fenômeno que, nesse mundo possível, se encontraria

constantemente correlacionado com a sensação que nós, neste mundo, exprimimos por meio da expressão ‘está calor’. Mas, como Kripke faz questão de insistir, o modo adequado de descrever esse estado de coisas seria o de dizer que, nesse outro mundo possível, a palavra ‘calor’ seria um designador rígido para um fenómeno físico que não o calor, fenómeno físico esse que, nesse mundo, se encontraria causalmente correlacionado com aquela sensação que nós, neste mundo, exprimimos por meio da expressão ‘está calor’, e que o designador rígido para o calor seria a palavra que, nesse outro mundo, seria usada em associação com a sensação que, nesse mundo, se encontraria causalmente correlacionada com o calor e que, neste mundo, se encontraria correlacionada com algum outro fenómeno físico que não o calor.

Resumindo o conjunto do argumento, então, o quadro geral que obtemos é o seguinte. No caso da alegada identificação entre a dor e o disparo das fibras-C temos uma intuição válida de acordo com a qual, mesmo assumindo que é um facto empírico que há uma correlação perfeita e constante entre a nossa experiência de episódios de dor e picos de actividade neural nas nossas fibras-C, a sensação a que chamamos ‘dor’ poderia, noutro mundo possível, ter estado correlacionada do mesmo modo perfeito e constante com outro fenómeno neurofisiológico que não o disparo das fibras-C. E temos também uma intuição válida de que o disparo das fibras-C poderia ter estado correlacionado, nesse outro mundo possível, com outra sensação que não a dor. As intuições aparentemente semelhantes a esta que temos no caso das identificações teóricas genuínas efectuadas pela ciência são, na realidade, dissemelhantes dela. A distinção crucial que se verifica entre estes dois tipos de intuição é a seguinte. As intuições de contingência associadas com as identificações teóricas válidas produzidas pela ciência são, na realidade, descrições enganadas de intuições válidas associadas ao carácter contingente do modo por meio do qual a referência dos designadores rígidos que figuram no lado esquerdo dessas expressões de identidade foi historicamente fixada. No caso das intuições acerca do carácter contingente da correlação que se verifica existir entre a dor e o disparo das fibras-C, essa contingência não é ilusória ou aparente, porque a origem da intuição não pode ser reconduzida à presença de um qualquer mediador contingente no estabelecimento da alegada identidade. Segundo Kripke, dor é o que o termo ‘dor’ significa. Não há qualquer mediador interno ou externo entre a palavra e o fenómeno mundano. Na realidade, não há qualquer fenómeno mundano: a

palavra designa rigidamente a própria sensação em vez de designar o que quer que seja de mundano com o que ela possa encontrar-se correlacionada (danos corporais, por exemplo). Todavia, se houvesse alguma identidade teórica que verdadeiramente relacionasse a dor com o disparo das fibras-C, uma tal identidade teria que ser necessária; por conseguinte, não poderia ter sido possível tanto poder-se conceber a dor como estando correlacionada com qualquer outro fenómeno neurofisiológico além do disparo das fibras-C como poder-se conceber o disparo das fibras-C como estando correlacionado com qualquer outro estado mental além da dor.

Ao fazer-nos seguir este percurso dialéctico, Kripke pretende então ter-nos mostrado duas coisas.

Primeiro, que a contenção, feita pelos teóricos originais da identidade tipo-tipo, de acordo com a qual a relação que se verificaría entre o calor e a energia cinética molecular média, ou entre a água e H₂O, seria uma identidade contingente, válida apenas no nosso mundo, em vez de uma identidade necessária, válida em todos os mundos possíveis, falha o alvo. Não há, de todo, identidades contingentes.

Segundo, que, dada a ausência de mediadores cognitivos contingentes no caso da descoberta da existência de uma correlação constante entre a dor e o disparo das fibras-C, não se coloca sequer a questão de ter sido feito um diagnóstico errado na avaliação original da intuição acima mencionada a respeito da natureza da conexão que se verificaría existir entre estes dois *relata*. Uma tal conexão é, de facto, contingente. E, como não existem quaisquer identidades contingentes, a correlação em causa não pode ser a da identidade.

Sendo este o caso, uma contradição emerge entre a intuição válida que temos do carácter contingente da correlação que se verifica existir entre a dor e o disparo das fibras-C e a tese da identidade teórica, a qual implica que uma tal correlação teria que ser necessária. Assim sendo, uma destas opções tem que ser deixada cair. De acordo com o ponto de vista de Kripke, a opção que devemos deixar cair tem que ser a segunda, uma vez que ninguém foi ainda capaz de esboçar ou, sequer, dar a entrever, um qualquer bom argumento que substanciasse a invalidade da nossa intuição original.

4. Ladear um obstáculo não é transpô-lo

Este é então o nó teórico que os defensores da renascida teoria da identidade tipo-tipo têm que enfrentar. Dado o modo como o mesmo se encontra atado, não é uma surpresa que as propostas por estes apresentadas para o desatar sejam propostas defendendo que a nossa intuição de que a correlação que se verifica existir entre a dor e o seu correlato neurofisiológico seria de carácter contingente é, na realidade, e contra Kripke, uma intuição inválida.

Vou chamar ‘agnosticismo modal’ a esta linha de ataque contra o desafio de Kripke. Os agnósticos modais defendem, basicamente, que as nossas intuições acerca do que é ou não concebível não constituem qualquer guia fiável para decidir questões de possibilidade metafísica. Deste ponto de vista, mesmo pressupondo que Kripke tem razão quando afirma que nós temos, de facto, uma intuição poderosa de acordo com a qual a dor e o disparo das fibras-C se encontram apenas contingentemente correlacionadas, a existência em nós de uma tal intuição é irrelevante para decidir questões acerca do modo como as coisas poderiam ou não ser.

Terão os agnósticos modais razão na sua alegação? Como sabê-lo? Em que fundamentos podemos basear os nossos juízos de que as intuições associadas às nossas capacidades de conceber um estado de coisas como possível não são um indicador fiável da necessidade metafísica? Deveremos, para o decidir, basear-nos em intuições acerca dessas intuições? E, nesse caso, o que devemos fazer com as intuições que apontam na direcção oposta? Esta é uma discussão que decorre num terreno muito armadilhado; como não quero desviar-me do meu objectivo, não vou agora entrar nela.

Vou por isso dar o benefício da dúvida ao agnosticismo modal e supor, sem mais discussão, que os agnósticos modais têm razão na sua alegação. A questão que quero realmente colocar é a seguinte: seguir-se-ão resultados substancialmente novos desta suposição? Na realidade, o que, no caso da avaliação da proposta de identificação das dores com o disparo das fibras-C, se segue do agnosticismo modal, enquanto tal, é apenas que é possível que esses *relata* sejam realmente idênticos, a despeito do facto de termos uma intuição bem distinta de que eles se encontram apenas contingentemente correlacionados, e que, portanto, não podem de todo ser idênticos. É claro que, dado o carácter agnóstico da perspectiva em causa, é também possível que, ainda de acordo com a mesma, as dores não sejam idênticas aos seus

correlatos neurais. Assim o que realmente se segue deste ponto de vista é, simplesmente, que não sabemos qual das possibilidades é, de facto, o caso.

Ora, esta parece-me ser uma resposta muito fraca a Kripke. É verdade que a objecção levantada por este último fica, por meio dela, bloqueada. Mas ela não fica, nem pouco mais ou menos, refutada. Além disso, mesmo aceitando esta resposta, Kripke continua a poder ter razão, mesmo que não tenha que tê-la. Na realidade, esta resposta constitui um lance puramente defensivo, cujo objectivo parece ser apenas o de evitar uma refutação poderosa da teoria da identidade tipo-tipo. Neste sentido, a única coisa que ela alcança é evitar a exclusão desta teoria do conjunto das posições viáveis no espaço lógico das respostas possíveis ao problema mente-corpo. Mesmo que pressuponhamos a sua validade, aquilo que dela obtemos é, simplesmente, que, tanto quanto somos capazes de saber, a teoria da identidade tipo-tipo mantém-se nesse conjunto, juntamente com todas as outras propostas de resolução do problema mente-corpo que já nos são familiares.

Mas a promessa associada ao ressurgimento desta teoria é a de que teríamos novas razões *positivas* para crer nela, e não apenas razões para operar uma suspensão agnóstica a respeito das objecções de Kripke contra ela. Que razões positivas poderão ser essas? Repare-se que, se, de acordo com o agnosticismo modal, não podemos recorrer às nossas intuições de concebibilidade para aceitar as objecções de Kripke, então tão-pouco deveremos poder recorrer a elas para comparar umas com as outras as diferentes sugestões de resposta ao problema metafísico que nos ocupa.

Chegados aqui, creio ser necessário enfatizar que uma das condições necessárias para a admissibilidade ao debate filosófico de uma sugestão de resposta a este problema é a de que a mesma seja compatível com a evidência empírica disponível. Portanto, todas as que continuam a participar nele o são. Por outro lado, deixai-me lembrar-vos que, se as intuições a que Kripke faz apelo são inúteis para o debate, então as intuições acerca de zumbis também deveriam sê-lo. Mas, se este é o caso, é legítimo levantar-se a questão de saber por que é que é suposto que, de acordo com, por exemplo, Block, não aceitemos as primeiras e as suas consequências, enquanto que é suposto que aceitemos que as segundas mostraram a incapacidade do funcionalismo em lidar com o problema da consciência e, por conseguinte, de se estabelecer como uma perspectiva materialista viável. De facto, se as intuições acerca da possibilidade metafísica são espúrias,

poderíamos ter simplesmente desconsiderado como igualmente espúrias e, portanto, irrelevantes, as intuições acerca de zumbis e mantido no activo uma parte substancial da parafernália funcionalista!

Este é precisamente o ponto do debate em que entra em cena a ideia de recorrer a argumentos construídos como inferências para a melhor explicação. Com efeito, se todas as posições que participam no debate são igualmente compatíveis com a evidência empírica, se as intuições de concebibilidade são inúteis para decidir questões de possibilidade metafísica, e se considerações acerca de simplicidade ontológica ou economia conceptual são extremamente dúbias, dado que tendem a incorrer em petições de princípio, então o único tipo de argumento que parece manter-se disponível para ser utilizado é, de facto, a inferência para a melhor explicação.

Como mencionei acima, creio que as inferências para a melhor explicação são notoriamente ineficientes para tratar deste tópico. Um dos problemas levantados pelo recurso a tais argumentos, e não o menor de entre eles, é o de que, neste contexto, não temos uma noção previamente definida daquilo que a expressão ‘boa explicação’ poderia querer dizer. Lembremo-nos que não devemos misturar a questão metafísica ou filosófica de determinar que espécie de relação formal se verifica entre os predicados pertencentes a dois tipos de discurso com questões empíricas relacionadas com o modo como deve alguma porção de nova evidência ser adequadamente tratada. Milhões e milhões de medicamentos são manufacturados todos os dias para evitar ou aliviar a dor. Não há dúvida que os cientistas que controlam os processos de fabrico na Bayer, ou outros conglomerados farmacêuticos, conhecem todas as questões relevantes a respeito da história química subjacente aos efeitos que o ácido acetilsalicílico, por exemplo, tem nas nossas células nervosas. Mas não é isso que procuramos aqui. O que queremos saber é se o nosso *conceito* de dor deve ou não ser *identificado* com o conceito de disparo de fibras-C ou com o conceito de um estado fenomenológico ou com o conceito de outra coisa qualquer. Por referência a o que é que uma melhor explicação deve, nestas circunstâncias, ser considerada melhor? Creio que qualquer resposta a esta pergunta está destinada, neste contexto, a ser tão circular quanto o são os apelos à simplicidade.

Como seria de esperar, o problema mencionado no parágrafo anterior acaba por vir à tona quando consideramos algumas das tentativas recentes de usar a estratégia que lhe dá origem. Assim, no âmbito do terreno

epistémico aberto pelo agnosticismo modal, podemos encontrar propostas de uso de inferências para a melhor explicação no tratamento do problema mente-corpo que apontam para direcções precisamente opostas. Um bom exemplo do que acabou de ser dito são os argumentos propostos por Block (e Stalnaker), por um lado, e por Levine, por outro lado. De acordo com Block e Stalnaker, o uso deste tipo de inferência permite-nos vindicar a verdade da teoria da identidade tipo-tipo a respeito da mente consciente; de acordo com Levine, o que o uso deste tipo de inferência nos mostra é que há um fosso explicativo intransponível a separar o modo como os fisicistas tratam os *qualia* e o carácter subjectivo dos mesmos. Objectivamente, Block (e Stalnaker) e Levine não podem estar ambos certos. Deixai-me introduzir o desacordo entre eles, começando por rever os argumentos de Levine.

5. Um caminho que não leva ao destino

Se alguma resposta particular ao problema mente-corpo for vindicada por meio de um argumento tendo a estrutura de uma inferência para a melhor explicação, então tem que ser o caso que o poder explicativo fornecido por uma tal resposta seja realmente o maior. Mas como podemos determinar isso? A resposta de Levine a esta questão é, à primeira vista, um tanto ou quanto desconcertante. Com efeito, e depois de tudo o que já ficou dito acima acerca do ‘agnosticismo modal’, Levine alega que podemos determinar qual de duas ou mais opções tem maior poder explicativo recorrendo a... intuições de concebibilidade. Mas, na realidade, a despeito de ser um agnóstico a respeito da capacidade das intuições de concebibilidade nos revelarem alguma coisa de relevante acerca da possibilidade metafísica, Levine sustenta que essas intuições não são despiciendas. Segundo ele, elas desempenham um importante papel *epistemológico*.

De acordo com o ponto de vista de Levine, pode produzir-se um bom resultado explicativo no problema mente-corpo se a redução por meio da qual se tentar resolvê-lo nos fornecer ganhos explicativos análogos àqueles que o empreendimento reducionista nos forneceu nas ciências naturais. Com efeito, a descoberta científica das identidades teóricas ‘água=H₂O’ e ‘calor=energia cinética molecular média’ melhorou grandemente a nossa compreensão do comportamento macroscópico da água e dos fenómenos térmicos. Assim, alega Levine, se as identidades ‘dor=disparo das fibras-C’

ou ‘consciência=actividade das células piramidais’ forem para ser entendidas de um modo análogo ao das identidades teóricas da ciência, como os defensores da teoria da identidade tipo-tipo defendem que são, então tem que ser o caso que o seu estabelecimento acabe por nos fornecer um poder explicativo semelhante ao destas últimas. Ainda segundo ele, é com vista a determinar se este é ou não o caso que o apelo para intuições de concebibilidade se revela crucial.

Vejamos como. Diz-nos Levine que, se estas identidades teóricas fossem todas do mesmo género, então as intuições de concebibilidade que poderíamos extrair delas deveriam também ser análogas. Mas não são. Em particular, Levine defende que se a água é, de facto, H₂O, então, apesar de podermos conceber que poderia existir por aí alguma matéria aquosa que se viesse a revelar não ser H₂O (e, portanto, se viesse a revelar tão-pouco ser água), não podemos, de modo semelhante, conceber, de modo coerente, a possibilidade de que a substância H₂O poderia não ter sido matéria aquosa, sem introduzir mudanças de monta à química terrestre. Isto é, dentro dos constrangimentos impostos pela nossa teoria química, a identificação da água com H₂O não é, simplesmente, uma questão de pura definição, cujo objectivo seria a obtenção de alguma espécie de economia conceptual. É também um meio de iluminar explicativamente o comportamento previamente conhecido da água em diferentes situações macroscópicas e de ampliar este conhecimento. Todavia, as intuições de concebibilidade suscitadas pelas alegações de que a dor seria idêntica ao disparo das fibras-C, ou de que a consciência seria idêntica à actividade das células piramidais, não nos colocam numa situação análoga.

Na realidade, e como o argumento de Kripke o mostra, no caso das identificações psicofísicas deste género continuamos a ser capazes de conceber distintamente a possibilidade de que, por exemplo, o disparo das fibras-C poderia não ter estado associado a experiências dolorosas sem que, com isso, nos sintarmos minimamente compelidos a introduzir mudanças importantes no nosso conhecimento neurofisiológico. Só que, em vez de nos mostrarem o que quer que seja de carácter metafísico, o que essas intuições de concebibilidade nos mostram é, segundo Levine, que nada há nessa identificação teórica que, por analogia com o caso da identificação da água com H₂O, contribua, mesmo que apenas minimamente, para *explicar* por que é que precisamente este género de mecanismo bioquímico está em nós associado ao género de experiências a que supostamente está associado.

A tese subjacente à interpretação de Levine destas nossas intuições é a de que uma boa explicação para um fenómeno deve tornar epistemologicamente necessária a descrição das suas principais características. Assim, o que, segundo ele, acontece quando tentamos identificar propriedades conscientes com propriedades físicas, do mesmo modo que identificamos substâncias e propriedades macrofísicas com elementos e propriedades microfísicos, é que, em cada um destes casos, obtemos intuições consideravelmente dissemelhantes a respeito de questões de necessidade epistemológica. Nas identidades teóricas genuínas, proporcionadas pela ciência, os redutores tornam epistemologicamente necessárias as principais propriedades dos materiais reduzidos; mas, ao contrário do que acontece nestas, as propriedades neurofisiológicas não tornam epistemologicamente necessárias as propriedades sentidas nos estados conscientes.

Em conclusão, Levine alega que tais identificações não são capazes de nos ajudar a explicar por que é que as propriedades conscientes são como são. E, se não são capazes de fazê-lo, então, ainda segundo Levine, não faz qualquer sentido defender que tais identificações poderiam ser alcançadas por meio de inferências para a melhor explicação.

Qual é a resposta de Block ao argumento de Levine? Juntamente com Stalnaker, ele argumenta que, em si mesmo, o propósito de uma vindicado de identidade não é providenciar uma qualquer explicação, mas, inversamente, impedir o surgimento da necessidade de procurá-la. Se é verdade que a dor é realmente idêntica ao disparo das fibras-C, então a questão de querer saber por que é que o disparo das nossas fibras-C estaria associado às nossas dores não deveria sequer ser levantada. Nessas circunstâncias, colocar esta questão seria algo de semelhante a colocar a questão de por que é que a água ocorre sempre em associação com H₂O. Obviamente, dada a identidade que se verifica existir entre a água e H₂O, colocar esta questão é o mesmo que colocar a questão por que é que a água é água, isto é, não faz sentido.

O recurso por Block e Stalnaker a este estratagema argumentativo não pode deixar de originar no participante atento neste debate o sentimento de que o que ele está a tentar obter por meio de uma tal jogada é um processo através do qual possa proibir-nos de formular uma questão difícil; mas vamos, por um momento, ignorar este sentimento.

Concentremo-nos antes na outra questão crucial suscitada por esta jogada de Block e Stalnaker e que fica por responder: se é isto o que se pretende obter com a identificação teórica que ele nos propõe, então por que é que a selecção filosófica de uma teoria da identidade tipo-tipo deveria resultar de uma inferência para a melhor explicação?

Com franqueza, creio que a resposta de Block e Stalnaker a esta questão em nada mais se apoia senão numa profissão de fé. Com efeito, o que eles acabam por dizer para lhe responder é que a investigação neurofisiológica futura encarregar-se-á de produzir um corpo de conhecimentos teoricamente consistente, no interior do qual a substituição da conversa acerca de conceitos fenomenalistas por conversa acerca dos seus equivalentes neurofisiológicos produzirá um ganho explicativo distintivo para a nossa compreensão dos fenómenos mentais.

Pode até ser que isto venha a ser verdade, mas como podemos sabê-lo agora? Será que o pomo da discórdia aqui não é precisamente a própria possibilidade de que esse possa alguma vez vir a ser o caso? E se a disputa filosófica/metafísica pode ser resolvida pelo simples aprofundamento da investigação empírica e pelas esperanças cognitivas nele depositadas, então para quê todo este espalhafato? O que estamos todos a fazer aqui afinal?

Deste modo, depois de terem posto em causa, de uma forma puramente negativa, a validade das intuições kripkeanas, ambas as linhas argumentativas do agnosticismo modal que apelam para o recurso a inferências para a melhor explicação terminam como um grande desapontamento.

6. Outro caminho

Mas, na realidade, os agnósticos modais dividem-se em dois tipos. Ao lado da linha de resposta inferencialista-explicativa ao desafio de Kripke, há também a linha de resposta a que chamarei da ‘disposição positiva’. Ao contrário da estratégia da linha inferencialista-explicativa, a estratégia da linha da disposição positiva não consiste em recorrer a inferências para a melhor explicação. Os filósofos dispostos positivamente, de acordo com o modo como estou aqui a usar esta expressão, têm tão pouca confiança no significado modal/metafísico das nossas intuições de concebibilidade quanto os outros agnósticos modais. Mas acrescentam um elemento novo a essa

desconfiança. Este elemento novo é a aceitação da tese de acordo com a qual sensações qualitativas, tais como a dor, são realmente representadas por conceitos fenomenais próprios.

David Papineau é o mais significativo e coerente dos agnósticos modais ‘dispostos positivamente’. Tal como Kripke, ele defende que os termos para conceitos fenomenais conseguem de algum modo atingir a sua referência directamente, em vez de o fazerem através de intermediários cognitivos. Por outras palavras, Papineau sustenta, por um lado, que há uma conexão directa especial entre um termo com essas características e a sua referência e, por outro lado, que esta conexão directa consegue, de algum modo, estabelecer uma ligação semântica durável entre os dois. Até aqui, Papineau está, portanto, de acordo com o anti-materialista Kripke. Mas Papineau desenvolve também uma teoria fisicista positiva com o objectivo de explicar como é que uma tal relação misteriosa de referência pode surgir. É este aspecto do seu empreendimento teórico que me faz seleccionar o rótulo ‘disposto positivamente’ para o caracterizar.

Esta é uma contribuição surpreendente para o debate. Com efeito, quando consideramos o desenvolvimento do materialismo naturalista ao longo do século XX, aquilo que nele é mais saliente é a substituição, tanto da velha interpretação fenomenológica do sentido dos termos mentais da nossa linguagem, como da sua rival comportamental-fisicista, por uma interpretação funcional dos mesmos. A consequência dessa mudança para o problema mente-corpo foi, precisamente, a de que, enquanto que o modo como este problema afligia as mentes dos fisicistas dos anos 50 parecia intratável, o modo como ele passou a apresentar-se para os filósofos funcionalistas tornou-o susceptível de uma solução. Na realidade, a tarefa destes últimos veio a revelar-se ser a de encontrar um modo de integrar a interpretação funcional do discurso mental com o discurso neurofisiológico, em vez de ser a de tentar reduzir o sentido de termos que pareciam referir conteúdos fenomenológicos ao sentido de termos que referem eventos neurofisiológicos. Tanto a relação de sobreveniência do mental no físico, introduzida por Davidson, como a relação de realização que se verificaria entre propriedades (mentais) de ordem superior e propriedades (fisiológicas) de ordem inferior, introduzida por Putnam, conseguiam operar uma tal integração; deste modo, os eventos mentais poderiam ser identificados exemplar a exemplar com eventos de natureza física, mesmo que as propriedades mentais não pudessem ser identificadas tipo a tipo com

propriedades de natureza física. Como disse acima, estas soluções para o problema mente-corpo eram, por um lado, intrinsecamente plausíveis; e eram, por outro lado, imunes ao argumento modal de Kripke, uma vez que, precisamente, cortavam a ligação semântica indissolúvel, pressuposta por Kripke, entre os termos mentais e as nossas experiências fenomenológicas. A sua desvantagem consistia, porém, em que faziam desaparecer do mapa teórico as experiências fenomenológicas que constantemente associamos ao nosso uso desses termos. Isto é, por seu intermédio, a linguagem da Psicologia via-se reconciliada com o modo científico de ver o mundo, mas, ao mesmo tempo, ficava completa e definitivamente divorciada da fenomenologia.

Como também já mencionei acima, a saída de cena do funcionalismo, no final do século XX, parece ter sido causada por um regresso irresistível da fenomenologia. E uma das motivações que subjazem ao trabalho de Papineau em Filosofia da Mente parece ser, de facto, a de se confrontar com esse regresso. Papineau reconhece a força da queixa fenomenológica contra o funcionalismo e defende, em consonância, que uma compreensão adequada da semântica dos nossos termos mentais nos deve levar a reconhecer que tais termos têm *tanto* um aspecto fenomenal *como* um aspecto funcional. Assim, segundo ele, a questão que deve pôr-se ao filósofo fisicista deixa de ser a de substituir a velha e mais intuitiva interpretação fenomenológica do sentido dos termos mentais pela interpretação fisicista dos mesmos para passar a ser a de integrar a primeira na segunda.

Ao argumentar deste modo, Papineau traz-nos de volta a um período ainda mais recuado da Filosofia da Mente do século XX. Mais concretamente, ele traz-nos de volta à distinção, para que Wittgenstein e os filósofos do Círculo de Viena chamaram a atenção nos anos 30 e 40 desse século, entre o modo como devem entender-se usos de termos mentais na primeira pessoa do singular e o modo como devem entender-se usos de termos mentais na terceira pessoa do singular. A ideia principal que foi então avançada por estes filósofos foi a de que a semântica desses termos funcionava de modo diferente nas suas diferentes condições de uso. Para Carnap, essa diferença era uma diferença de referência: os usos na primeira pessoa eram usos fenomenológicos, que referiam as experiências privadas do sujeito, enquanto que os usos na terceira pessoa eram usos comportamental-fisicistas referindo eventos que ocorriam no mundo público. Para Wittgenstein, essa diferença era uma diferença no uso da linguagem:

nos seus usos na primeira pessoa, frases com conteúdo psicológico do género de 'Estou com dores de dentes' estariam a ser usadas expressivamente, como formas comportamentais mais sofisticadas do que o choro, o gemido ou o gesto de levar a mão à bochecha, mas não essencialmente diferentes destes; em contrapartida, nos usos na terceira pessoa, as frases com conteúdo psicológico estariam a ser usadas descriptivamente, e estariam, de facto, a identificar o que mais tarde ficou a ser conhecido como papéis causais-funcionais.

Dado que Carnap se interessava principalmente pelo estudo da linguagem da ciência, ele acabou por considerar que os usos dos termos e expressões mentais na primeira pessoa eram descriptivamente irrelevantes, mesmo que fossem comportamentalmente relevantes. E dado que Wittgenstein considerava os usos na primeira pessoa de tais palavras e frases como sendo apenas expressivos, acerca desses usos não se punha sequer a questão de tentar descortinar qual seria a natureza da referência desses termos nesses contextos: pura e simplesmente, não havia aí qualquer referência para descobrir. Assim, a questão de tentar encontrar uma referência comum para usos fenomenais (na primeira pessoa) e psicológicos (na terceira pessoa) de termos mentais não se levantava sequer. Por outro lado, para esses filósofos, a linguagem, científica ou outra, era um jogo normativo, público e intersubjectivo; assim sendo, a semântica deveria procurar os sentidos dos termos que dela faziam parte fora do domínio da experiência privada. Para alguém com inclinações místicas, como Wittgenstein, esta deveria permanecer virgem de todo o armadilhamento conceptual e deveria ser deixada livre e intocada para a contemplação e apreciação silenciosa e não mediada de todo o ser consciente.

Mas este não é, de todo, o caso de Papineau. Ele regressa à análise clássica da distinção primeira pessoa/terceira pessoa em termos de uma distinção entre conceitos fenomenais e conceitos de papel funcional/causal, mas, ao contrário do Carnap dos anos 30, ele defende que um termo como 'dor' tem uma referência única, comum a ambos os seus usos conceptuais, a saber, o evento físico que realiza o papel funcional da dor em nós, seja ele o disparo das fibras-C, ou o que quer que seja que a Neurofisiologia acabe por vir a descobrir a este respeito. Esta é uma variação engenhosa da tese clássica da identidade tipo-tipo, que, além de pretender vindicar o materialismo, pretende também, e em simultâneo, vindicar um papel causal-

explicativo para a chamada ‘Psicologia Popular’, ou seja, para uma teoria da acção humana construída a partir do silogismo prático aristotélico.

7. O segundo obstáculo

Já disse acima que, no contexto deste ensaio, não irei pôr em causa a validade da alegação anti-kripkeana de acordo com a qual as intuições de concebibilidade não têm as capacidades metafísicas que é costume serem-lhes atribuídas. Todavia, além da objecção de Kripke, há outra objecção que a perspectiva defendida por Papineau tem que enfrentar, a saber, a objecção da privacidade. Esta objecção pode ser expressa da seguinte forma: como é possível atribuir um valor semântico à ideia do estabelecimento de uma relação de referência privada e directa entre aquilo que Papineau considera serem os nossos termos para conceitos fenomenais e os seus referentes, qualquer que possa ser a sua natureza ontológica última? Note-se que, uma vez que recusa o materialismo, este não é, necessariamente, um problema para Kripke, pelo menos enquanto este não quiser confrontar-se com o problema da existência de outras mentes e do solipsismo.

Mas Papineau é um materialista; ele não pode, por isso, passar simplesmente por cima deste problema. Para lhe responder, ele chama então a nossa atenção para o facto de que as objecções que mencionam o problema da privacidade, tipicamente inspiradas no argumento da linguagem privada de Wittgenstein, pressupõem que a linguagem é um jogo social normativo. Todavia, e apesar de não usar exactamente estas palavras, Papineau defende que, desde que Chomsky introduziu com apreciável sucesso os conceitos de linguagem-I, faculdade da linguagem e gramática interna em Linguística, a nossa ideia de o que é uma linguagem humana deixou de ser a de que a mesma seria um jogo social e passou a ser a de que ela seria um órgão natural. Ora, a partir desta perspectiva mais recente de o que é a linguagem, segue-se que o modo como os itens mentais são mobilizados pelo órgão da linguagem e se relacionam com a realidade neurofisiológica subjacente é uma questão totalmente biológica acerca da mecânica da qual a filosofia pré-chomskiana da linguagem nada tem de relevante a dizer.

Tal como no caso do agnosticismo modal, esta resposta ao problema da privacidade, em vez de o enfrentar nos seus próprios termos, afirma apenas

uma mudança de território. E tal como o fiz no caso do agnosticismo modal, também aqui vou limitar-me a aceitar esta mudança de território sem mais discussão. Ponhamos então de lado a objecção da privacidade, aceitemos que Kripke e Wittgenstein estão ambos fora de jogo, e aceitemos também que quaisquer argumentos baseados no seu trabalho filosófico não constituem razões genuínas que impeçam a teoria da identidade tipo-tipo de ser uma posição viável no debate em torno do problema mente-corpo. Mas, mais uma vez, aquilo que me preocupa, peço desculpa por insistir, é encontrar razões *positivas* para crer na bondade da teoria da identidade tipo-tipo, e não propriamente encontrar razões para não a excluir à partida do âmbito do debate sério em torno do problema mente-corpo. O que tem então Papineau a dizer-nos de novo a este respeito?

A sua contribuição original para o debate consiste na apresentação de uma teoria positiva dos conceitos mentais, a qual permitiria conciliar a indissolubilidade da conexão semântica entre os termos fenomenais e os seus referentes, pressuposta por Kripke, com a defesa da teoria da identidade tipo-tipo. Deixai-me começar por dar-vos conta em traços largos desta teoria positiva. Trata-se, basicamente, de uma actualização naturalizada da Psicologia de Hume. Na realidade, Papineau diz-nos que os conceitos fenomenais que ele introduz na discussão do problema mente-corpo são conceitos que têm um componente indexical e que tanto podem ser usados imaginativamente como introspectivamente. No primeiro destes usos, o pensador traz à sua mente aquilo a que Hume chamou uma ‘cópia esmaecida’ da própria experiência; no segundo destes usos, o pensador amplifica a experiência em curso naquilo a que Hume chamou uma ‘cópia vívida’ de si mesma. Deste modo, em qualquer um dos usos possíveis de um conceito fenomenal, a experiência à qual o conceito se reporta é, ela própria, usada no pensamento; o ponto crucial associado a esta ideia é o de que pensar fenomenalmente numa experiência dá a sensação de se estar a ter a própria experiência. Ao contrário, quando pensamos no modo material, referimos descritivamente; as experiências não entram, elas próprias, no processo de pensamento.

Papineau sustenta, então, que seria uma falácia partir desta diferença entre o modo material de pensar e o modo fenomenal de pensar para argumentar que o primeiro deixaria algo de fora de um ponto de vista semântico. Ele baptizou esta nova falácia com a expressão ‘falácia antipatética’ e analisou-a em termos da falácia uso/menção. A ideia aqui

seria a de que, quando as pessoas argumentam que o modo material de pensar deixaria a experiência de fora, o que elas estariam realmente a fazer seria expressar a sua percepção do facto de que este modo de pensar, ao contrário do modo fenomenal de pensar, não usaria as próprias experiências no processo de pensamento. Mas, de acordo com Papineau, mesmo que este seja um fenómeno intrinsecamente interessante, ele não introduz qualquer novidade essencial na questão que nos ocupa. Sendo verdade que a maioria dos nossos conceitos se comporta como os conceitos materiais, tão-pouco deixa de ser o caso que, quando pensamos no modo fenomenal, e usamos, por conseguinte, um item para se representar a si mesmo, aquilo que fazemos não deixa de ser um acto de representação, tal como quando usamos símbolos arbitrários para produzir uma representação. O facto de um pensamento no modo fenomenal nos dar a sensação de estarmos a ter a experiência acerca da qual ele é não é então mais do que um efeito colateral de uma estratégia representacional particular do cérebro. Abreviando: do facto de haver mais do que uma forma possível de representar nada se segue a respeito do objecto da representação. Não haveria então nenhuma razão de princípio pela qual este objecto não poderia ser o mesmo em ambos os casos.

Na realidade, a história apresentada acima não é acerca de quaisquer termos presentes na nossa linguagem natural. Ela é acerca do modo como a nossa linguagem do pensamento, no sentido do termo originalmente introduzido por Fodor, consegue representar usando experiências qualitativas. Dada a posição teórica na qual se coloca, aquilo de que Papineau está aqui à procura é de uma teoria empiricamente verdadeira acerca do modo como o cérebro representa o mundo e, nele, os estados do corpo do qual é a unidade de processamento central. Tratar-se-ia então de mostrar como a verdade desta história empírica tornaria plausível a tese da identidade tipo-tipo. Neste sentido, a plausibilidade da tese metafísica que nos interessa testar é feita depender da verdade de uma hipótese empírica particular. Nestas circunstâncias, para avaliar dessa plausibilidade, necessitamos de tomar contacto com a evidência empírica que poderia sustentar o acerto da hipótese referida. Onde, então, se encontra ela? Tanto quanto eu saiba, em lado nenhum.

Como poderemos, então, avaliar a contenção de Papineau? Uma resposta possível a esta pergunta poderia ser a de se alegar que, não podendo ainda ser avaliada em termos de confirmação empírica, a hipótese

proposta poderia, todavia, ser provisoriamente avaliada em termos da sua plausibilidade à luz do conhecimento de fundo disponível. Mas, mesmo que nos mantenhamos no âmbito das intuições impressionísticas acerca de o que essa evidência poderá, eventualmente, vir a ser, dado esse conhecimento de fundo, não creio que o panorama que aí consigamos discernir favoreça as pretensões empíricas de Papineau. Com efeito, os nossos pensamentos acerca da dor, por exemplo, não são, em geral, dolorosos, nem sequer levemente (apesar de poderem ser assustadores). Neste sentido, não consigo ver como é que esta história representacional se poderia adequar ao que podemos imaginar acerca de o que poderia ser a evidência apropriada ao território teórico recortado pela abordagem de Papineau. Ou, dizendo o mesmo por outras palavras, não vejo que tenhamos quaisquer fundamentos empíricos para acreditar, mesmo que apenas provisoriamente, que esta história possa ser verdadeira. Lembremo-nos, por outro lado, que todas as outras posições avançadas no âmbito do problema mente-corpo nos apresentam igualmente histórias em termos das quais uma semântica compatível com elas poderia ser desenvolvida. Na ausência de evidência empírica a seu favor, por que razão, então, escolher esta? Mais uma vez me vejo aqui forçado a relembrar que o que me interessa obter são razões *positivas* para acreditar na teoria da identidade tipo-tipo. E o interesse que possamos ter em vindicá-la não constitui, obviamente, uma dessas razões.

Na realidade, e como, aliás, o próprio Papineau reconhece, o verdadeiro papel inferencial no seu argumento é desempenhado por um passo adicional de natureza conceptual. Este passo contém o desfecho do chamado ‘argumento causal’. Em traços gerais, este argumento tem o seguinte aspecto. A primeira premissa afirma que os estados mentais, qualitativos ou não, têm efeitos físicos característicos. A segunda premissa afirma que todos os efeitos físicos são integralmente originados por causas físicas. A terceira premissa exclui a sobredeterminação como uma opção viável. A conclusão que se segue da conjunção destas três premissas é a conclusão fisicista de que *as causas mentais de efeitos físicos são, na realidade, eventos físicos*.

Obviamente, nenhuma das três premissas do argumento causal é neutra. A primeira premissa exclui, desde logo, o epifenomenismo. A segunda premissa é o sumário de uma tese particular em Filosofia da Física, a saber, a tese da completude da Física. Mais uma vez, uma tal tese necessita de ser justificada a partir de uma base independente. É óbvio que ela exclui desde logo o dualismo cartesiano. E a terceira tese introduz pressupostos

relevantes acerca da causação que também necessitam de ser objecto de uma argumentação independente.

Nenhuma destas premissas me incomoda. Não sou um epifenomenista, nem sou um dualista cartesiano. O que me incomoda neste argumento é que o mesmo sempre foi do nosso conhecimento. É um argumento a favor do materialismo. Enquanto tal, nada faz para discriminar entre as diferentes variedades de materialismo. É tão compatível com a teoria da identidade tipo-tipo como é compatível com o monismo anómalo ou com qualquer outra abordagem materialista que faça uso da ideia da identidade exemplar-exemplar. Seja como for, se nós já nos inclinamos de antemão para o materialismo, já sabemos o que é que, de acordo com as nossas convicções, deve ser o caso. O que queremos saber é *como* o que achamos que deve ser o caso *pode*, de facto, ser o caso. Porque, intuitivamente, parece estranho que o seja. E nada de novo surge aqui no horizonte a este respeito. Mais ainda, dado o carácter à partida não neutral das premissas do argumento, nenhum não-materialista alguma vez se deixará convencer por seu intermédio.

Assim, só por si, a história representacional de Papineau não nos leva muito longe. Ela aponta numa direcção, isto é, pode ajudar-nos a compreender, em abstracto, como é que representações pertencendo a diferentes modos de pensar – o modo fenomenal e o modo material – podem acabar por revelar-se ser representações de uma mesma coisa. Mas, dado o facto de situar-se num terreno puramente empírico, ela necessita de ser sustentada por alguma evidência que contribua, mesmo que apenas minimamente, para a sua confirmação. Na ausência de qualquer evidência empírica deste género, e na presença de o que parece ser evidência que a desconfirma, ficamos apenas com o argumento causal. Mas este, além de não permitir discriminar entre as diferentes variedades de materialismo, só consegue convencer aqueles que, desde o princípio, já se encontravam convencidos da justeza da sua conclusão.

8. Outro obstáculo ladeado mas não transposto

Na secção anterior, argumentei que, na ausência de qualquer evidência empírica que a sustente, a história cerebral-representacional de acordo com a qual o nosso cérebro usaria as próprias experiências que produz como

sinais referenciais não constitui qualquer argumento positivo a favor da teoria da identidade tipo-tipo. E argumentei, também, que, tanto quanto nos é possível ter uma ideia de o que essa evidência possa vir a ser, ela parece ir contra as pretensões de Papineau. Agora, porém, gostaria de ir um pouco mais longe. Nomeadamente, gostaria de defender que a tese de carácter mais geral de que a referência dos conceitos fenomenais, presumindo que tal coisa existe, e a referência dos conceitos psicológicos, poderia ser idêntica, para além de não ser suportada pela evidência disponível, parece também ser contraditória com, pelo menos, parte dela. Neste sentido, defendo que ela é implausível.

De um modo consistente com a sua perspectiva, Papineau defende, mais uma vez, que esta é uma questão inteiramente *a posteriori*. Também mais uma vez, não vou disputar a definição do território que ele escolheu para tratar desta questão; vou por isso tentar abordá-la a partir do ângulo empírico em que ele se coloca. Ora, quando consideramos esta contenção a partir desta perspectiva, e olhamos para a evidência que se pressupõe ser relevante, acabamos por nos aperceber que parte dela, pelo menos, aponta, sem margem para dúvidas, na direcção oposta à contenção de Papineau. Com efeito, existem resultados empíricos, como aqueles que foram obtidos por Libet nas suas experiências com o cérebro humano, ou como aqueles que foram obtidos em programas de investigação que tentam identificar as heurísticas rápidas e frugais subjacentes à implementação de uma proporção significativa dos nossos comportamentos e acções, que parecem sugerir que os mecanismos neurais que subservem um número importante das nossas respostas, decisões e intenções de agir funcionam antes de, e independentemente de, os nossos sentimentos, percepções e racionalizações conscientes.

Não estou em posição de atribuir um valor definitivo a estes resultados empíricos. Mas eles tocam um ponto sensível, na medida em que parecem sustentar uma teorização psicológica que nos permite atribuir sentido a importantes fragmentos da nossa experiência de nós mesmos. Estas considerações não são, obviamente, de cariz anti-físicista. As propriedades fenomenais que caracterizam os nossos sentimentos, percepções e racionalizações conscientes têm efeitos físicos no mundo, nomeadamente, o que nos fazem dizer, acções que nós executamos por causa daquilo que nos fazem dizer, etc. Deste modo, se o argumento causal estiver, de facto, correcto, estas propriedades fenomenais terão que ser físicas.

Imaginemos, entretanto, que já conseguimos, de algum modo, não interessa agora qual, encontrar uma confirmação independente que sustente as premissas do argumento causal. Dada a validade lógica do argumento, isto significaria que teríamos encontrado uma confirmação independente para a sua conclusão, isto é, para a proposição que afirma a identidade entre propriedades fenomenais com eficácia causal e propriedades materiais. Se, nessas circunstâncias, levássemos em conta os resultados empíricos acima referidos, o que eles pareceriam então sugerir seria que, pelo menos num grande número de casos, as propriedades neurofisiológicas subjacentes com as quais as propriedades fenomenais se viriam a revelar ser idênticas seriam *distintas* das propriedades neurofisiológicas subjacentes com as quais as propriedades psicológicas, definidas em termos dos seus papéis causais/funcionais, se viriam a revelar ser idênticas.

Portanto, e pelo menos num número relevante de casos, os conceitos fenomenais, partindo, mais uma vez, do princípio algo duvidoso de que tal coisa existe, viriam a revelar-se ter uma referência diferente da referência dos conceitos psicológicos. Sob tais circunstâncias, não seria de todo uma surpresa que não se conseguisse encontrar evidência empírica capaz de sustentar a história psico-semântica proposta por Papineau.

A minha aposta, então, é que, se, como o defende a versão do fisicismo que ficou conhecida como teoria da identidade tipo-tipo, as propriedades fenomenais se vierem de facto a revelar ser identificáveis com propriedades neurofisiológicas, – e este é um *grande se* – acabaremos por ver-nos confrontados com uma Psicologia muito diferente daquela que temos hoje. Apesar de Papineau considerar este desfecho como teoricamente possível, ele desconsidera-o como sendo altamente improvável. Todavia, parece-me que a sua intuição de improbabilidade se encontra fortemente dependente do seu desejo de apresentar o seu próprio tipo de materialismo como sendo a única alternativa filosófica ao dualismo cartesiano que é igualmente capaz de preservar o essencial da psicologia do senso comum.

Portanto, para além de estar convencido que ele não apresentou quaisquer novos argumentos de carácter positivo que nos permitam ver a teoria da identidade tipo-tipo a uma nova luz, creio também que Papineau tão-pouco conseguiu convencer-nos que, caso a teoria da identidade tipo-tipo venha, de facto, a ser vindicada de algum modo que ainda não conhecemos nem imaginamos, então essa vindicação permitir-nos-ia, em simultâneo, salvar a Psicologia Popular como uma teoria explicativa da nossa acção

intencional. Pelo contrário, o que a evidência empírica disponível parece sugerir é que se uma tal vindicação vier, de algum modo, a ser possível, então ela acarretará, com toda a probabilidade, o abandono da Psicologia Popular como teoria explicativa da acção intencional.

9. Conclusão: estamos onde já estávamos

Espero ter tornado claro por que é que penso que os filósofos cujo trabalho passei aqui em revista não apresentaram quaisquer argumentos *positivos* realmente novos a favor do género particular de fisicismo de que tratei neste ensaio – a teoria da identidade tipo-tipo. E também por que é que penso que eles não responderam efectivamente às objecções que há meio século perseguem esta primeira variante do fisicismo moderno, tendo-se apenas limitado a abandonar o território onde a confrontação com elas é inevitável. Na realidade, creio que o problema da consciência se mantém tão duro de roer como sempre foi para os materialistas; e que, enquanto assim permanecer, não há grande esperança de que qualquer nova versão da teoria da identidade tipo-tipo nos possa fazer avançar para além do ponto onde já nos deixaram as variantes anteriores.

Referências Bibliográficas

- Block, N. & Stalnaker; R., 1999. Conceptual Analysis, Dualism and the Explanatory Gap. *Philosophical Review*, 108, 1-46.
- Carnap, R., 1931. Die physikalische Sprache als universal Sprache der Wissenschaft. *Erkenntnis*, 2, 219-241.
-- 1932/33. Psychologie in physikalischer Sprache. *Erkenntnis*, 3.
- Chalmers, D., 1995. Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, vol. 2, 3, 200-219.
-- 1996, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford, Oxford University Press.
- Chomsky, N., 1986, *Knowledge of Language: Its Nature, Origins and Use*, New York, Praeger.
-- 1998, *On Language*, New York, The New Press.
- Crick, F., 1994, *The Astonishing Hypothesis*, New York, Scribner.
- Crick, F. & Koch, C. 1990. Towards a Neurobiological Theory of Consciousness. *Seminars in Neuroscience*, 2, 263-275.
-- 1995. Are We Aware of Neural Activity in Primary Visual Cortex?. *Nature*, 375, 121-123.
- Davidson, D., 1980, *Essays on Actions and Events*, Oxford, Clarendon Press.
- Feigl, H., 1958. The 'Mental' and the 'Physical'. In: Feigl, H., Scriven, M. & Maxwell, G. (eds.), *The Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis, University of Minnesota Press.
- Fodor, J., 1976, *The Language of Thought*, Brighton, The Harvester Press.
- Gigerenzer, G., 2000, *Adaptive Thinking: Rationality in the Real World*, Oxford, Oxford University Press.
- Gigerenzer, G., Todd, P., The ABC Group, 1999, *Simple Heuristics that Make Us Smart*, Oxford, Oxford University Press.
- Hume, D., 1739/40, *A Treatise of Human Nature: being an attempt to introduce the experimental Method of Reasoning into Moral Subjects*, London. (Reprinted: London: Penguin Books, 1984).
- Jackson, F., 1982. Epiphenomenal Qualia. *Philosophical Quarterly*, 32, 127-136.
--1986. What Mary Didn't Know. *Journal of Philosophy*, 83, 291-295.
- Kim, J., 1998 *Mind in a Physical World*, Cambridge (MA), The MIT Press.
-- 2005 *Physicalism, or Something Near Enough*, Princeton (NJ), Princeton University Press.
- Kripke, S., 1980, *Naming and Necessity*, Cambridge (MA), Harvard University Press.

- Levine, J., 1993, On Leaving Out What It's Like. In: Davies, M. & Humphreys, G. H. (eds.), *Consciousness – Psychological and Philosophical Essays*, Oxford, Blackwell, 121-136.
- 2001, *Purple Haze*, Oxford, Oxford University Press.
- Libet, B., 1993, *Neurophysiology of Consciousness: selected papers and new essays*, Boston (MA), Birkhauser.
- 2004, *Mind Time. The temporal factor in consciousness*, Cambridge (MA), Harvard University Press.
- Papineau, D., 1993. Consciousness, Physicalism and the Antipathetic Fallacy. *Australasian Journal of Philosophy*, 71, 2, 169-183.
- 2002, *Thinking about Consciousness*, Oxford, Clarendon Press.
- Place, U. T., 1956. Is Consciousness a Brain Process?. *British Journal of Psychology*, 47.
- Putnam, H., 1975, Minds and Machines. In: Putnam, H., *Mind, Language and Reality*, Cambridge, Cambridge University Press.
- 1975, The Nature of Mental States. In: *Philosophical Papers*, vol. 2., Cambridge, Cambridge University Press.
- Smart, J. J., 1959. Sensations and Brain Processes. *Philosophical Review*, 68, 141-156.
- Wittgenstein, L., 1958, *The Blue and Brown Books*, Oxford, Blackwell.
- 1971, Notes for Lectures on 'Private Experience' and 'Sense Data'. In: Jones, O. R., (ed.), *The Private Language Argument*, London, MacMillan, 226-275.
- 1984, *Philosophische Untersuchungen*. In *Werkausgabe - Band 1*, Frankfurt a.M., Suhrkamp.

ISSN (on-line): 1647-659X
ISSN (print): 2182-2824



Kairos
Revista de Filosofia & Ciência
Journal of Philosophy & Science
<http://kairos.fc.ul.pt>



CFCUL

Centro de Filosofia das Ciências
da Universidade de Lisboa
<http://cfcul.fc.ul.pt>