Kairos

Revista de Filosofia & Ciência
Journal of Philosophy & Science

Abril / April 2013

6

Artigos
Papers

Nietzsche: Science and Truth
Danny Smith

Arqueología: arte, historia, antropología. Análisis filosófico de la génesis y desarrollo de una disciplina
Anna Estany

Freud, a concepção do descentramento e a Física Moderna
Lino Machado

Understanding Admissibility
George Masterton

Truth and Historicism in Kuhn's Thesis of Methodological Incommensurability
Marco Marletta

Are Colors Real?
Emiliano Boccardi

Dossier
Introspection and Intuition in Mathematics

Introduction: *Where experience matters*
Alexandra Van-Quynh

Les possibilités de l'expérience. Mathématiques, aperception pure et aperception empirique dans la *Critique de la raison pure* de Kant
Matthieu Haumesser

On the possibility and reality of introspection
Michel Bitbol & Claire Petitmengin

La psycho-phénoménologie, théorie de l'explicitation
Maryse Maurel

A Mathematician's View on Mathematical Creation
Pedro J. Freitas

On The Source of Mathematical Intuition
António Machiavelo

Kairos. Revista de Filosofia & Ciência

Kairos. Journal of Philosophy & Science

ISSN: 1647-659X

Edição: Centro de Filosofia das Ciências da Universidade de Lisboa

Índice / Index

Resumos / Abstracts 5

Artigos / Papers

Nietzsche: Science and Truth
Danny Smith 13

Arqueología: arte, historia, antropología.
Análisis filosófico de la génesis y desarrollo de una disciplina
Anna Estany 27

Freud, a concepção do descentramento e a Física Moderna
Lino Machado 49

Understanding Admissibility
George Masterton 71

Truth and Historicism
in Kuhn's Thesis of Methodological Incommensurability
Marco Marletta 91

Are Colors Real?
Emiliano Boccardi 111

Dossier: Introspection and Intuition in Mathematics

Introduction: *Where experience matters*
Alexandra Van-Quynh 159

Les possibilités de l'expérience. Mathématiques, aperception pure et aperception empirique dans la *Critique de la raison pure* de Kant
Matthieu Haumesser 161

On the possibility and reality of introspection
Michel Bitbol & Claire Petitmengin 173

La psycho-phénoménologie, théorie de l'explicitation
Maryse Maurel 199

A Mathematician's View on Mathematical Creation
Pedro J. Freitas 213

On The Source of Mathematical Intuition
António Machiavelo 223

Resumos / Abstracts

Nietzsche: Science and Truth
Danny Smith

This paper uses some of Lacan's concepts to discuss a number of questions regarding Nietzsche's writings on the problem of science, particularly focusing on the relationship between science and truth. Nietzsche's position, it is argued, is a deeply paradoxical one, entailing both a strong antirealism (Nietzsche takes all scientific statements to be 'false'), *and* an understanding that the scientific discourse nevertheless has massive and undeniable consequences in the real world. How is it possible to bring together these two seemingly incompatible views? Nietzsche himself does not provide us with a detailed resolution of this problem, but, it is suggested, Lacan also holds this same position, and *does* provide us with a formal answer to this problem. Rather than following what Lacan would call a 'male' reading of Nietzsche's position, in which his theory of truth provides us with a 'meta'-perspective, it is argued that Nietzsche ought to be read according to the 'feminine' side of Lacan's formulas. Rather than taking scientific statements to be 'false' because they do not perfectly capture whatever ultimately *is* true, Nietzsche and Lacan instead question the idea that truth itself is something 'out there' in a fully constituted state, waiting to be 'captured' by the scientific discourse. Instead, they both argue for a picture of reality as 'not-whole', as constitutively incomplete, which in turn allows them to explain how it is that the 'false' scientific discourse nonetheless succeeds in creating its 'real' object.

Arqueología: arte, historia, antropología.
Análisis filosófico de la génesis y desarrollo de una disciplina
Anna Estany

Este artículo se sitúa en el campo de la filosofía de la arqueología o meta-arqueología. El objetivo es el análisis de los principales cambios ocurridos en la arqueología en su constitución como disciplina científica. Para ello vamos a centrarnos, en primer lugar, en la evolución de la arqueología desde la etapa de los anticuarios hasta su constitución como ciencia social pasando por la etapa histórica; en segundo lugar, vamos a examinar el cambio que supuso el paso de la arqueología tradicional a la llamada "Nueva Arqueología" o arqueología procesal surgida en la década de los sesenta, y, finalmente, analizaremos el surgimiento de la arqueología post-procesual, haciendo un balance de la controversia entre los dos enfoques.

Lo que intento defender en este trabajo es que el paso de la arqueología tradicional a la Nueva Arqueología fue un cambio que podemos llamar "revolución metodológica" y que, por tanto, dicho enfoque no ha fracasado a menos que estemos dispuestos a renunciar al estudio sistemático y científico del pasado. Por tanto, la arqueología post-procesual no constituye un nuevo paradigma que pueda sustituir la Nueva Arqueología.

Freud, a concepção do descentramento e a Física Moderna
Lino Machado

A noção filosófica de descentramento é oriunda de textos de Sigmund Freud de 1916-17, baseando-se na física, na biologia e estendendo-se à psicanálise; todavia, os "golpes cosmológico", "biológico" e "psicológico" no narcisismo humano, que ele atribuiu respectivamente a Copérnico, à dupla Darwin-Wallace e à psicanálise (vale dizer, ao próprio Freud), precisam ser revistos, pois o primeiro se sustenta numa cosmologia apenas tridimensional (copernicana e não quadrimensional-relativística), o segundo, numa visão biológica ainda sem mecanismos quânticos, que, aos poucos, estão sendo descobertos, e o terceiro, numa concepção da psique que a isola demais do universo. Ao invés de um descentramento freudiano do sujeito, precisamos pensar num *modelo multicentrado* (quadrimensional) da existência como um todo, nisto incluída a subjetividade, sem retroagir a concepções da psique ultrapassadas por Freud e seus seguidores. As argumentações sobre o descentramento de Michel Foucault, Jacques Lacan e Jacques Derrida (sobretudo o último, que explicitou o termo) também são examinadas no artigo.

Understanding Admissibility
George Masterton

Lewis' concept of admissibility was introduced as an integral part of his famous Principal Principle; the principle that initial rational/reasonable belief should conform to objective chance unless there is evidence to the contrary. At that time Lewis offered only the rough and ready characterization that evidence not to the contrary of such dependence is admissible. This, together with some sufficiency conditions, served well enough until it became clear that admissibility was central to debates on the viability of Humean Supervenience and the analysis of objective chance. In response, Thau and Lewis refined the concept of admissibility in various ways. Since the mid 90's those who have employed the concept have, with minor variations and additions, followed the Thau/Lewis line. Yet, in the 30 years since its introduction what has been all too conspicuous by its absence is a full formal definition of admissibility and its degrees. Herein a family of definitions – all in terms of screening off by chance – that capture much that has been agreed about admissibility are proposed and evaluated; one of which is ultimately found to be serviceable as a definition schema for relative admissibility and its degrees.

Truth and Historicism in Kuhn's Thesis of Methodological Incommensurability

Marco Marletta

Methodological incommensurability is a Thomas Kuhn's thesis affirming that there are no shared, objective methodological rules or neutral scientific standards for theory comparison and choice. This thesis has often been interpreted as a relativistic and irrationalist claim on the incomparability of scientific theories. Since every paradigm refers to its standards, problem-field and aims, theory choice is subjective and arbitrary. Moreover it seems that, in his latest works, Kuhn abandons this aspect of incommensurability to focus on semantic incommensurability. On the contrary I will argue against the interpretation of methodological incommensurability as a source of epistemological relativism. The relativistic feature of incommensurability, rather, must be looked for in Kuhn's skepticism on the concept of truth as correspondence. From this point of view methodological incommensurability is consistent with semantic incommensurability, because they are both rooted in the intra-theoretical nature of truth. According to Kuhn, incommensurability and truth are historical concepts. The rational explanation of scientific conviction change cannot aim to something above the historical situation and the concrete scientific practice (such as the correspondence between theory and reality): truth is not correspondence, but an historical function of scientific community's agreement. We can evaluate the accuracy, fruitfulness, consistency, scope or simplicity of a theory and make a rational decision; but none of these parameters can measure the theory likeness to truth. Theory choice is always a theory-theory match, not a theory- reality match.

Are Colors Real?

Emiliano Boccardi

In this paper I argue that the properties that we represent in our color experiences should best be thought of as relational properties of physical objects and perceptual apparatuses. In particular, I argue that color properties are the (physical) properties that instantiate the operators that projects the infinite-dimensional space of spectral reflectances onto the finite color spaces that organisms perceive. Colors, under this account, are objective, mind-independent (albeit relational) properties of the world.

Artigos / Papers

Nietzsche: Science and Truth

Daniel Smith
(University of Warwick)
online@danny-smith.com

Gutting has argued that the great French philosophies of the twentieth century can be read primarily as different responses to the problems opened up in Nietzsche's thought[1]. Derrida, Foucault and Deleuze have each written important texts on Nietzsche, in each case taking one of Nietzsche's concepts as the major springboard for their own work (the attempt to escape metaphysics, the procedure of genealogy, the philosophy of immanence). These French philosophers can be seen as taking up Nietzsche's challenge, developing further some of the paths only hinted at in his own work. For example, it could be claimed that Nietzsche's claim that 'the "apparent" world is the only world, the "true" world is just *added to it by a lie...*'[2] only finds its full philosophical expression in Deleuze: in Nietzsche it remains a provocative thought rather than a fully worked-out ontology. Lacan might seem at first to be an exception to this narrative: Lacan almost never mentions Nietzsche, and is always disparaging when he does. I will argue on the contrary that, despite Lacan's minimal engagement with Nietzsche, the two thinkers are much closer than they may appear, in ways which will consequently problematise the 'official' readings of Nietzsche (whether Anglo-American or 'poststructuralist').

Our aim will not be simply to describe Nietzsche's influence on Lacan, rather it will be to use Lacan's conceptual apparatus to re-read some of the 'difficult' or 'problematic' aspects of Nietzsche's thought. Nietzsche's thought, particularly on this topic, undergoes far-reaching changes over the course of

[1] Gutting, 2010, chapter 5: 'How they are all Nietzscheans'.
[2] Nietzsche, 1997,"'Reason" in Philosophy', 2. Where reference is made to Nietzsche's aphoristic texts, I refer to the number of the aphorism, rather than the page.

his writing. 'Science' at some points stands for something invaluable 'with regard to everything one will afterwards do'[3], and at others for mere dogmatic anti-aesthetic thinking[4]. The aim of this essay is not so much to give a textually accurate description of Nietzsche's views, but to explore some interesting open questions that emerge from his writings. For most of the quotes discussed, it would probably be possible to find in another text the opposite view being stated. But this need not overly concern us, since our focus here is on the examination and development of some of Nietzsche's concepts, working through some of the difficulties and paradoxes he has left us. Our specific focus will be the question of science: what does the famous anti-metaphysician think is the relationship between science and truth?

As is made absolutely clear by statements like 'physics... is only an arrangement and interpretation of the world'[5], Nietzsche is an *antirealist* with respect to science. Nietzsche has a number of reasons for taking up this position, but we will focus on one in particular: his belief that in science, 'we operate only with things that do not exist: lines, planes, bodies, atoms, divisible time spans, divisible spaces'[6]. Nietzsche's point is that these objects do not have any reality 'in themselves': there is no 'pure' line or atom in the Platonic sense. Again, Nietzsche has a number of reasons for saying this: his conception of truth as 'perspective'[7], his valorisation of flux and becoming over static being[8], his rejection of the idea that 'there are identical things'[9]. But, perhaps most importantly, it is his suspicion that 'it is still a *metaphysical faith* upon which our faith in science rests'[10] that prevents him from subscribing to this position. The straightforwardly realist belief that the objects studied by physics are the ultimate 'foundation' of reality cannot but remain a *metaphysical* idea. He sees it as an unjustified presupposition that, as Cox puts it, 'truth is "already there" waiting to be discovered'[11]. Nietzsche mocks the naive scientific man thus:

[3] Nietzsche, 1974, 256.
[4] Nietzsche, 2000, 18.
[5] Nietzsche, 2003, 14.
[6] Nietzsche, 1974, 112.
[7] See ibid. 354.
[8] See Nietzsche, 1996a, 16.
[9] See ibid. 19.
[10] Nietzsche, 1974, 344.
[11] Cox, 1999, 49.

> He has concluded that so far as we can penetrate here – from the telescopic heights to the microscopic depths – everything is secure, complete, infinite, regular, and without any gaps[12].

To assume that reality is 'out there' in a fully constituted state, with relations, identities, and mathematical structures already built into it is to anthropomorphise, to naively assume that our own human categories must also be valid for reality itself: 'to a world which is *not* our idea the laws of numbers are wholly inapplicable: these are valid only in the human world'[13].

In fact, according to Jean-Claude Milner, the revolution brought about by modern (i.e. post-Galileian) science allowed us to dispense with precisely this notion that reality is 'in itself' mathematically constituted[14]. The difference between ancient and post-Galileian science, Milner argues, lies in their respective understandings of 'nature': for ancient science, 'nature' designated 'the order of the world that exists independently of man's conventions', whereas for Galileo it needed only to designate 'the empirical object of science'[15]. Thus for the ancients (and, we could also say, according to spontaneous common sense), the object of science had to be 'really real': if we accurately describe how something functions, and can correctly predict how it will continue to function, then we understand what it is in its essence. Modern mathematical science, by contrast, only 'requires *the mathematisation of the object;* it does not require that the object be a mathematical essence'[16]. In other words, we can analyse something scientifically without necessarily committing ourselves to any ontological claims about how the object is 'in itself'. This is the reason for the proliferation of different 'regional' sciences: it is possible for disciplines like economics, anthropology and psychology to exist as legitimate sciences, even though very few people think their specific objects (laws of human behaviour) actually 'exist' in any straightforward sense. To put this yet another way, modern science no longer has to rely on the distinction between 'natural law' and 'conventional law' *(physis/thesis* – 'what is according to natural necessities and what is according to man's conventions'[17]). Because 'nature' now means simply 'the object of science', 'mere' human conventions can be made the object of inquiry just as easily as can 'real' physical phenomena.

[12] Nietzsche, 2010, 120.
[13] Nietzsche, 1996a, 19.
[14] Milner, 2002, cited in Chiesa, 2010, 163-164.
[15] Ibid. p. 163.
[16] Milner, 2002, 289, cited in ibid., 164 (my italics).
[17] Ibid.

This 'de-ontologised' idea of modern science looks like a promising step towards a Nietzschean 'antirealist' position. However, this conception still remains an *epistemology* of science, a historically-specific description of the break associated with the Galileian revolution, rather than an 'absolute' account of the relation between science and reality. Even if modern science allows us to create new disciplines, investigate new kinds of objects and so on, this does not tell us anything about the *metaphysical status* of the scientific object itself: as Milner suggests, we can 'do' science whilst remaining completely agnostic about the actual nature of the object under consideration (this is why, unlike in pre-Galileian times, science and metaphysics are able to function completely independently of one another). As famously scathing as he was about metaphysics, Nietzsche has arguably even harsher words to say about any philosophy which remains at the level of epistemology. He calls it a 'timid epochism and abstinence doctrine', which does not even have the boldness to 'get over the threshold' and 'painfully *denies* itself the right of entry' to the proper philosophical questions: it is 'philosophy at its last gasp, an end, an agony, something that arouses pity'[18]. In other words, even an intricate study of the scientific method will remain 'merely' epistemological: what is needed here is an account of the relationship between science and 'the real'.

The Reality of the Illusion

What, then, is Nietzsche's metaphysical position *vis-a-vis* science? In order to answer this question, we will have to take a brief detour through Nietzsche's conception of truth, first asking the related question: how does Nietzsche conceptualise the *truth* of scientific statements? The most interesting text, for our purposes, is *On Truth and Lies in an Extra-Moral Sense*, where Nietzsche develops his position in a very unexpected (and often missed) direction. At the end of the sentence where Nietzsche famously states that 'truths are illusions we have forgotten are illusions'[19], he makes the following interesting analogy:

[18] Nietzsche, 2003, 204.
[19] Nietzsche, 2010, 117.

> [truths] are metaphors that have become worn out and have been drained of sensuous force, coins which have lost their embossing and are now considered as metal and no longer as coins[20].

Nietzsche's point is clear: coins 'in themselves' are simply round pieces of metal. However, once we endow them with the property of being coins (marked by their embossing), they cease to be mere pieces of metal and 'magically' become money, the universal equivalent. This takes place not because of any inherent property, but only through (what on Marx's analysis is) a *'salto mortale'*[21], a 'leap of faith' on the part of the users. As a result of this 'false semblance', there nevertheless comes about a real change in the way money functions: this is the famous analysis in chapter 4 of *Capital* where the original process of C-M-C transforms itself into the apparently 'irrational' M-C-M. Where money originally functioned in a simple, 'utilitarian' way to ease transactions in a barter system, it eventually comes to have a logic and dynamics of its own, completely independent of this original function. Nietzsche is of course not referring to Marx's theory of money, but it is a useful analogy; Nietzsche is making exactly the same point about the functioning of language. On Nietzsche's evolutionary account, language first arose because it was 'useful'; man does not have 'horns' or 'sharp teeth' like the other animals and so instead used language to better his chances of survival[22]. This original creation is, as in Marx, a seemingly 'magical' moment, which, once it has taken place, brings about all the 'metaphysical subtleties and theological niceties'[23] that Nietzsche spends the rest of his essay exposing.

To push the similarity further, we can say that, just as the illusions brought about by our 'false' understanding of money have a massive material effect in the world, according to Nietzsche, the 'errors' that are generated by our systematic misunderstandings of the nature of language also bring about real material changes. As Nietzsche puts it in a crucial passage (which strongly resonates with Milner's description of modern science):

> Everything which distinguishes man from the animals depends on this ability to volatilize perceptual metaphors in a schema, and thus to dissolve an image into

20 Ibid.

21 Marx, 1990, 200.
22 Nietzsche, 2010, 115, but see also Nietzsche, 1974, 110-111.
23 Marx, 1990, 163.

> a concept. For something is possible in the real of these schemata which could never be achieved with the vivid first impressions[24].

This formulation emphasises the deeply paradoxical nature of Nietzsche's antirealism. Certainly, words are always falsifications, never truly capturing the essence of things. However, this inevitable transformation of an original image into a 'false' concept nevertheless opens up an entirely new space, marked by Nietzsche with his curious phrase 'the real of these schemata'. This new space is, according to Nietzsche, nothing other than *human culture itself,* that which distinguishes us from the animals. Nietzsche evidently does not think that this conceptual view totally 'captures' the real (he certainly thinks that it engenders a number of dangerous beliefs which he spends most of the rest of his essay denouncing), but, and this will turn out to be an important formulation, this conceptualisation does nevertheless *have real consequences*.

We see a fuller development of this idea in Gay Science 58, which Nietzsche begins by proclaiming that 'what things are *called* is incomparably more important than what they are'[25]. In this aphorism, Nietzsche once again states a very strong antirealist position; his starting point is that the name of a thing as well as the usual properties we assign it are merely conventional, 'thrown over things like a dress and altogether foreign to their nature'. In other words, Nietzsche is against Russell's descriptivism (which connects the name of a thing to the description that speakers would give of it), and for Kripke's thesis that a name is bestowed in an 'initial baptism', subsequently acting as a 'rigid designator', so that the name remains even if all the predicates we previously associated with the thing change[26]. The name does not point to the essence of the thing, only to the tautological fact that it has been named in such a way. Nietzsche's next point is that this initially 'false' designation gives rise to a movement where 'what at first was appearance becomes in the end, almost invariably, the essence, and is effective as such'[27]. However, just as we saw before, Nietzsche's position is considerably more interesting than *just* being anti-essentialist: he asks us 'how foolish it would be to suppose that one only needs to point out this origin and this misty shroud of delusion in

[24] Nietzsche, 2010, 118.
[25] Nietzsche, 1974, 58.
[26] Kripke, 1981, 96-97. For further elaboration of this point, see the classic discussion in Žižek, 2008, 97-101.
[27] Nietzsche, 1974, 58.

order to *destroy* the world that counts for real, so called *"reality"*[28]. We can point out reifying, essentialising illusions as much as we like, but this does not stop them from continuing to function. Nietzsche thus concludes with the insight that 'it is enough to create new names and estimations and probabilities in order to create in the long run new "things"'[29]. Far from unconditionally denouncing the 'false' world, Nietzsche suggests that we make creative use of this peculiar feature of language in order to bring about effects which are 'really real'.

Truth is a Woman (who does not exist)

How are we to understand this paradoxical position of Nietzsche's? On the one hand, he tells us that 'reality' is an illusion: we are consistently led by language to group unlike things together, thereby giving ourselves the false impression that they have an underlying essence. On the other, he fully understands that one cannot simply 'denounce' reality; even as 'false' it exerts a certain efficiency on whatever ultimately *is* real. We come to a better understanding of these issues, I would suggest, by reading Nietzsche along the lines of Lacan's formulas of sexuation[30]. Nietzsche's position is usually read in what Lacan would have called a 'male' way, that is, as making a universal claim to which there is a constitutive exception. This is the well-known problem of self-reference in Nietzsche's conception of truth, nicely formulated by Clark: 'if it is supposed to be true that there is no truth, then there is apparently a truth after all'[31]. One can only universalise the claim that 'there is no truth' by allowing for an exception, which is this statement itself. The usual Nietzschean response to this apparent contradiction is to appeal to different 'levels' of truth: for Clark, Nietzsche rejects 'the existence of

[28] Ibid.
[29] Ibid.
[30] Lacan's enigmatic formulae are as follows ('male' is on the left, 'female' is on the right):

For the original formulation, see Lacan, 1999. Whilst there is an important connection between this table and sexual difference, for our purposes we will only be considering the logic of Lacan's formulae and not their connection to the 'male' and 'female' subject-positions.
[31] Clark, 1990, 3.

metaphysical truth... but not truth itself'[32]; for Danto, Nietzsche rejects the correspondence theory of truth, but accepts a 'pragmatic' theory[33]; for Schacht, Nietzsche assigns to his own writing a 'meta-level perspective'[34] from which he is able to pronounce the *real* truth.

Nietzsche's own position, I would argue, follows rather the 'feminine' logic of the 'not-whole' [*pas-tout*]. With his 'perspectivism', Nietzsche seems to be trying precisely to *avoid* adopting the kind of 'meta'-standpoint described above[35]. His position is exactly the opposite one: 'perspectivism' means that truth is always absolutely *immanent* to a situation, what is prohibited is precisely any kind of appeal to a different or 'higher' level. Such an appeal would miss Nietzsche's point that even the 'immediate' presentation of a situation is always minimally subjectively mediated: there can be no 'pure' datum of experience which has not always-already been interpreted. His point is thus not so much 'there is no such thing as truth' as 'it is never possible to tell the truth of a situation from outside of that situation'. There is, then (according to Lacan's formulae), no exception to this rule: *every* situation finds its 'truth' from an engaged perspective within it, but, and for this very reason, it is impossible to know the 'full' truth. The truth that one grasps is 'not-whole', it can only ever be partial and incomplete[36]. But, and this is crucial, this is not because of any epistemological limitation. What Nietzsche aims at in his 'perspectivism' is the idea that the concept of a 'full' truth which 'says it all' is a *metaphysical* impossibility. Unlike on the 'male' side, there is no point of exception from which the whole truth can be spoken: *the absolute itself* is lacking, inconsistent, incomplete. Lacan's own version of this point is, of course, his notorious dictum that 'Woman doesn't exist' [*la femme n'existe pas*][37].

---

[32] Ibid. 21.
[33] Danto, 1965, discussed in Clark, 1990, 31-34.
[34] Schacht, 1983, 10, cited in Clark, 1990, 152.
[35] I rely here on the analysis in Zupančič, 2003.
[36] The usual example of the 'not-whole' is late Wittgenstein, whose understanding is similar to Nietzsche's on this point. With his concept of 'language games' we do indeed have an absolutely universal account of language. The mystical exception famously mentioned at the end of the 'masculine' Tractatus has disappeared; there is no utterance which cannot be understood as a 'language game', no matter how mysterious it may appear. However (and for that very reason) what is stated remains 'not-whole': language describes partial connections and 'family resemblances', not 'the absolute truth', as it did in Wittgenstein's earlier work.
[37] See Lacan, 1999. A formula which, to return for a moment to the problematic of sexual difference, should of course be read alongside Nietzsche's infamous claim that

This insistence on the *metaphysical* status of our partial knowledge of truth is, as Žižek suggests[38], the key difference between the usual 'post-structuralist' position and Lacan: the difference lies in how to understand the claim that 'there is no metalanguage'[39] (Lacan's version of 'perspectivism'). For a 'deconstructivist', this would mean that there is no 'pure' literal meaning in a text; it will always contain elements that destabilise it, that undermine any final interpretation. Žižek criticises this position thus:

> the position from which the deconstructivist can always make sure of the fact that 'there is no metalanguage', that no utterance can say precisely what it intended to say, that the process of enunciation always subverts the utterance, *is the position of metalanguage in its purest, most radical form*[40].

In other words, the typical post-structuralist appropriation of Nietzsche remains squarely on the 'male' side: 'there is no metalanguage' is taken precisely as a metalinguistic statement, rather than, as Lacan has it, designating that the field of language is incomplete, incapable of being totalised because it does not have a full, positive reality to begin with.

A Discourse with Consequences

What, then, does this 'feminine' understanding of truth mean for our conception of science? As Nietzsche puts it in a crucial aphorism:

> It is a profound and fundamental good fortune that scientific discoveries stand up under examination and furnish the basis, again and again, for further discoveries. After all, this could be otherwise. Indeed, we are so convinced of the uncertainty and fantasies of our judgements and of the eternal change of all human laws and concepts that we are really amazed how *well* the results of science stand up[41].

It is clear that, for Nietzsche, scientific formulations are in some basic sense 'false'. This is a problem that any conception of science will have to deal with at some point: how to account for the efficiency of 'wrong' theories. If theory X is superseded by theory Y, we nonetheless still require some explanation of the previous efficacy of theory X. The usual approach is to

---

'Truth is a Woman'. The combination 'Truth is a Woman (who does not exist)' could serve as the basic formula for our Lacanian reading of Nietzsche.

[38] Žižek, 2008, 172.

[39] See e.g. Lacan, 2006, .688.

[40] Žižek, 2008, 173 (my italics).

[41] Nietzsche, 1974, 46.

show how these old theories nevertheless correctly grasped *some* aspect of the real. This presupposes, however, that there is an underlying 'absolute truth' of the matter, of which our theories are only better or worse approximations. As we have seen in his theory of truth, Nietzsche's position is much more radical: he repeatedly denies the existence of *any* metaphysics which grounds 'our world' of change and becoming in 'another' stable, unchanging world[42]. He is thus faced with the problem raised in the quote: how are we to explain the undeniable efficiency of the apparently 'false' scientific discourse?

Lacan agrees with Nietzsche's basic antirealist standpoint: modern science 'posits' rather than 'discovers' the reality it works with[43]. Lacan states, for example, that: 'energy is not a substance... it's a numerical constant that the physicist has to find in his calculations, so as to be able to work'[44]. 'Energy' is a model we use in order to understand the results of scientific experimentation: it is a discursive formation, not a material thing. This scientific object, then, is 'a fact experimentally produced by a theory'[45], rather than something which pre-exists the theory. However, as in Nietzsche, this does not necessarily lead to the 'postmodern' relativist position, where science is taken to be just one discourse among others: unlike many of his contemporaries, Lacan does think there is something unique about the discourse of science. Following the work of the French epistemologists, Lacan sees mathematisation and formalisation as the most important aspect of modern science, much more so than the focus on experimentation usually highlighted by the Anglo-American tradition. In stark contrast to a traditional British empiricist view, Lacan sees the key breakthrough of modern science in its ability precisely to 'allow oneself a free-fall from any recourse to evidence'[46]. The ability to reduce the richness of experiential data to a letter or a number means, in Lacan's terminology, that science no longer needs to be subject to 'imaginary capture': science is able to function perfectly well *even when its object cannot be thought*. One can simply 'do the maths' and obtain the correct result without having to have any mental representation of what it 'means'.

[42] See e.g. Nietzsche, 1997, 'Reason in Philosophy', 2, Nietzsche, 1974, 344.
[43] The following analysis of Lacan relies heavily on the interpretations presented in Nobus, 2002, Fink, 1995, Verhaege, 2002, and especially Zupančič, 2011.
[44] Lacan, 1990, 18.
[45] Ibid.
[46] Lacan, 1990, 39 (my italics).

A good example of this is the infamous number *i*, the square root of -1. If there is *anything* which can be said not to exist, this is surely it: a number whose impossibility is built into its very *definition*. And yet, even though it is nothing but a fictitious and 'false' construction, nobody could seriously deny the material efficiency of this purely symbolic entity. Even though it 'doesn't exist', we can nonetheless use this number in calculations which allow us to build buildings. Even if its referent is 'false' in some sense, the discourse it is a part of literally changes the real, material world:

> scientific discourse was able to bring about the moon landings, where thought becomes witness to a performance of the real... using no apparatus other than a form of language[47].

Modern science is for Lacan not the progressive unfolding of the absolute truth, but a historical event, something which emerged at a particular moment in time. This event nonetheless opened up a new space, and this is the central point of Lacan's conception of modern science: it is a form of discourse *which has real consequences*.

This aspect of Lacan's thought, I would suggest then, is an attempt to *formalise* a basic ontological conception of science which he shares with Nietzsche. Nietzsche, as we have seen, is thoroughly sceptical about the real existence of the scientific object, but does not for a moment question that the discourse of science has 'real' effects. On the one hand, we know that science doesn't present us with 'the real as such'. However, we also cannot deny that nature does at least seem to follow the laws we posit with some regularity. In another crucial quote, Lacan deals with this problem of the relation between science and nature:

> We cannot resist the idea that nature is always there whether we are there or not, we and our science, as if science were indeed ours and we weren't determined by it. Of course I won't dispute this. Nature is there. But what distinguishes it from physics is that it is worth saying something about physics, and that discourse has consequences in it, whereas everybody knows no discourse has any consequences in nature[48].

Zupančič immediately relates this quote to the anecdote of Hegel being dragged to the Alps by his friends: they wanted him to see the sublime grandeur of the mountains, and to reassess his thesis according to which only the products of human Spirit can attain real beauty. Hegel's ironic response

[47] Ibid., 36.
[48] Lacan, Seminar XVI second lesson (unpublished), cited in Zupančič, 2011.

was 'the sight of these eternally dead masses provokes nothing in me but the uniform and at length boring idea: it is [*es ist so*]'[49]. It is not that we can't understand the deep mysteries of nature, rather that *there simply is nothing there to understand:* 'it is' is all that can be meaningfully said. We can talk about the geological processes which formed the mountains, the chemical reactions which produced the different types of rock and so on, but then we have entered a different kind of discourse, a scientific one, one which precisely *does* have consequences.

Conclusion

We have seen, then, how it is that the scientific discourse produces its object in both Lacan and Nietzsche; this produced object is 'false' in the absolute sense, but it does have undeniable effects in the real (like allowing us to land on the moon). Nietzsche certainly does think that science 'falsifies' reality, but his position is, as we have seen, much more refined than the relativism of which he is often accused. Nietzsche was of course not interested in the kind of formalisation carried out by Lacan and his followers, he was no 'structuralist', but what I am suggesting is that Lacan's structuralism (or 'hyper-structuralism' as Chiesa designates it[50]) could be seen as a development of this aspect of Nietzsche's thought (just like deconstruction and genealogy are developments of other aspects). Even though Nietzsche announces the 'end of metaphysics', putting an end to all philosophies which aim to fully capture the absolute, this does not at all mean that we have to give up on 'the real as such'. Modern science, as we have seen, in its own way *produces* a new real, which, even if it is not fully complete or even 'correct', nonetheless functions. Of course, our investigation has been limited to the discourse of science: we have not dealt with the broader question of how we are to conceive of what ultimately *is* real. But is Nietzsche's anti-metaphysical position not opposed to precisely this kind of gesture? His opposition to atomism, for example, is not a result of his belief that there is some 'deeper' level of substance: if, as has been argued, Nietzsche thinks that reality *itself* is 'not-whole', then all such 'foundationalist' enterprises must be mistaken. This is precisely what Lacan's conception of science avoids: the presupposition that there is a true underlying reality of

[49] Hegel, 1997, 53.
[50] Chiesa, 2010, 159.

natural laws 'out there' waiting for us to discover them. If Nietzsche is right, then there can *only* be 'regional ontologies', different forms of discourse which somehow touch on the real; modern science may prove to be only one among many.

Bibliography

Lorenzo Chiesa, 'Hyperstructuralism's Necessity of Contingency', in *S: Journal of the Jan Van Eyck Circle for Lacanian Ideology Critique* (3), 159-177, 2010.

Maudemarie Clark, *Nietzsche on Truth and Philosophy*, Cambridge: Cambridge University Press, 1990.

Christopher Cox, *Nietzsche: Naturalism and Interpretation*, California: University of California Press, 1999.

Arthur Danto, *Nietzsche as Philosopher*, New York: Macmillan, 1965.

Bruce Fink, *The Lacanian Subject*, Princeton: Princeton University Press, 1995.

Gary Gutting, *Thinking the Impossible: French Philosophy since 1960*, Oxford: Oxford University Press, 2011.

Georg Wilhelm Friedrich Hegel, trans. Robert Legros and Fabienne Verstraeten, *Journal d'un Voyage dans les Alpes Bernoises (du 25 au 31 Juillet 1796)*, Grenoble: Millon, 1997.

Saul Kripke, *Naming and Necessity*, Oxford: Blackwell, 1981.

Jacques Lacan, trans. Bruce Fink, *Écrits*, London: Norton, 2006.
--- trans. Bruce Fink, *Seminar XX*, London: Norton, 1999.
--- trans. Denis Holler, Rosalind Krauss and Annette Michelson, ed. Joan Copjec, *Television*, London: Norton, 1990.

Darian Leader, 'The Not-All', in *Lacanian Ink* (8), 1994, 43-50.

Karl Marx, trans. Ben Fowkes, *Capital Volume I*, London: Penguin, 1990.

Jean-Claude Milner, *Le Périple Structural: Figures et Paradigme*, Paris: Seuil, 2002.
Friedrich Nietzsche, trans. Brittain Smith, *Dawn*, Stanford: Stanford University Press, 2011.
--- trans. Daniel Breazeale, 'On Truth and Lies in a Nonmoral Sense', in *The Nietzsche Reader,* Oxford: Blackwell, 2010, 114-123.
--- trans. R. J. Hollingdale, *Beyond Good and Evil,* London: Penguin, 2003.
--- trans. Douglas Smith, *The Birth of Tragedy,* Oxford: Oxford University Press, 2000.
--- trans. Richard Polt, *Twilight of the Idols*, Indianapolis: Hackett, 1997.
--- trans. R. J. Hollingdale, *Human, All too Human*, Cambridge: Cambridge University Press, 1996a.

--- trans. Douglas Smith, *On the Genealogy of Morals*, Oxford: Oxford University Press, 1996b.
--- trans. Walter Kaufmann, *The Gay Science,* New York: Vintage, 1974.
Danny Nobus, 'A Matter of Cause: Reflections on Lacan's Science and Truth', in *Lacan and Science* ed. Jason Glynos and Yannis Stavrakakis, London: Karnac Books, 2002, 89-118.

Richard Schacht, *Nietzsche*, London: Routledge and Kegan Paul, 1983.

Paul Verhaege, 'Causality in Science and Psychoanalysis', in *Lacan and Science* ed. Jason Glynos and Yannis Stavrakakis, London: Karnac Books, 2002, 119-145.

Slavoj Žižek, *The Sublime Object of Ideology,* London: Verso, 2008.

Alenka Zupančič, 'Realism in Psychoanalysis', in *Journal of European Psychoanalysis* (32:1), 2011 accessible at:
http://www.psychomedia.it/jep/number32/zupancic.pdf
--- *The Shortest Shadow: Nietzsche's Philosophy of the Two*, Massachusetts: MIT Press, 2003.

Arqueología: arte, historia, antropología. Análisis filosófico de la génesis y desarrollo de una disciplina[1]

Anna Estany
(Departamento de Filosofía, Universitat Autònoma de Barcelona)
Anna.Estany@uab.es

1. Introducción

La arqueología es la depositaria de nuestra memoria colectiva. Uno de los deseos más arraigados en nuestra especie es la explicación del mundo, el otro deseo es la curiosidad por conocer quiénes eran y cómo eran nuestros ancestros. "Arqueología" es el término acuñado por la cultura occidental para referirse a todo el conocimiento sobre nuestros antepasados y su cultura, entendiendo ésta en su sentido más amplio. Pero la arqueología, aún cuando ha seguido con la misma denominación, ha sufrido profundos cambios a lo largo del siglo XX. Sigue teniendo el mismo objeto de estudio pero todo lo demás es distinto: la forma de abordar dicho objeto, los grupos interesados en ello, los instrumentos utilizados y los objetivos a largo plazo.

El objetivo de este trabajo es analizar los cambios más significativos que ha experimentado la disciplina en el siglo XX, centrándonos en los siguientes puntos: la evolución de la disciplina, entroncada primero con el arte, después con la historia y, finalmente, con la antropología; la arqueología procesual conocida como "Nueva Arqueología"; la arqueología post-procesual y su crítica a la "Nueva Arqueología".

[1] Este trabajo se enmarca en el proyecto financiado por el Ministerio de Ciencia e Innovación de España “Innovación en la práctica científica: enfoques cognitivos y sus consecuencias filosóficas (Referencia FFI2011-23238). Además, este trabajo es resultado del trabajo del grupo consolidado y reconocido por la Generalitat de Catalunya (España) “Grup de Estudios Humanísticos sobre Ciencia y Tecnología” (GEHUCT).

Desde el punto de vista del análisis filosófico de la dinámica científica y partiendo de la evolución que ha experimentado la arqueología propongo las siguientes hipótesis de trabajo que argumentaré a lo largo de este trabajo:

La "Nueva Arqueología" (NA) supuso un cambio significativo del tipo en que la metodología es el motor del cambio y, por tanto, la que determina el campo de acción y las líneas de investigación.

Una buena parte de los autores de teoría arqueológica desarrollada en los noventa argumenta que la arqueología procesual fue una "moda" de los sesenta pero fracasó y que las nuevas tendencias de los ochenta han acabado con ella. En este trabajo propongo una interpretación distinta del surgimiento de la arqueología post-procesual teniendo en cuenta criterios epistemológicos. Desde un punto de vista de la práctica científica, si se abandonara la arqueología procesual se debería pagar un precio muy elevado, algo a lo que los arqueólogos no parecen estar dispuestos si nos atenemos a su trabajo de campo.

Si la arqueología procesual fue, fundamentalmente, una revolución metodológica, dicha arqueología no ha sucumbido, al menos no en sus rasgos más esenciales. Se han abandonado algunos esquemas metodológicos concretos pero subsisten los principios básicos que subyacen a toda investigación científica. Al menos subsisten para aquellos arqueólogos empeñados en explicar la sociedad de nuestros antepasados.

2. Génesis de la arqueología

Al abordar el análisis filosófico del desarrollo de la arqueología surgen las preguntas de qué es la arqueología y cuáles son los fines de la misma. En sentido general podemos decir que es el estudio del pasado de los humanos, pero la perspectiva puede ser radicalmente distinta: desde descubrir aspectos maravillosos del pasado – objetivo de la etapa de los anticuarios y directamente ligada al arte – hasta explicar el pasado – objetivo de la arqueología actual – pasando por la reconstrucción del pasado – objetivo de la etapa histórica. Para el tema que nos ocupa vamos a centrarnos en las etapas histórica y científica pero vamos a hacer una incursión a los orígenes de la disciplina vinculados al arte[2].

[2] Quiero señalar que la referencia a estas tres etapas de la arqueología no tiene como objetivo hacer un estudio exhaustivo de la historia de la arqueología, sino proporcionar las principales características de las diversas etapas por las que ha pasado esta disciplina.

2.1. Arte

La arqueología como disciplina académica nació hace poco más de cien años pero como actividad de "hurgar" en el pasado tenemos datos de mucho antes de nuestra era. Nabonidus, último rey de Babilonia (555-538 AC) estaba muy interesado por el pasado de la cultura babilónica y llevó a cabo una serie de excavaciones construyendo un museo en el que se exponían todos sus descubrimientos (Hole y Heizer, 1973:41).

A finales del siglo XIV se inició una etapa denominada, a veces, "caza de tesoros" cuya finalidad principal era coleccionar objetos de arte y catalogarlos, más por interés personal que público. Esta labor se llevó a cabo por aventureros con aire romántico y movidos por el interés en la antigua Grecia y Roma. Italia fue especialmente importante durante el siglo XV en la actividad de buscar tesoros y tanto los papas como la nobleza decoraban sus casas con estatuas antiguas. Este interés se extendió por toda Europa. Los españoles en su conquista del Nuevo Mundo también realizaron numerosas excavaciones tal como relata uno los cronistas Fernández de Oviedo (Hole y Heizer, 1973: 42). Hay que señalar que la mayoría de estas excavaciones eran auténticos saqueos. En el siglo XVII muchos ingleses fueron al Mediterráneo en busca de tesoros para confeccionar sus propias colecciones. Uno de los más importantes fue Thomas Howard que visitó Italia. También se inició la búsqueda en otros lugares del suroeste de Asia.

En el siglo XIX las colecciones a gran escala surgieron de otros lugares como el valle del Nilo, Tigris y Éufrates. Pero también siguió la búsqueda en el suroeste asiático. Especial importancia tuvo el establecimiento del consulado en Bagdad en 1802 que marcó el inicio de la búsqueda de tesoros al suroeste asiático. Claudius Rich era un estudiante de lenguas y un político astuto (Hole y Heizer, 1973: 43) que ocupó la residencia británica de Bagdad durante veinticinco años. Cuando murió en 1821 había conseguido unos tres mil quinientos kilos de antigüedades. Desde mediados del siglo XIX los gobiernos británico y francés, viendo el gran tesoro que podía encontrarse, decidieron financiar las excavaciones retribuyendo económicamente a los que trabajaban en dichas excavaciones. De alguna forma había nacido el oficio de arqueólogo. Austein Henry Layard (británico) Paul Emile Botta (francés) son sólo una muestra de los individuos que se dedicaron a la búsqueda de tesoros financiados por sus gobiernos respectivos. Al final del siglo XIX, cuando los museos estaban repletos y las cabezas de los excavadores también, la arqueología empezó a preocuparse por la historia

de la zona donde se encontraban los restos arqueológicos (Hole y Heizer, 1973: 49). Empezaba a surgir la arqueología como historia de los pueblos del pasado y, por tanto, una nueva etapa en su establecimiento como disciplina académica.

2.2. Historia

El enfoque histórico tiene una preocupación por ordenar los acontecimientos pasados secuencialmente e interpretar los eventos como únicos, lo cual hace que cada hecho histórico sea distinto. La arqueología como historia, o arqueología prehistórica, tiene como objetivo investigar el pasado del hombre en aquellos periodos en que los documentos escritos son escasos o no existen. La falta de documentos históricos hace que los prehistoriadores recurran a los artefactos y, en general, al registro arqueológico del mismo modo que los paleontólogos recurren a los fósiles y los de historia geológica a los estratos geológicos. Pero esta circunstancia no los hace menos historiadores. El sentir de muchos arqueólogos de aquella época era que todos eran historiadores, con o sin texto escrito.

El enfoque histórico fue predominante en arqueología hasta finales de la década de los cuarenta en que, como veremos en el próximo apartado, se cuestiona el enfoque histórico como puramente descriptivo y en cierto modo como no-científico. Sin embargo, hay que señalar que la dicotomía historia/ciencia no se desvaneció con la implantación de la arqueología como ciencia de la cultura sino que muchos arqueólogos siguieron planteándose la relación entre arqueología e historia, aunque con un concepto de la disciplina histórica muy distinta de la que manejaban los pre-historiadores de principios del siglo XX.

2.3. Antropología

Los antecedentes de una arqueología entroncada con la antropología hay que situarla a finales de la década de los cuarenta con el surgimiento de voces que consideraban que hasta entonces la arqueología había sido un área de conocimiento dedicada exclusivamente a detalles de cronologías y a la distribución de rasgos o características del registro arqueológico. Walter Taylor criticó esta concepción en su obra *A study o archaeology* (1948) en la que se hace un análisis de las ciencias de la cultura, incluyendo la arqueología, la antropología y la historia, y dice que los arqueólogos, al menos hasta 1948, no han hecho otra cosa que coleccionar datos,

proponiendo la utilización de métodos científicos para la investigación arqueológica.

Otros precursores son Gordon Willey y Philip Phillips en su obra *Method and theory in american archaeology* (1958). Willey y Phillips dintinguen tres niveles de organización conceptual. El primer nivel corresponde al trabajo de campo y se mueve en el plano de la observación, siendo el producto de este trabajo el material obtenido en una excavación. El segundo nivel se mueve en el plano de la descripción y corresponde a la integración histórico-cultural que consiste en la organización de los datos primarios, a saber: tipología, formulación de la unidades arqueológica etc. Y el tercer nivel corresponde a la explicación. Señalan estos autores que se ha hecho tan poco en el tercer nivel que difícilmente puede hablarse de explicación en arqueología.

Y llegamos a la década de los sesenta. ¿Qué pasó en este "década prodigiosa", no sólo para la arqueología? Según muchos arqueólogos una revolución, un cambio de paradigma en sentido kuhniano. R.A. Watson (1972) dice que lo que se dijo en los 40 se hizo en los 60 y espera (esto lo dijo en 1972) que en los 70 se encuentren definitivamente la teoría y la práctica. La arqueología surgida del cuestionamiento de la etapa descriptivo-histórica, denominada “Arqueología Tradicional” (AT), es la "arqueología procesual", normalmente denominada "Nueva Arqueología" (NA). La NA no cuenta con un texto referencial global de su propuesta teórico-metodológica, sino que este cuadro general se fue construyendo, fundamentalmente, a partir de numerosos artículos desde 1962 a 1972. Entre las figuras más importantes destacan L.R. Binford, K.V. Flannery, J.N. Hill, P.J. Watson, S.A. LeBlanc, Ch.L. Redman, J.M. Fritz y F.T. Plog. Sin embargo, el artículo de Binford en 1962 "Archaeology as Anthropology" en *American Antiquity* se considera como el punto de partida y, en cierto sentido, el manifiesto de la NA. Podría decirse que Binford sintetizó las nuevas ideas y críticas que se habían ido gestando durante las dos últimas décadas[3].

El hecho de que lo que se consideró un manifiesto de la NA fuera publicado en *American Antiquity* y buena parte de las principales figuras estuvieran en universidades de Estados Unidos no significa que la NA sea una corriente limitada a este país. En realidad se expandió e influenció la arqueología en general. Sin embargo, hay que señalar que la crítica a la AT no se encauzó solamente a través de la NA, sino que, entre otras corrientes, está la surgida en América Latina y que tomó el nombre de “Arqueología

[3] Dejo abierta la cuestión de si Binford es a la arqueología lo que Newton es a la física, Lavoisier a la química y Darwin a la biología.

Social Latinoamericana" (ASL), en un intento de aplicar en su investigaciones una teoría y una metodología basadas en el Materialismo histórico, y con nombres como L. Bates, L. Lumbreras y M. Sanoja, entre otros. El caso de Manuel Gándara también forma parte de los arqueólogos latinoamericanos que cuestionaron la AT a fin de darle estatus científico, aunque su postura respecto al modelo teórico va más allá del materialismo histórico, centrándose en un modelo teórico que pueda fundamentar la arqueología como ciencia, como muestra al señalar la necesidad de "un conjunto de supuestos valorativos, ontológicos y epistemológico-metodológico, que guían el trabajo de una comunidad académica particular, y que permiten la generación y el desarrollo de teorías sustantivas." (Gándara 1993: 8). Hay que tener en cuenta que así como la ASL surge en los años sesenta, parte del trabajo de Gándara se lleva a cabo en las décadas de los ochenta y noventa.

3. La arqueología histórica vs. la arqueología procesual

Con el surgimiento de la "Nueva Arqueología" la arqueología histórica pasó a ser considerada como la "Arqueología Tradicional" (AT). M. Leone (Leone, 1972) ha apuntado que la identidad de la AT es más una consecuencia de la caracterización de la NA que de la propia identidad de la AT, es decir, que el viejo paradigma se identifica por contraposición al nuevo. Vamos a señalar las características más relevantes de la NA a partir de los trabajos de Binford como uno de los representantes de esta corriente.

La NA rechaza el enfoque puramente empiricista o inductivista estrecho. Binford califica de metafísica la premisa, propia de la AT, de que la causa de la variabilidad en los utensilios hay que buscarla en la variabilidad de la identidad social de sus productores y de que las causas de la identidad social de los pueblos hay que buscarla en la historia (Binford, 1983:4). Según Binford, estas premisas van parejas a la idea de F. Boas (Boas 1966:273) de que puede haber cosas similares que tengan significados distintos para pueblos distintos. Según Boas la investigación antropológica no puede presuponer que los fenómenos etnológicos se han desarrollado de la misma forma y, por tanto, no tiene sentido que se intente descubrir las leyes históricas universales. Boas dice: "Aquí reside el defecto del argumento del nuevo método, ya que no es posible dar la prueba que dice dar. Incluso el más superficial de los informes muestra que los mismos fenómenos pueden desarrollarse en múltiples formas" (Boas, 1966:273). Frente al estricto

empirismo de Boas y a las inferencias puramente inductivas, los arqueólogos de la NA argumentan que la disciplina debería adoptar el método científico y tomar la inferencia deductiva como forma de razonamiento. Otra cuestión importante es lo referente a la utilización de analogías. La NA no atribuye a la analogía el papel que la AT le asigna en la interpretación del pasado. La postura de Binford es clara: "mientras la analogía sea el instrumento para justificar las interpretaciones del pasado, la arqueología adolecerá de métodos apropiados para hacer afirmaciones rigurosas sobre el pasado" (Binford, 1983:8).

La diferencia más importante entre la AT y la NA es en lo referente a la explicación. Mientras la AT crea el pasado para explicar el registro arqueológico actual, la NA exige que para explicar un evento, éste pueda ser insertado en un cuerpo de conocimiento más general. Estas diferencias en cuanto a la explicación quedan reflejadas en las críticas que Binford hace a posturas como las de J.A.Safloff y G.R. Willey (1967) por primar el enfoque histórico y, en consecuencia, la descripción frente a la explicación. Para Binford toda esta literatura escrita bajo el enfoque histórico no es más que exposiciones descriptivas de nuestro conocimiento del registro arqueológico y no resúmenes de nuestro conocimiento del pasado. El objetivo último de la arqueología es explicar el pasado y el enfoque histórico sólo lo describe ya que se ocupa de lo ideográfico o particular, por oposición a lo gnomotécnico o general. Por ejemplo, para Sabloff y Willey (1967) el colapso de la civilización Maya es atribuido a una invasión, por tanto, es una acontecimiento histórico el que da cuenta de dicho colapso. Pero Binford, entre otros, dice que esto no es una explicación, ya que para que hubiera una explicación, en primer lugar, habría que demostrar que hubo una invasión, en segundo lugar, la invasión tendría que explicar el colapso de los Maya, y finalmente, si la invasión tiene que explicar el colapso de los Mayas tienen que haberse confirmado leyes generales sobre procesos culturales, de las que el ejemplo de los Maya sería una instancia. Sin embargo, hasta el momento ninguna ley de este tipo ha sido confirmada. Por tanto, a fin de explicar lo que ocurrió con los Maya la prioridad tiene que ser la confirmación de estas leyes procesuales, no la reconstrucción histórica como piensan Sabloff y Willey. Mientras tanto, dice Binford, no tenemos explicación[4].

[4] Hay que señalar que la crítica de Binford a la arqueología como historia se hace desde una concepción de la historia como pura descripción de acontecimientos, concepción cuestionada por diversas corrientes historiográficas que quieren incorporar en la investigación histórica los resultados de las ciencias sociales. Un

Binford reconoce sus débitos a Taylor pero su obra es un monumento al trabajo de C. G. Hempel (1979). La NA consiste en la aceptación del modelo de ley cubriente con énfasis en el método hipotético-deductivo para confirmar hipótesis formuladas a partir de los datos arqueológicos. Según Binford, la AT interpreta los rasgos o características en un vacío teórico, explicando las diferencias y similitudes entre rasgos como resultado de la armonización, de las influencias direccionales y de la estimulación entre tradiciones históricas. Frente a la AT Binford propone que estas explicaciones sean en términos de nuestro conocimiento de las características estructurales y funcionales de los sistemas culturales.

Según Binford, el objeto de estudio de la arqueología no es la conducta humana, ni los códigos simbólicos, ni los sistemas sociales, ni las culturas antiguas, ni el pasado, sino los "artefactos". El arqueólogo estudia los artefactos en tres dimensiones: forma, espacio y tiempo. Todo lo que digamos sobre la conducta de los pueblos antiguos, de los sistemas sociales etc. es una inferencia a partir de los artefactos, que son la única evidencia arqueológica que poseemos a partir de la cual construimos hipótesis que luego hay que confirmar. Los artefactos son datos culturales, elementos de un sistema cultural.

Binford no sólo tiene desavenencias con el enfoque histórico, sino también con algunos de sus más inmediatos precursores de la NA, tales como Willey, Phillips, Taylor, Ford, Rouse, etc., pertenecientes a la denominada "escuela normativa". Todos, incluido Binford, están de acuerdo en que el sujeto de la arqueología es la cultura, pero el desacuerdo está en la definición de las unidades de análisis[5] y en cómo se concibe la dinámica entre dichas unidades. La escuela normativa pone el acento en las características comunes de la conducta humana, considerando que las variaciones en la vida y cultura humanas tienen una base ideológica y la función de los arqueólogos consiste en abstraer de los productos culturales las normas por las que se regían los humanos del periodo estudiado.

---

ejemplo sería B.G.Trigger (1970, 1978) que representa una postura de síntesis según la cual el enfoque histórico y el procesual son dos caras de una misma moneda. Queda fuera de los objetivos de este trabajo analizar la evolución de las ciencias históricas y ver cómo repercutieron en la arqueología y hasta qué punto la NA tuvo en cuenta la evolución de las corrientes historiográficas.

[5] La discusión sobre las unidades de análisis es una discusión sobre la ontología teórica de la ciencia, entendiendo por ontología las unidades mínimas de una disciplina sobre las que se construyen las teorías. Para un análisis de la ontología de la ciencia, ver Estany (1993), cap. 1.

La cultura es vista por los normativistas como un todo y cualquier intento de romper este todo se considera arbitrario. Las diferencias y similitudes culturales se expresan en términos de "relaciones culturales" que se "resuelven" en un modelo interpretativo general. Este enfoque deja al arqueólogo como un historiador cultural o un paleo-psicólogo y ésta no es la mejor situación para explicar el pasado. La NA propone un nuevo concepto de cultura para abordar la explicación de los procesos culturales. Así, la cultura sería "el medio extra somático de adaptación del hombre" (White, 1959). Por tanto, la cultura no puede medirse con una sola variable como puede ser la transmisión de ideas espacio-temporalmente, sino que en la cultura influyen muchas variables que actúan independientemente y la labor de los arqueólogos es aislar estos factores causales e investigar las relaciones entre dichos factores, su regularidad y su poder predictivo (Binford, 1965: 205). A través de esta búsqueda podremos establecer leyes de procesos culturales.

La NA insiste en la importancia de las técnicas de investigación que van desde las técnicas de datación hasta la construcción de modelos matemáticos y programas informáticos. Es decir, la NA apuesta por la utilización de las técnicas de investigación que las ciencias sociales ponen a su alcance. Además recurre a otras disciplinas como la química, la biología y la geología que le proporcionan medios para las técnicas de datación[6]. Por su parte, la AT utiliza mayormente los métodos propios de la investigación en historia.[7]

4. La "Nueva arqueología" como revolución Kuhniana

Uno de los temas más debatidos en la filosofía de la arqueología es la valoración de los cambios ocurridos en la década de los sesenta. Esta cuestión se concreta en si dichos cambios fueron o no una revolución y en si es factible aplicarles el modelo kuhniano. Análisis de este tipo los encontramos en Adams (1968), Martin (1971), Hill (1972), Zubrow (1972) y

[6] Ver Brothwell y Higgs (1963), editores, *Ciencia en arqueología* para las técnicas de datación, Orton (1988) *Matemáticas para arqueólogos* para la construcción de modelos matemáticos, Shennan (1992) *Arqueología cuantitativa*, para técnicas de investigación en general y J.A. Barceló (1996*) Arqueología automática. Inteligencia artificial*, para el papel de la inteligencia artificial en la arqueologia.

[7] Para un análisis de la NA desde la perspectiva de la tradición latinoamericana véase el trabajo de M. Gádara "La Vieja nueva Arqueología" (1980).

Fitting (1973). Hay prácticamente unanimidad en que hubo cambios significativos en la disciplina en la década de los sesenta. Varios de estos análisis ven estos cambios como revolucionarios pero hay desacuerdo sobre si se ajusta o no al modelo kuhniano. La mayoría de estos autores arguyen que los cambios ocurridos en la década de los sesenta en arqueología se refieren a cuestiones metodológicas. Esta última consideración es la que toma D. Meltzer (1979) para argumentar que si los cambios fueron, fundamentalmente, metodológicos entonces el aspecto revolucionario de la "Nueva Arqueología" queda seriamente debilitado.

Otro de los argumentos aducidos por Meltzer para no considerar la NA como una revolución es que no encaja con la concepción metacientífica de Kuhn. Uno de los motivos por los que no puede ser una revolución kuhniana es porque la concepción de la ciencia de la NA está basada en el empirismo lógico, concepción ampliamente criticada por Kuhn. Tenemos, pues, que el análisis del paso de la AT a la NA da lugar a dos posiciones: 1) este cambio fue una revolución kuhniana (Zubrow, 1972); 2) no hubo revolución kuhniana (Meltzer, 1979). La postura 2) alega dos cuestiones fundamentales: a) la incompatibilidad entre la concepción metateórica de Kuhn y la de Hempel, que fue quien inspiró la NA; y b) el hecho de que fuera, fundamentalmente, un cambio de metodología.

El argumento a) se refiere a la incoherencia atribuida a los autores de la NA por considerar que su trabajo, inspirado en el empirismo lógico, desencadenó una revolución kuhniana, esencialmente antipositivista. Pero esto es sólo una consecuencia de no haber considerado la obra de Kuhn en sus diversas facetas. Dos de estas facetas son fácilmente diferenciables: una es la crítica al empirismo lógico, faceta que discurre en el contexto de la justificación, y otra es una propuesta de análisis filosófico de la historia de la ciencia, faceta que discurre en el contexto del descubrimiento. No es habitual encontrar en la literatura filosófica la separación conceptual de estas dos facetas, con lo cual la adhesión o no al pensamiento de Kuhn se plantea siempre de forma global cuando, en realidad, aunque interrelacionadas, estas facetas discurren en planos distintos. Los arqueólogos de la NA toman la segunda faceta cuando califican la NA como una revolución kuhniana y, aunque de forma implícita, rechazan las críticas al empirismo lógico.

El argumento b) se refiere a que Kuhn no contempla revoluciones en que los cambios metodológicos sean centrales para el desarrollo de la disciplina. Por tanto, la objeción de Mertzer es pertinente, pero no las conclusiones que saca, diciendo que no hubo revolución. Mi propuesta es que hubo revolución

pero no precisamente kuhniana, sino una revolución metodológica. Kuhn introduce los compromisos metodológicos e instrumentales como parte del paradigma pero no contempla que el programa metodológico sea el motor del cambio.

La tesis que subyace a muchos de estos análisis de la historia de la arqueología es que las revoluciones científicas o son kuhnianas o no son. Es hora de revisar esta tesis y hoy con más razón que nunca después de más de cuatro décadas de la publicación *de La estructura de las revoluciones científicas* (1962), durante las cuales se han cuestionado y revisado algunas de las tesis defendidas en esta obra. Además, han surgido nuevos enfoques en el campo de estudio de la dinámica científica que suplen las carencias del modelo kuhniano. Sin embargo, en la década de los setenta, que es la época en que se realizaron muchos de los estudios del paso de la AT a la NA, el modelo kuhniano era predominante en filosofía de la ciencia, con lo cual se partía del supuesto de que las revoluciones científicas o eran kuhnianas o no podían considerarse revoluciones. Pero, ya en el siglo XXI este supuesto es insostenible ya que el modelo de Kuhn ha sido cuestionado en muchos puntos, uno de los cuales es precisamente que no es aplicable a todos los cambios históricos.

Entre los diversos modelos de cambio surgidos con posterioridad al de Kuhn (Lakatos, Hanson, Toulmin, Laudan etc.) el que más explícitamente introduce la metodología como un elemento de las Tradiciones de Investigación (TI)[8] es el de Laudan. Laudan (1977) distingue tres elementos en una TI: los problemas y su solución en el seno de las teorías, la ontología y la metodología; y prevé cambios parciales que afecten sólo a alguno de estos elementos, en el caso de la arqueología, a la metodología. No hay duda de que el modelo de Laudan supuso un avance en el análisis de los cambios científicos porque permite dar razón del desarrollo paso a paso de una disciplina. Sin embargo, lo que no parece considerar Laudan es que un cambio en el programa metodológico pueda desencadenar un giro de ciento ochenta grados en la investigación de la disciplina en cuestión. En Estany (1990) se propone la introducción de una tipología de cambios científicos, y en Estany (1996) se analizan las revoluciones metodológicas, siendo el paso de la AT a la NA un ejemplo claro de este tipo de revoluciones.

Otro punto a considerar es la valoración epistemológica de este cambio. Lo cual significa valorar la influencia del empirismo lógico como modelo

[8] Equivalentes a los paradigmas de Kuhn.

metodológico, teniendo en cuenta lo que son principios generales y lo que son las concreciones de dichos principios. En este punto es importante distinguir los valores epistemológicos generales de las formas concretas con las que se piensa hacer prevalecer dichos valores. Como valores epistémicos aceptados como guía de la ciencia podemos señalar la objetividad, la simplicidad, el poder explicativo etc. Aunque la idea de unos valores epistémicos haya sido motivo de debate en la filosofía de la ciencia, no cabe duda de que hay cierto consenso sobre los criterios epistémicos que guían la investigación científica.

Estos principios básicos se plasmaron en la arqueología de la década de los sesenta en una serie de patrones cuyo modelo fue la teoría de la ciencia procedente del empirismo lógico en su versión hempeliana. Así tomaron la explicación nomológico-deductiva como modelo de explicación y, consecuente con ello, se plantearon la formulación de leyes generales sobre las relaciones interculturales, tarea clave dado el papel que las leyes generales juegan en el modelo de explicación de Hempel-Oppenheim.

Poner en práctica el programa metodológico de Hempel requería la utilización de los métodos cuantitativos para los cuales pusieron especial énfasis en las técnicas de datación y en la utilización de modelos matemáticos. Es decir, la NA supuso un cambio en todos los niveles metodológicos[9]: en los principios generales, en el sentido de valores epistemológicos, en la concepción de lo que debe ser una ley, una teoría, o una explicación científica y en las técnicas de investigación con la utilización de análisis químicos, programas informáticos, etc.

5. La arqueología postprocesual (APP): la arqueología contextual

Si bien siempre subsistió una parte de arqueólogos que no se sumó al nuevo paradigma, podemos decir que la NA predominó durante las décadas de los sesenta y los setenta. Las críticas a la NA comenzaron a finales de los setenta pero se desarrollaron, sobre todo, en los ochenta. Vamos a examinar esta crítica a través del pensamiento de uno de sus máximos exponentes: I. Hodder (1994)[10]. Hay que señalar que una de las características de la APP es que la arqueología deja de tener un modelo unificado de investigación y se

[9] Ver Estany, 1993, cap.1 para un análisis de los diversos sentidos de metodología.
[10] En Interpretación en arqueología Hodder expone su pensamiento sobre el modelo metodológico en arqueología.

presenta como pluralista en cuanto al enfoque. Sin embargo, podemos señalar algunas características comunes a los distintos enfoques pos-procesuales, cuyo denominador común es su oposición a la NA:

a) Uno de los puntos de divergencia reside en la importancia de la generalización, primordial para la NA e irrelevante para la APP. La APP pone el énfasis en el individuo, en lo idiosincrático en contraposición a la generalización: "¿Hasta qué punto podemos generalizar a partir de contextos culturales únicos, y por qué esforzarnos en generalizar, en cualquier caso?"[11]. Esto le lleva a una crítica del enfoque marxista, estructuralista y sistémico, y a todos los que intentan establecer relaciones interculturales.[12]

b) Crítica a la determinación de la cultura a partir de los resultados materiales. Según la APP hay que tener en cuenta los elementos subjetivos, las ideas, es decir, la mente del individuo: "la cultura no es reducible a los resultados materiales"[13].

c) Imposibilidad de contrastación y de utilización de medios objetivos de medición: "es imposible la contrastación de la teoría con los datos, un mecanismo independiente de medición y un conocimiento cierto del pasado"[14]

d) Crítica del supuesto positivista de que midiendo la covariancia entre variables observables en el mundo real, el sistema puede ser identificado y verificado. Esta confianza en los datos es lo que Hodder considera ilusorio.

A partir de estas críticas Hodder propone la "Arqueología Contextual"(AC), señalando que contextualismo no significa particularismo y que el análisis contextual no es incompatible con la teoría y la generalización: "'contextualismo' no significa 'particularismo', un término que, en arqueología, ha venido a asociarse al rechazo o a la falta de interés por la teoría general"[15]. Sin embargo, estas afirmaciones encajan mal con lo dicho anteriormente criticando a las generalizaciones y a las relaciones interculturales.

El concepto de "contexto" es fundamental para la propuesta de Hodder ya que todo se refiere a este concepto:

---

[11] Ibid., 20.
[12] Como hemos señalado anteriormente, han habido críticas a la arqueología procesual que no podemos integrarlas totalmente en los postulados de la APP en la línea de Hodder, quien también cuestiona el enfoque maxista en el que se enmarca la Arqueología Social Latinoamericana.
[13] Ibid., 25.
[14] Ibid., 32.
[15] Ibid., 165.

> La arqueología contextual implica el estudio de los datos contextuales, utilizando métodos contextuales de análisis, para llegar a dos tipos de significado contextual, analizados en función de una teoría general"[16].

> El contexto relevante de un objeto x al que queremos dar un significado (de cualquier tipo) son todos aquellos aspectos de los datos que tienen relación con x, y que obedecen a una pauta significativa según la descripción anterior. (...) el contexto de una característica arqueológica es la totalidad del medio relevante, donde "relevante" se refiere a una relación significativa con el objeto, esto es, una relación necesaria para discernir el significado del objeto"[17].

Una consecuencia inmediata es que no es posible analizar una característica de un objeto aislada de todas las demás, es decir, la concepción de Hodder es holista en el sentido de que, según él mismo reconoce, todo depende de todo y cualquier característica que queramos definir depende de las características de todas las demás.

Para valorar en su justa medida la propuesta de Hodder hay que tener en cuenta que el significado del objeto de estudio depende no sólo del contexto de dicho objeto sino también del contexto del arqueólogo como persona. Así, según el contexto del investigador tendríamos dos perspectivas arqueológicas: "establecidas" y "alternativas". Por arqueología establecida entiende Hodder la arqueología escrita por el sexo masculino, de clase media alta y, en su mayor parte, anglosajona. Los enfoques alternativos son los que corresponden a las arqueologías indígenas, la arqueología feminista y la arqueología de la clase obrera entre otras. Dice Hodder: “En todas ellas cabe destacar dos cosas: primero, el pasado se construye subjetivamente en el presente y, segundo, el pasado subjetivo está implicado en las actuales estrategias de poder"[18]. A pesar del rechazo de la perspectiva arqueológica establecida, Hodder es un perfecto representante de ella ya que tiene todas las características: sexo masculino, clase media alta y anglosajón.

Hodder (1994) resume así los rasgos fundamentales de la APP:

> La arqueología postprocesual, al revés de la procesual, no defiende un solo enfoque, ni afirma que la arqueología debe desarrollar una metodología aceptada. Por ello la arqueología postprocesual es sencillamente "post". Parte de una crítica de lo anterior construyendo sobre esta vía, pero al mismo tiempo divergiendo de ella. Supone diversidad y falta de consenso. Se caracteriza por el debate y la incertidumbre acerca de los problemas fundamentales poco

[16] Ibid., 165.
[17] Ibid., 154.
[18] Ibid., 176.

> discutidos anteriormente en arqueología. Es más un planteamiento de preguntas que una provisión de respuestas.[19]

Hemos expuesto las principales características de la APP a través del trabajo de Hodder, uno de los más fieles representantes. Tenemos, pues, una arqueología que rechaza los ideales epistémicos de la objetividad, contrastación de hipótesis, observación y medición de los datos y que aboga por la subjetividad, lo idiosincrático, lo individual y lo contextual.

De la crítica de Hodder a los postulados de la NA no puede inferirse que todos los enfoques que marcan distancias o difieren más o menos radicalmente de la NA pueda atribuírseles falta de los más elementales valores epistémicos que requieren cualquier disciplina científica. En este sentido, la crítica a la NA desde la Arqueología Social Latinoamericana no puede incluirse en el enfoque propuesto por Hodder, a pesar de que sean cuestionables algunos de sus postulados. Sólo hay que tener en cuenta algunas de las afirmaciones de L. F. Bate quien, a pesar de sus críticas al positivismo, señala que la fase de obtención de información sobre el registro arqueológico permite la "obtención, procesamiento analítico, ordenación, descripción y comunicación de la información generada a partir de los datos arqueológicos empíricamente observables" (Bate 1989: 12), lo cual implica formular "protocolos de registro y procedimientos técnicos y analíticos que sistematicen los trabajos de campo y laboratorio, así como la creación de acervos y de procedimientos de comunicación de la información producida" (Bate 1989: 12). Estas afirmaciones están más cerca del enfoque empirista, aunque Bates no lo muestre de forma explícita, que de posicionamientos relativistas.

6. ¿Constituye la arqueología Post-procesual una nueva revolucíon en arqueología?

¿Es la APP un paradigma en competencia con la AP? El programa de Hodder puede producir distintos productos culturales pero no el producto cultural de lo que entendemos por ciencia. Dichos productos culturales pueden proporcionar conocimiento sobre nuestros ancestros, pero no el tipo de conocimiento sistemático y que tiene como objetivo ser lo más fiel posible al mundo real. La ciencia no agota la forma de acercarse al mundo, pero es la mejor forma de hacerlo cuando el objetivo primordial es el conocimiento del

[19] Ibid., 190.

mundo. La APP bien podría llamarse "arte arqueológico" o "novela arqueológica" pero no ciencia arqueológica ya que ni comparte los fines generales de la ciencia ni acepta sus reglas de juego.

¿Ha fracasado la NA? El fracaso significaría que el trabajo arqueológico ha renunciado al conocimiento lo más objetivo posible del pasado y que sigue estrictamente los dictados de Hodder. ¿Es esto lo que realmente ocurre en la investigación arqueológica? ¿Hasta qué punto la investigación empírica en arqueología se ciñe a los principios contextuales?

La arqueología no ha dejado de ser considerada una ciencia social, cuyo objetivo es el explicar las sociedades prehistóricas. Coexisten comunidades de arqueólogos con un interés histórico pero esto no invalida lo anterior, también la historia económica, la historia natural tienen su nicho disciplinario sin que nadie cuestione la economía como una ciencia social y la biología como una ciencia natural.

En cuanto a los principios metodológicos fundamentales siguen tan vigentes hoy día como cuando Binford los formuló en la década de los sesenta. Los únicos que cuestionan estos principios metodológicos son los arqueólogos que se sitúan en corrientes sociologistas como el "programa radical en sociología del conocimiento". Pero entonces el problema no es que cuestionen los principios metodológicos que subyacen a la investigación arqueológica sino que se cuestiona cualquier principio metodológico y cualquier conocimiento científico. Según esta corriente el mundo no proporciona ningún límite a nuestras creencias. La crítica del programa radical no supone ningún reto a la NA.

Una de las críticas a la NA es que se acogió al empirismo lógico y, en concreto, a la versión de Hempel. La argumentación discurre en los términos siguientes: la NA se fundamenta en una concepción de la filosofía de la ciencia que ha sido abandonada por la propia comunidad de filósofos, por tanto, no tiene sentido continuar defendiendo la NA cuando sus cimientos se han desmoronado. En este punto es dónde adquiere especial importancia la distinción entre principios metodológicos fundamentales y sus concreciones. Es cierto que la filosofía de la ciencia actual es crítica con muchos de los presupuestos del empirismo lógico, por ejemplo, la distinción teórico-observacional, la concepción sintáctica de las teorías, el modelo de explicación nomológico-deductivo, el énfasis en las reconstrucciones formales de las teorías etc. Sin embargo, ¿podemos afirmar que los filósofos de la ciencia han renunciado a los valores epistémicos como guía de la

investigación científica? La respuesta es no, al menos para una buena parte de la comunidad de filósofos de la ciencia.

En la propuesta original de Binford, éste proponía tomar la concepción del empirismo lógico en la versión de Hempel como guía para la investigación arqueológica. En este punto sí podemos decir que la NA se equivocó o, al menos, este presupuesto ya no es válido actualmente porque ha habido críticas bastante definitivas y, lo más importante, existen alternativas que se adecuan mucho mejor a la práctica científica. Aquí deberíamos incluir desde la concepción semántica[20] hasta filósofos como P. Kitcher (1993) N. Nersessian (1992) y P. Thagard (1992) que proporcionan esquemas que, aun manteniendo los valores epistémicos, son capaces de dar cuenta de realidades complejas.

En cuanto a la utilización de técnicas de datación gracias al desarrollo de otras ciencias no sólo no ha sido abandonado sino que se ha incrementado. Por ejemplo, el descubrimiento del tesoro de Troya en Moscú puede aportar datos muy importantes gracias a técnicas de datación muy sofisticadas y muy fiables[21]. Ningún arqueólogo hace ascuas a la utilización de dichas técnicas, antes al contrario se considera una oportunidad única para poder desvelar información inalcanzable hasta el momento. Como dice Binford, los principios metodológicos y las técnicas de investigación introducidas por la NA han contribuido a resolver problemas planteados por la arqueología tradicional. Valorar estas técnicas para la arqueología significa valorar positivamente los ideales epistémicos que subyacen en la investigación científica y que los arqueólogos de la NA hicieron suyos.

7. Conclusiones

1) La historia cultural y la ciencia de la cultura no son dos disciplinas en competencia de las que hay que tomar partido por una de ellas en detrimento de la otra. Binford se equivocó en contraponerlas sin tener en cuenta los cambios que las disciplinas históricas habían sufrido durante el siglo XX.

2) Los cambios ocurridos en la arqueología en la década de los sesenta fueron suficientemente importantes como para hablar de una revolución,

[20] Me refiero a la corriente propugnada por R. Giere, B. van Frassen, P. Kitcher, entre otros.

[21] Otra muestra de la utilización de técnicas rigurosas en le resolución de problemas hasta ahora no desentrañados se encuentra en la obra de B. Fagan (1995) *Time detectives*.

fundamentalmente de una revolución metodológica. Si no encaja con el modelo de revolución científica de Kuhn, lo que han que replantear no es si Binford llevó a cabo una revolución sino si Kuhn fue demasiado simplista en sus planteamientos y si su modelo es capaz de abordar determinados cambios significativos de la historia de la ciencia.

3) Si calificamos la NA como una revolución metodológica, la NA no ha fracasado, al menos en sus principios fundamentales.

4) La llamada "arqueología post-procesual" que, en realidad, como dice Renfrew, habría que llamarla "anti-procesual" no hace ninguna aportación interesante a la arqueología como estudio sistemático y científico del pasado de nuestra especie. Ningún arqueólogo, cuyos principios metodológicos sean los de Hodder, puede sentirse muy motivado en su trabajo. Mi impresión es que cuando los arqueólogos post-procesualistas salen a hacer trabajo de campo, consciente o inconscientemente siguen los principios metodológicos procesualistas, utilizando todas las técnicas de datación disponibles, lo cual entra en contradicción con sus principios teóricos, ya que, ¿para qué utilizar técnicas de datación si no es posible la objetividad?

5) Últimamente la meta-arqueología se ha centrado demasiado en las corrientes más sociologistas que no son ni mucho menos predominantes en filosofía de la ciencia, olvidando otros enfoques que, aún enlazándose con la tradición positivista, son capaces de abordar campos mucho más complejos. Por ejemplo, la concepción semántica de las teorías de R. Giere, o los modelos de explicación científica de autores como P. Kitcher y W. Salmon.

6) De la antigua URSS se decía que era un gigante con los pies de barro. Utilizando esta metáfora podríamos decir que la arqueología es un enano con los pies de acero. Es una ciencia social joven pero anclada en las ciencias naturales más asentadas.

7) El pasado puede ser abordado desde perspectivas distintas: como objetos de arte, como historia cultural y como ciencia de la cultura, no son incompatibles pero tampoco pueden tomarse como paradigmas distintos y en competencia en arqueología. Tenemos otros ejemplos donde un objeto puede ser abordado desde diversas perspectivas. Tal es el caso de los minerales que pueden abordarse desde el arte: piedras preciosas utilizadas en joyería; como historia natural y como ciencia que es la geología.

Esto significa que muchas de las corrientes post-procesuales no constituyen paradigmas en competencia con la NA. Creo que es un error por parte de procesualistas considerarlas como tales. Tiene una explicación porque en disciplinas jóvenes (inmaduras o preparadigmáticas, diría Kuhn) el

debate filosófico-metodológico juega un papel muy importante e influye directamente en su evolución. Algunos post-procesualistas se lamentan de la influencia de la filosofía de la ciencia, en concreto del positivismo lógico, en la comunidad de arqueólogos porque consideran que fue perjudicial para la arqueología y señalan que por fin la arqueología se ve libre de la influencia de los filósofos de la ciencia. Nada más lejos de la realidad, la arqueología post-procesual, en su mayor parte, está impregnada de relativismo y de sociologismo, versiones actuales del escepticismo filosófico que empezó en Grecia con los pirrónicos. ¿Es necesario elegir entre el positivismo lógico y el relativismo? Creo que existen alternativas de equilibrio. Dejo al lector la elección entre positivismo y relativismo como mejor aliado en la práctica científica con el objetivo de satisfacer uno de los anhelos de nuestra especie, a saber: el conocimiento del mundo que nos rodea.

Bibliografia

Adams, R. McC., "Archaeological research strategies: past and present". *Science* 160: 1187-1192, 1968.

Adams, W. Y. y E. W. Adams, E. W., *Archaeological typology and practical reality. A dialectical approach to artifact classification and sorting*. Cambridge University Press, Cambridge, 1991.

Barceló, J. A., *Arqueología automática. El uso de la inteligencia artificial en arqueología*. Editorial Ausa, Barcelona, 1996.

Bate, L. F., "Notas sobre el materialismo histórico en el proceso de investigación arqueológica". *Boletín de antropología americana* 19: 5-27, 1989a.
---- Bate, L. F. 2007b. "Teorías y métodos en Arqueología ¿Criticar o proponer?". En: *Boletín Electrónico Arqueología y Marxismo. Ediciones Las Armas de la Crítica, pp: 105-115.*

Binford, L. R., "Archaeology as anthropology". *American Antiquity* 28(2): 217-225 1962.
--- "Archaeological systematics and the study of cultural process". *American Antiquity* 31(2): 203-210, 1965.
--- "Archaeological perspectives. En *New perspectives in Archaeology,* editado por S. R. Binford y L. R. Binford, pp. 5-23. Aldine Publishing House, Chicago, 1968a.
--- "Some comments on historical versus processual Archaeology". *Southwestern Journal of Anthropology* 24 (3): 267-275, 1968b.
--- "General introduction. En *For theory building in Archaeology*, editado por L. R. Binford. Academic Press, New York, 1977.
--- *Working in archaeology*. Academic Press, New York, 1983.
--- *Debating Archaeology*. Academic Press, New York, 1989.

Boas, F. *Race, language and culture*. The McMillan Company, New York, 1966 (primera edición en 1940).

Brothwell, D. R. y E. Higgs (editores), *Science in archaeology: a comprehensive survey of progress and research with a foreword by Grahame Clark*. Thames & Hudson, London,1963.

Embree, L. (editor), *Metaarchaeology. Reflections by archaeologists and philosophers*. Kluwer Academic Publishers, Boston, 1992.

Estany, A., *Modelos de cambio científico*. Crítica, Barcelona, 1990.
--- *Introducción a la filosofía de la ciencia*. Crítica, Barcelona, 1993.
--- The role of methodology in models of scientific change. En *Spanish Studies in the Philosophy of Science*, volumen 186, editado por G. Munevar pp. 275-288, Kluver Academic Publishers, 1996.

Finley, M. I., "Archaeology and History". *Proceedings of the American Academy of Arts and Sciences*, volumen 100, pp. 168-186, 1971.

Fitting, J. E., "History and crisis in Archaeology". En *The development of north-american Archaeology,* editado por J.E. Fitting, pp. 1-13. Doubleday, New York, 1973.

Flagan, B., *Time detectives. How archaeology use technology to recapture the past*. Simon & Schuster, Ney York, 1995.

Fritz, J. M. y F. T. Plog, "The nature of archaeological explanation". *American Antiquity* 35(4): 405-412, 1970.

Gandara, M., "La Vieja nueva Arqueología". *Boletín de Antropología Americana* 2: 7-45,1981.
--- "El análisis de posiciones teóricas: aplicaciones a la arqueología social". *Boletín de antropología americana* 27: 5-20, 1993.

Gibson, G., *Explanation in Archaeology*. Blackwell, Oxford, 1989.

Giedymin, J., "Antipositivism in contemporary philosophy of social sciences and humanities". *British Journal for the Philosophy of Science*, 26: 275-301, 1975.

Heizer, R. F. y S. F. Cook (editores), *The application of quantitative methods in Archaeology*. Quadrangle Books. New York, 1960.

Hempel, C. G., *La explicación científica. Estudios sobre la filosofía de la ciencia. Paidós*, Buenos Aires, 1979.

Hill, J. N., "The methodological debate in contemporary Archaeology: a model". En *Models in Archaeology*, editado por D. L. Clarke, pp. 61-108. Methnen, London, 1972.

Hodder, I., *Interpretación en arqueología. Corrientes actuales*. Crítica, Barcelona, 1974.

Hole, F. y R. F. Heizer, *An introduction to prehistoric Archaeology*. Rinehart and Winston, New York: Holt, 1973.

Kelley, J. H. y M. P .Hanen (editores), *Archaeology and the methodology of science*. University of New Mexico Press, Alburquerque, 1988.

Kitcher, P., *The advancement of science*. Oxford University Press, Oxford, 1993.

Krober, A. L., "History and Science in Anthropology". *American Anthropologist*, 37: 539-569, 1935.

Kuhn, T., *La estructura de las revoluciones científicas*. Fondo de Cultura Económica, México, 1971 (primera edición 1962).

Laudan, L., *El progreso y sus problemas*. Encuentro, Madrid, 1986.

Leone, M. P., "Issues in Anthropological Archaeology". En *Contemporary Archaeology: a guide to theory and contributions*, editado por M. P. Leone, pp. 14-27. Southern Illinois University Press, Carbondale, 1972.

Levin, M. E., "On explanation in Archaeology: a rebuttal to Fritz and Plog". *American Antiquity*, 37(4): 387-395, 1973.

Martin, P.S., "The revolution in Archaeology". *American Antiquity*, 36 (1): 1-8, 1971.

Meltzer, D. J., "Paradigms and the nature of change in American Archaeology". *American Antiquity*, 44 (4): 644-657, 1979.

Nersessian, N., "How do scientists think? Capturing the dynamics of conceptual change in science". En *Cognitive models of science*, editado por R. Giere, pp. 3-44. University of Minnesota Press, Minneapolis, 1992.

Orton, C., *Matemáticas para arqueólogos*. Alianza Universidad, Madrid, 1988.

Pandit, G. L. (editor), *Methodological variance. Essays in Epistemological Ontology and the Methodology of Science*. Kluwer, Boston, 1991.

Reid, J. J., M. Chiffer y W. Rathje, "Behavioral Archaeology: four strategies". *American Anthropology*, 77: 864-869, 1975.

Renfrew, C., "Coments on Archaeology into the 1990's". *Norwegian Archaeological Review*, 22 (1): 33-41, 1989.

Sabloff, J. A. y G. R. Willey, "The collapse of Maya civilization in the southern lowlands, a consideration of history and process". *Southwestern Journal of Anthropology,* 23: 311-336, 1967.

Salmon, M. H., "Deductive" versus "inductive" Archaeology. *American Antiquity*, 41 (3): 376-381, 1976.

Shennan, S., *Arqueología cuantitativa*. Crítica, Barcelona, 1992.

Taylor, W. W., *A study of Archaeology*. Memoir 69 of the American Anthropological Association, 1948.

Thagard, P., *Conceptual revolutions*. Princeton University Press, Princeton (N. J.), 1992.

Thompson, R. H., Interpretative trends and linear models in American Archaeology. En *Contemporary Archaeology: a guide to theory and contributions*, editado por M.P. Leone. Southern Illinois University Press, Carbonale, 1972.

Trigger, B. G., "Aims in prehistoric Archaeology". *Antiquity*, 44 (173): 26-37, 1970.
--- *Time and tradition: essays in Archaeological interpretation*. Columbia University Press, New York, 1978.

Tuggle,H. D., N. P., S. Price y B. G. Trigger, “Trigger and prehistoric Archaeology”. *Antiquity*, 45 (178): 130-145, 1971.

Tuggle,H. D., D., A. H. Townsend y R. J. Riley, “Laws, systems and research design: a discussion of explanation in Archaeology”. *American Antiquity*, 37 (1): 3-12, 1972.

Watson, R. A., “The 'New Archaeology' of the 1960's”. *Antiquity*, 46 (183): 210-215, 1972.

White, L., “History, evolutionism and funtionalism”. *Southwestern Journal of Anthropology*, I: 221-248, 1945.

White, L., “The concept of culture”. *American Anthropology*, 61(2): 227-251, 1959.

Willey, G. R. y P. Phillips, *Method and theory in American archaeology*. University of Chicago Press, Chicago, 1958.

Wylie, A., “The interplay of evidential constraints and political interests: recent archaeological research on gender”. *American Antiquity*, 57 (1): 15-35, 1992.

Zubrow, E. B. W., “Environment, subsistence and society: the changing Archaeological perspective”. *Annual Review of Anthropology*, 1:179-206, 1972.

Freud, a concepção do Descentramento e a Física Moderna

Lino Machado
(Universidade Federal do Espírito Santo)
lino@npd.ufes.br

1. A dívida parcial de certa reflexão freudiana para com a física clássica pré-einsteiniana

Em 1916 e 1917, Sigmund Freud proferiu as *Conferências introdutórias sobre psicanálise*, as quais estão entre os seus textos mais lidos. No final da XVIII Conferência ("Fixação em traumas – O inconsciente"), há uma página cheia de implicações filosóficas, que diz respeito à posição dos homens no interior do cosmo, do reino animal e, para nada ficar em falta, de si mesmos. Esta página seria expandida em cerca de sete outras, publicadas ainda em 1917, no periódico húngaro *Nuygat*, sob o título "Uma dificuldade no caminho da psicanálise".

Abordando a problemática dos traumas no contexto da Primeira Guerra Mundial, Freud encerra a sua mencionada XVIII Conferência com a lembrança de dois grandes golpes no *homo sapiens*, aos quais ele acrescenta uma nova pancada, desferida agora pela psicanálise, ou seja, pela sua pessoa autoral (reforçada com o prestígio adquirido ao liderar um movimento ainda na vanguarda das ideias), que igualmente vem afligir o amor próprio, o narcisismo do gênero humano.

Golpe inicial, na visão de Freud:

> O primeiro foi quando souberam que a nossa Terra não era o *centro do universo*, mas o diminuto fragmento de um sistema cósmico de uma vastidão que mal se pode imaginar. Isto estabelece *conexão*, em nossas *mentes***,** com *o*

> *nome de Copérnico*, embora algo de semelhante já tivesse sido afirmado pela *ciência* de Alexandria[1].

Baque seguinte, na versão de Freud:

> O segundo golpe foi dado quando a *investigação biológica* destruiu o lugar supostamente privilegiado do homem na criação, e provou sua descendência do reino animal e sua inextirpável natureza animal. Esta nova avaliação foi realizada [...] por *Darwin*, *Wallace* e seus predecessores, embora não sem a mais *violenta oposição* [...][2].

Terceiro choque, oriundo agora do tinteiro afiado do próprio Freud:

> Mas a megalomania humana terá sofrido o seu terceiro golpe, o *mais violento*, a partir da pesquisa psicológica [...] que procura provar [a]o ego que ele não é senhor nem mesmo *em sua própria casa*, devendo, porém, contentar-se com *escassas informações* acerca do que acontece *inconscientemente* em sua mente[3].

À última pancada, o autor ajunta:

> Os *psicanalistas* não foram os primeiros e nem os únicos que fizeram essa invocação à introspecção; todavia, parece ser *nosso destino* [...] apoiá-la com material empírico que é encontrado em *todas as pessoas*. Em consequência, surge a *revolta geral contra nossa ciência*, o desrespeito contra todas as noções de civilidade acadêmica[4].

Com elegância, Freud não cita o seu próprio nome neste contexto da XVIII Conferência, como citara os de Copérnico, Darwin e Wallace. O decoro do pensador austríaco, todavia, não deve impedir-nos de supor um duplo desejo organizando as palavras da página final (lugar privilegiado, "coda") da Conferência em exame: o de associar, de um lado, a psicanálise a ciências ("não humanas") como a física e a biologia e, de outro, o de conectar o (implícito) nome próprio de Freud aos de Copérnico, Darwin e Wallace, como desbravadores de temas que os homens em geral prefeririam manter em silêncio. Desejo duplo: institucional e pessoal ao mesmo tempo, legitimamente narcísico, aliás.

Na modernidade e na pós-modernidade do século XX (a partir do estruturalismo e do pós-estruturalismo), o que viria a ser denominado

[1] Sigmund Freud, *Conferências introdutórias sobre psicanálise*. Edição standard brasileira das obras psicológicas completas. Vol. XVI. Trad. José Luiz Meurer. Rio de Janeiro: Imago, 2006a, 292, destaques nossos.
[2] Sigmund Freud, op. cit., 2006a, 292, destaques nossos.
[3] Sigmund Freud, op. cit., 2006a, 292, destaques nossos.
[4] Sigmund Freud, op. cit., 2006a, 292, destaques nossos.

descentramento não deixou de prender-se ao entrelaçar de signos onomásticos efetuado antes, com mão de mestre, por Freud: o emaranhamento do seu patronímico, da obra cuja paternidade era sua, aos dos pais de famosas teorias do passado (e mesmo do futuro, em diálogos póstumos com os eventuais "Freuds" do devir e os seus discípulos, numa "conexão, em nossas mentes, com o nome de" quem quer que consiga fazer-se associado a uma façanha mental da espécie). O que, afinal das contas, é o que se entende como autoria, no sentido forte da palavra – ou era, posta que foi sob suspeita por (anti)autores como Roland Barthes e Michel Foucault[5] (paradoxalmente hiper-autorais em alguns dos textos que assinaram) e os seus herdeiros, que tentaram limar os pedestais nos quais, em geral, os criadores são colocados, sejam exploradores das psiques, dos átomos ou de quaisquer outros domínios. Como veremos, Foucault não deixará de ser fiel, em parte, ao criador da psicanálise, ao intitular "Nietzsche, Freud, Marx" uma conferência da sua lavra, no Colóquio de Royaumont (julho de 1964). Com certeza, Sigmund apreciaria ler o seu patronímico ladeando o de Friedrich, mas é duvidoso que se entusiasmasse vendo-o em companhia do de Karl[6].

Abordemos agora o artigo publicado no periódico *Nuygat*.

Em "Uma dificuldade no caminho da psicanálise", Freud trata do embaraço do ego em lidar com a "vida instintual da mente", de acordo com tudo o que a sua "teoria da libido" descobrira. Em muitas ocasiões, o mesmo ego, sentindo-se ameaçado pelas pulsões eróticas, coloca-se "na defensiva" e "nega aos instintos sexuais a satisfação que almejam"[7]. Tentando libertar as pessoas desses distúrbios, Freud percebeu um fato com valor geral: uma "distribuição primeva da libido dos seres humanos".

---

[5] Referimo-nos, sobretudo, a Roland Barthes, "A morte do autor". In: *Rumor da língua*. Trad. António Gonçalves. Lisboa: Edições 70, 1987, 49-53 e Michel Foucault, *O que é um autor?* Trad. António Fernando Cascais e Edmundo Cordeiro. [S.l.]: Passagens, 1992, passim.

[6] Não por acaso, no belo prefácio que escreveu para a sua antologia de textos ligados ao estruturalismo, Eduardo Prado Coelho acrescentou Marx à problemática do descentramento. Influenciado por Foucault, o importante teórico e crítico português associou o nome do autor de *O capital* aos do pai da psicanálise e de Nietzsche. Cf. Eduardo Prado Coelho, "Introdução a um pensamento cruel: estruturas, estruturalidade e estruturalismos". In: *Estruturalismo: antologia de textos teóricos*. Trad. Maria Eduarda Reis Colares et al. Lisboa: Portugália, 1968, I-LXXV, esp. XXXIX.

[7] Sigmund Freud, *Uma neurose infantil e outros trabalhos*. Edição standard brasileira das obras psicológicas completas. Vol. XVII. Trad. José Luiz Meurer. Rio de Janeiro: Imago, 2006b, 148.

> Fomos levados a presumir que, no início do desenvolvimento do indivíduo, toda a sua libido (todas as tendências eróticas, toda a sua capacidade de amar) está vinculada a si mesma – ou, como dizemos, catexiza o seu próprio ego. É somente mais tarde que, ligando-se à satisfação das principais necessidades vitais, a libido flui do ego para os objetos externos. [...] Para a libido, é possível desvincular-se desses objetos e regressar [...] ao ego[8].

Dessa percepção do ir e vir da libido, Freud dá um salto conceitual para o "narcisismo *universal* dos homens", o seu "amor-próprio", que "sofreu até o presente três severos golpes por parte das *pesquisas científicas*"[9], entre as quais ele inscreveu as da psicanálise, desde a redação da XVIII Conferência. E sabemos que na sua derradeira página foram recordados os onomásticos Copérnico, Darwin e Wallace, sendo que os dois primeiros retornarão ao artigo presente (com o acréscimo do nome do alexandrino Aristarco de Samos ao de Copérnico).

Em alíneas (a), (b) e (c), Freud retomará a trinca de golpes no que agora classifica de "ilusão narcísica" dos homens como um todo: um baque "associa-se, em nossas mentes, com o nome e a obra de Copérnico" (redação muito semelhante à já fixada na XVIII Conferência), cujo heliocentrismo fora antecedido pelo de Aristarco, que "havia declarado que a Terra era muito menor que o sol e movia-se ao redor deste corpo celeste"[10]; outro baque veio com "as pesquisas de Charles Darwin", que puseram fim à presunção do homem de "colocar um abismo entre a sua natureza e a dos animais"[11]; por fim, o que "talvez seja o que por natureza *mais fere*"[12], o da psicanálise, naturalmente. Por ordem retrospectiva, aguentemos um "golpe cosmológico", um "golpe biológico" e um "de natureza psicológica": somos retirados do centro do universo, expulsos do centro da natureza e confrontados com o "labirinto de impulsos" das nossas mentes. Tríade de desgraças, que, caindo dos céus mais elevados, chega ao interior das cabeças que os observam.

Nos anos 1960, em "Nietzsche, Freud, Marx", como dissemos, Michel Foucault retomará esses textos freudianos, sobretudo o de 1917, sem nomeá-los: "Freud fala, em algum lugar, que há *três grandes feridas narcísicas* na cultura ocidental", afirmativa a que Foucault acrescenta um interessante comentário em forma de indagação retórica (erótema):

---

[8] Sigmund Freud, op. cit., 2006b, 148.
[9] Sigmund Freud, op. cit., 2006b, 149, destaques nossos.
[10] Sigmund Freud, op. cit., 2006b, 149.
[11] Sigmund Freud, op. cit., 2006b, 149.
[12] Sigmund Freud, op. cit., 2006b, 149, destaques nossos.

> Eu *me pergunto* se não seria possível dizer que Freud, Nietzsche e Marx, nos envolvendo em uma tarefa de *interpretação que sempre se reflete sobre si mesma*, constituíram à nossa volta, e para nós, esses *espelhos*, de onde nos são enviadas as imagens, cujas figuras inesgotáveis formam o *nosso narcisismo atual*[13].

Foucault, entretanto, não se responde (de modo *direto*, ao menos), seguindo por outra via, em seu texto ("Em todo o caso..."). No caso presente, somos tentados a responder por ele, aproveitando a sua questão, mas de um modo que o filósofo francês talvez desautorizasse.

Sim, parece-nos aqui haver uma "interpretação que sempre se reflete sobre si mesma", em "espelhos" nos quais o "nosso narcisismo atual" (já velho, num pós-1960 de terceiro milênio), mais do que "formar-se", repete-se, e tudo isto – ferimentos narcisistas requentados, tornados clichês intelectuais sem qualquer poder de subversão, vulgata fácil de referir ou repetir, trabalho de exegese que segue auto-espelhando-se – está enredado com uma visão mais antiga, mas renitente, da realidade: o seu nome é física clássica newtoniana, ou melhor, *certa cosmovisão que passou a acompanhá-la*. O livro *O campo*, de Lynne McTaggart, contém uma boa descrição do que ela seja:

> [...] Tudo que acreditamos a respeito do nosso mundo e do lugar que ocupamos nele deriva de ideias formuladas do [sic] século XVII [por Isaac Newton, sobretudo], mas que ainda compõem a espinha dorsal da ciência moderna – teorias que apresentam todos os elementos do Universo como sendo *isolados uns dos outros, divisíveis e de todo independentes*.
>
> Essas concepções, em sua essência, criaram uma visão de mundo de *separação*. Newton descreveu um mundo material em que as partículas individuais da matéria seguem certas leis de movimento através do espaço e do tempo, ou seja, o Universo como uma máquina. [...]
>
> Esse mundo de *separações* deveria ter sido destruído[14] de uma vez por todas pela descoberta da física quântica [e da teoria da relatividade] na primeira parte do século XX[15].

---

[13] Michel Foucault, "Nietzsche, Freud, Marx". In: *Ditos e escritos II*. 2ª ed. Trad. Elisa Monteiro. Rio de Janeiro: Forense Universitária, 2008, 43, destaques nossos.

[14] Há exagero aqui, se considerarmos a dimensão do Universo que habitamos, na nossa escala humana, na qual a Física de Newton atua muito bem, embora não em escalas maiores, envolvendo objetos muito grandes e massivos e velocidades mais altas, nas quais os conceitos de Einstein (teoria da relatividade geral) são necessários. A cosmovisão de tal física é que precisa ser superada de uma vez por todas, nos domínios do cotidiano, não apenas nos da ciência – ou nos do chamado "misticismo quântico" ou nos dos seguidores da Nova Era (New Age)...

[15] Lynne McTaggart, *O campo*. Trad. Claudia Gerpe Duarte. Rio de Janeiro: Rocco, 2008, 16 e 18, destaques nossos.

Esta concepção, inclusive em parcela dos meios intelectualizados, prossegue impondo, quando não uma visão apenas determinista dos fatos, um modelo de "separabilidade" das coisas, uma noção de tempo dissociada fisicamente da de espaço (este é apenas o palco onde aquele transcorre), etc. Exagero? A quem Freud recorreu para expor o nosso hematoma cosmológico, senão a Copérnico, um dos iniciadores da mesma cosmovisão clássica, revolucionada por Einstein e pelos (demais) físicos quânticos? Certo, em 1916-17 ele apenas podia proceder assim, por razão óbvia de constituição e divulgação lenta da física moderna[16]. Quanto a nós, não temos mais tal (boa) desculpa.

Interessante como Freud elaborou o seu argumento já "descentrante" por meio de três "círculos concêntricos" (digamos assim): o cosmológico, o da vida e o da psique ("casa" em que o ego já não é o senhor, na metáfora arquitetônica do autor de *Psicopatologia da vida cotidiana*). De modo inadvertido, eles formam um autêntico mandala, mas de um jeito paradoxal: um mandala que, ao contrário dos tradicionais, que simbolizam totalidade e ordem, enfatiza o completo desamparo do sujeito. Coloquemos em questão o primeiro dos "círculos", parece-nos que o mais poderoso deles, *não egocentricamente falando*.

Num dos seus textos de divulgação científica, o bioquímico e ficcionista Isaac Asimov propôs-se a pergunta: "Existe um centro do universo?". Resposta dada:

> Apesar de todas as evidências, o fato é que *não existe tal centro do universo*, porque a expansão do universo não ocorre no costumeiro *padrão tridimensional*, mas num *quadrimensional*, o qual inclui, além das *três dimensões normais do espaço comum* (comprimento, largura e altura), a *quarta dimensão do tempo*. É difícil imaginar uma expansão em quarta dimensão [...].
>
> [...] o local no universo em que a expansão *se iniciou* não está *em nenhuma parte do espaço tridimensional* do universo que podemos percorrer, mas *bilhões de anos no passado*, e não podemos visitá-lo, embora tenhamos *informações* a seu respeito [...].[17]

[16] Para evitar qualquer absurdo anacronismo aqui, frisemos: a teoria da relatividade restrita ou especial foi apresentada por Einstein ao mundo em 1905; a geral, em 1916; embora iniciada em 1900 por Max Planck, a mecânica quântica se estruturou de maneira mais completa tão-só em 1927, com a chamada Escola de Copenhague, liderada por Niels Böhr.

[17] Isaac Asimov, *111 questões sobre a terra e o espaço*. Trad. Ieda Morriya. São Paulo: Best Seller, Círculo do Livro, [s.d.], 259, destaques nossos.

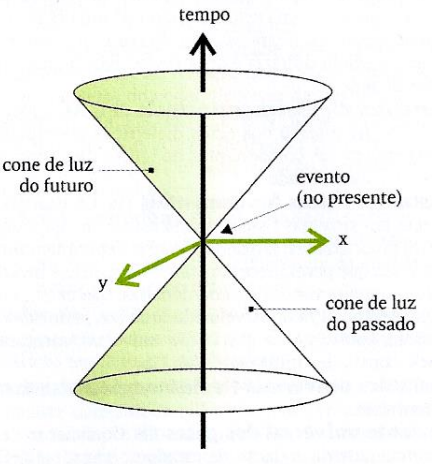
Estamos fora do "centro do universo", segundo o Freud copernicano; todavia, de acordo com o juízo do Asimov (implicitamente) einsteiniano, o cosmo atual é de outro modo descentrado (ou *multicentrado*, como notaremos adiante, com o auxílio dos físicos Lawrence M. Kraus e Roger Penrose, o que muda um bocado de coisas, em matéria de consideração da realidade). O que decorre daí? Ora, muito do *pathos* intelectual do descentramento, verdadeiro drama laico em que não deixa de haver uma nova "queda" da humanidade como um todo, perde a razão de ser, "desdramatiza-se", uma vez que agora *não parece existir em parte alguma a centralidade cosmofísica tridimensional* (ou mesmo algo como a sua *ausência*, em termos de *mera tridimensionalidade*) de que teríamos sido despejados ou em cujo interior um dia (que durou séculos) tivemos a ilusão de habitar. De fato, por causa de, na aparência, o sol, a lua e mais alguns astros moverem-se sobre as cabeças dos humanos, os quais não viam o chão sob os seus pés movimentar-se junto com os objetos celestes, ao menos desde a Antiguidade eles acreditaram na miragem do geocentrismo, já implícito na filosofia de Anaximandro de Mileto, que supunha fosse a Terra uma coluna cilíndrica, flutuando no centro de tudo. No Ocidente pagão, nomes como Eudoxo de Cnide, Platão, Heráclides do Ponto, Aristóteles e Ptolomeu, entre outros, consolidaram essa miragem compreensível (porque baseada na nossa percepção costumeira do mundo), desafiada pelo minoritário Aristarco de Samos, cujo heliocentrismo será retomado, no século XVI, por Copérnico, sem menção ao seu formulador na era antiga. Sabemos como – unindo as concepções aristotélicas às cristãs, que, na Idade Média, herdaram parte da cultura do paganismo – Santo Tomás de Aquino solidificou ainda mais a ilusão de que o empíreo girava em torno da Terra. Isto fortaleceu, sem dúvida, o narcisismo de criaturas cristãs que passaram a crer na hipótese de que um Criador arquitetou um macrocosmo ao redor delas.

O geo-, o bio- e por fim o psiconarcicismo teriam que vir mesmo abaixo, com o avanço científico de que Copérnico se transformou num símbolo, seguido logo por Galileu, condenado pela Igreja por dar sequência lógica e experimental às ideias copernicanas. Tal avanço se fez, em geral, *contra as aparências* de a realidade ser assim ou de qualquer outro modo, desde que correspondesse consideravelmente ao mundo observado pelos nossos órgãos sensórios, correspondência logo generalizada em normas pela razão (aliás, com notável sucesso prático, não poucas vezes), o que passava a viver-se como algo *intuitivo*. (Mesmo o heliocentrismo copernicano será

invalidado, entretanto, pois irá descobrir-se que à roda do sol giram apenas alguns corpos celestes.)

O narcisismo cósmico foi ferido de modo mortal, sem dúvida, ao se (re)questionar o movimento ilusório dos astros em volta do nosso planeta, no quinhentismo. O antinarcisismo filosófico extremo *da área das ciências humanas*, que, no século XX, surgiu com a brilhante intervenção de Freud em 1916-17, também precisa ser golpeado, com a *crítica às aparências* de que o cosmo possua deveras um "padrão tridimensional", como a passagem citada de Asimov assinala. *Não para regredirmos a uma concepção do sujeito pré-freudiana*, da espécie que volta e meia irrompe no horizonte intelectual, pouco (ou nada) perturbado em relação à sexualidade, mas nem sempre aberto ao inconsciente, fora das mesmas ciências humanas. Temos, porém, direito a uma noção de subjetividade que esteja mais de acordo com o que a física pós-newtoniana diz a propósito da realidade. (Tal física, afinal, está muitíssimo bem embasada em "pesquisas científicas", para usarmos a expressão freudiana de "Uma dificuldade no caminho da psicanálise").

Precisamos de algo novo, que seja mais do que uma fantasia *science-fiction* ou, como vem ocorrendo faz tempo, auto-ajuda empacotada com a terminologia das ciências. Haja então o (contra-intuitivo) *cone de luz*:



Cone de luz do espaçotempo (Figura 1)

Trata-se de um diagrama (um conjunto sígnico gráfico-linguístico)[18] que representa, de maneira *simplificada*, as três dimensões do espaço e a dimensão de tempo, de coordenadas não independentes umas das outras. Tal diagrama é oriundo da teoria da relatividade especial ou restrita, lançada em 1905 por Albert Einstein e logo (1907) interpretada por Hermann Minkowski (antigo professor de Einstein) no sentido de um *continuum* espaçotemporal, sem precedente no quadro da física newtoniana.

Por não conseguirmos, com o nosso aparato sensorial, visualizar uma verdadeira realidade de quatro dimensões, aceitemos uma simplificação gráfica: apenas um *par* de eixos dispostos na horizontal (*X* e *Y*) representa as três dimensões *espaciais* (comprimento, largura e altura); um eixo vertical simboliza o *tempo*. Há, na verdade, dois cones no diagrama, unidos pelos seus vértices: o de cima retrata os eventos do futuro; o ponto em que os seus vértices se encontram é o presente, onde se acha o observador; o cone da parte inferior da figura vale pelos eventos do passado.

O ângulo de inclinação de 45º do cone decorre do fato de a luz viajar a cerca de 300.000 quilômetros por segundo no vácuo. A teoria da relatividade restrita exige que a velocidade da luz seja absoluta (invariante) para todos os observadores, o mesmo ângulo servindo para todos os cones luminosos ou eventos envolvendo o presente, o passado e o futuro deste ou daquele indivíduo no universo.

Se uma informação veio do *interior* do cone de luz do passado, atingirá o observador postado no ponto em que os vértices dos dois cones se encontram (presente).

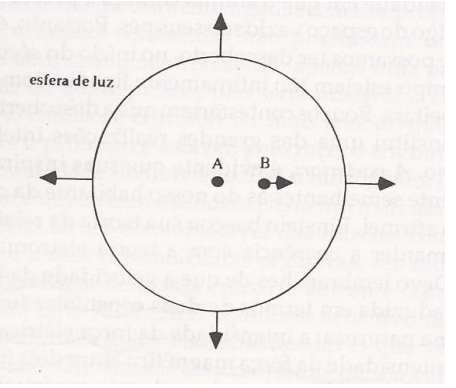
Se uma informação proveio do *exterior* do cone luminoso do passado, apenas atingirá o observador quando este se achar colocado num ponto que se situe *além* dos vértices onde os dois cones se encontram (um local no futuro).

Vejamos como o espaçotempo envolvido no diagrama revela os seus efeitos relativísticos. Uma boa explicação dessa espécie de fenômeno aparece no livro *Sem medo da física*, de Lawrence M. Kraus:

> Imagine dois observadores [uma Sra. A e um Sr. B] em movimento relativo que passam um pelo outro no instante em que um deles está acendendo um interruptor de luz. Sairá uma concha esférica de luz em todas as direções para iluminar a noite. A luz se desloca com tanta rapidez que nós normalmente não temos consciência que ela leve qualquer tempo para sair da fonte, mas leva. A

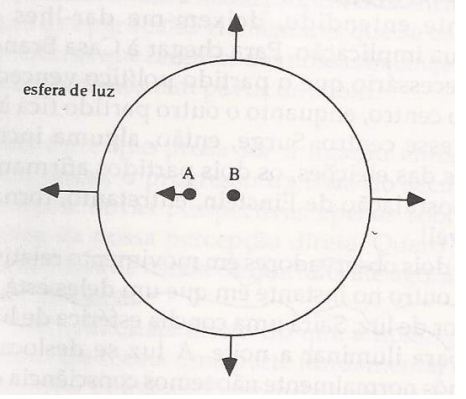
[18] In: Itzhak Roditi, *Dicionário Houaiss de física*. Ilustrações de Veronica Françoise Teicher. Rio de Janeiro: Objetiva, 2005, 49.

observadora A, em repouso em relação à lâmpada, veria o seguinte logo depois de a luz ser acesa:



Esfera de luz da perspectiva da observadora A (Figura 2)

Ela [A] se veria no centro da esfera de luz, e o observador B, que está se deslocando para a direita em relação a ela, teria se deslocado um pouco no tempo que a luz levou para se propagar até a sua posição atual. As medições do observador B, por outro lado, revelarão que esses mesmos raios de luz que estão se deslocando para fora [da esfera de luz] têm a mesma velocidade fixa em relação a ele e, portanto, percorrem a mesma distância para fora em relação a ele, segundo a postulação de Einstein. Portanto, ele se verá no centro da esfera, e A se deslocando para a esquerda do centro.



Esfera de luz da perspectiva do observador B (Figura 3)

Em outras palavras, ambos os observadores afirmarão estar no centro da esfera. A nossa intuição nos diz que isso é impossível [embora seja verdadeiro, relativisticamente falando!]. [...]

> [...] [Há um] absolutismo da velocidade da luz. [...] Não podemos estar *aqui* e *lá* ao mesmo tempo. A única maneira de descobrirmos o que está acontecendo *lá* agora é receber algum sinal, como um raio de luz. Mas se o recebermos *agora*, ele terá sido emitido então [...].
>
> A relatividade nos diz que, na realidade, os observadores que estão se deslocando em relação um ao outro *não* podem sentir o mesmo *agora*, mesmo que ambos estejam *aqui* no mesmo instante.[19]

Leiamos também Roger Penrose a respeito do assunto, em *A nova mente do imperador*:

> É conveniente, muitas vezes, descrever a luz em termos de *partículas* – chamadas *fótons* – e não em termos de ondas eletromagnéticas. [...] No espaço livre, os fótons viajam sempre em linhas retas com a velocidade fundamental *c*. Isso significa que no quadro do espaçotempo de Minkowski a linha-mundo de um fóton é sempre mostrada como linha reta com inclinação de 45º na vertical. [...]
>
> Essas propriedades são válidas, geralmente, em todos os pontos do espaçotempo. Não há nada de especial sobre a origem [de propagação de um fóton]: o ponto *O* [do observador no presente] não é diferente de nenhum outro ponto. Assim, deve haver um cone de luz em todos os pontos do espaçotempo, com o mesmo significado que tem o cone de luz na origem. A história de qualquer raio de luz [...] está sempre ao longo do cone de luz em cada ponto, ao passo que a história de qualquer partícula material deve estar sempre dentro do cone de luz em cada ponto [...]. A família de cones de luz em todos os pontos deve ser considerada como parte da *geometria minkowskiana* do espaço-tempo.[20]

Estas explicações tanto podem ser fastidiosas quanto fascinantes, dependendo da expectativa dos leitores. Perante as mesmas, uma coisa parece inegável, porém: em matéria de golpe cosmológico, o mandala freudiano paradoxal (por já colocar o ser humano numa situação de descentramento) perde um dos seus "círculos", o mais amplo deles, aliás, pois, queiramos ou não, no presente da vida de cada um de nós, *estamos sempre no ponto central de um cone de luz, no espaçotempo de Einstein e Minkowski*. E um novo mandala surge aqui – ou melhor, *muitíssimos*.

Lemos as afirmativas de Lawrence M. Kraus. Por sua vez, o físico Luiz Carlos de Menezes assinala:

---

[19] Lawrence Maxwell Kraus, *Sem medo da física*. Trad. Luiz Euclides Trindade Frazão Filho. Rio de Janeiro: Campus, 1995, 113-116, destaques do autor. (As duas ilustrações aqui utilizadas são deste livro.)

[20] Roger Penrose, *A nova mente do rei: computadores, mentes e as leis da física*. Trad. Waltensir Dutra. Rio de Janeiro: Campus, 1991, 216-217, destaques do autor.

> [...] Sobre o cone de luz, estão os fótons gerados no evento representado pelo vértice do cone. Os fótons são partículas de luz [...], que interessa aqui considerar por uma característica extrema: como têm a velocidade da luz, para elas o tempo não passa [...]. Se o vértice fosse o "Big Bang" e se detectássemos hoje um fóton "daquela época", para nós teriam se passado bilhões de anos, mas, para o fóton, o universo começou naquele instante![21]

Biografando Einstein, Jürgen Neffe ressalta:

> Como as partículas de luz não se movem no tempo, mas com o tempo, podemos dizer que elas não envelhecem. Para elas o "agora" significa o mesmo que "eterno". Elas vivem para sempre em seu instante[22].

O "tempo não passa para os fótons" (Menezes); eles "não envelhecem" (Neffe); há um "absolutismo da velocidade da luz" (Kraus). Abramos, portanto, bem os olhos para o sentido maior de tudo isto. Deveras, *primeiro* tiramos a coroa do geocentrismo e entronizamos o heliocentrismo, nenhum dos dois merecedores dessa majestade afinal, já que assentados em apenas três tridimensões, entre mais aparências. A teoria da relatividade restrita nos impõe *agora* (ou *há décadas*!) uma espécie de "fotocentrismo"**,** de caráter "hiperdemocrático", numa geometria minkowskiana que abriga a família de todos os cones luminosos com os respectivos aquis-e-agoras de cada evento particular, para adaptarmos os termos de Penrose.

O que resulta do tremendo achado relativista de 1905 (Einstein) e 1907 (Minkowski) é um *modelo multicentrado* para os observadores do universo, humanos e não humanos.

Quanto ao "golpe biológico", embora ainda seja muito cedo para extrairmos maiores conclusões a respeito, há algumas indicações de que a vida se vale de processos ligados a outra área da física moderna: não a relatividade, mas a mecânica dos *quanta*, como assinala o cientista Vlatko Vedral em "A vida em um mundo quântico"[23]. (Apenas ressaltemos que ali, na esfera atômica e subatômica, as coisas se revelam ainda mais surpreendentes!)

---

[21] Luis Carlos de Menezes, *A matéria: uma aventura do espírito*. São Paulo: Livraria da Física, 2005, 128.
[22] Jürgen Neffe, *Einstein: uma biografia*. Trad. Inês Antonia Lohbauer. Barureri, SP: Novo Século, 2012, 183.
[23] Vladko Vedral, "A vida em um mundo quântico". In: *Scientific American Brasil*. São Paulo: n. 110, Ediouro Duetto Editorial, 30-35, julho de 2011. Ver também o ótimo artigo eletrônico: Osvaldo Pessoa Jr., "A nascente biologia quântica", 02-07-2012 15:17. Disponível em: http://www2.uol.com.br/vyaestelar/biologia_quantica.htm.

E o "golpe psicológico"? Sigmund Freud garantiu que o ego não é *senhor* sequer *em sua casa*, precisando contentar-se com poucas informações a respeito do que ocorre na psique inconsciente. Ora, por obra e graça do pensador austríaco, sem falar em diversas outras contribuições, parte dos seres humanos tem a seu dispor informações, de variados graus de cientificidade, que lhe permitam tentar lidar melhor (ainda que esta tarefa não seja fácil) com a alteridade interna, com os seus "labirintos" emotivos, para lembrar a outra metáfora arquitetônica utilizada por Freud.

O traumatismo triplo sofrido pelo "narcisismo universal" pede um específico cuidado: algo que seja, em simultâneo, pós-traumático e pós-narcísico, ou melhor, não regressivo a respeito do que está correto, na crítica que acabamos de criticar.

Bem pe(n)sado tudo o que foi dito acima, não haver um centro psíquico em sentido clássico ou "cartesiano" (em que o ego fosse o *senhor* do nosso universo mental) é algo coerente com o fato de que o cosmo não apresenta, no final das contas, *um* centro situado num espaço de três dimensões, mas produz, a todo e qualquer momento (ou espaçotempo) da sua história, uma *profusão* de centros criados pelos cones luminosos em que os observadores, relativisticamente falando, acabem por localizar-se (ou ser localizados), nas quatro dimensões em que lhes cabe existir, mesmo que não as percebam assim, na sua vivência cotidiana. Que aprendamos, portanto, com uma postura mais acurada do ponto de vista científico, a sair de cosmovisões superadas, sem recaídas em egocentrismos também perecidos.

A "casa" (sem "senhor") a assumir-se é o universo, a realidade cada vez mais ampla que vamos conhecendo (mesmo ao ponto de colocar a Terra em risco, no processo!). O seu limite é luminoso, ao pé da letra: a luz, com os fótons que a compõem. Não somos mais especiais do que os outros componentes do cosmo, mas somos, relativística e quanticamente considerando, tão especiais quanto esses mesmos componentes, pois aqui nenhum deles é irrelevante, se atentarmos bem para a física moderna, não para os atuais volumes de auto-ajuda (ou os de autodepreciação datada).

Em suma: pensemos num *universo quadrimensional multicentrado*, sim, algo que *nos centra* a todo o instante – um cosmo com excesso de centros…

2. Abordagem de dois herdeiros (bem diversos) da problemática freudiana

Esperamos ter evidenciado que as reflexões de Freud, das quais deriva muito da ideia de descentramento, têm como um dos seus três suportes uma visão física do cosmo: a de Copérnico (e de seu precursor Aristarco), que foi absorvida pela mecânica de Newton, a qual, por sua vez, foi revolucionada pela teoria da relatividade, no que diz respeito a alguns aspectos decisivos, entre os quais o tempo e o espaço tridimensional, transformados num espaçotempo de quatro dimensões.

A concepção de descentramento desenvolveu-se no âmbito do estruturalismo e do pós-estruturalismo franceses, em especial por Jacques Lacan e por Jacques Derrida (neste bem explicitamente). Tal concepção continua causando impacto no terreno das ciências humanas, a julgar pela quantidade de material que se pode ler, em termos de textos impressos ou postados na *Internet*.

Quanto a Lacan, apesar de a sua produção ser a de um psicanalista (notoriamente *inovador*, que, por igual, preconizava um *retorno* a Freud), ela carrega implicações filosóficas consideráveis, pois ali o sujeito humano é visto a partir da perspectiva de um inconsciente que o ego cartesiano (o célebre "eu do *cogito*") não consegue dominar, disto resultando uma subjetividade não centrada na consciência, noção que acatamos, mas com restrições: não nos parece mais aceitável que, para a defesa teórica desse tipo de sujeito (ou psique), se permaneça atado a uma visão física do universo (a "revolução dita copernicana"[24]) que se tornou obsoleta, embora ainda siga parecendo evidente para muitos; julgamos que o referido ideal de sujeito deva ser repensado, *matizado* ao menos, a partir do que a ciência moderna do cosmo tem a dizer-nos, em especial no que concerne ao modelo multicentrado (envolvendo cada ponto do espaçotempo) que, com a ajuda dos textos de Kraus e Penrose, buscamos explicitar. Alguma coisa não combina, não vai bem conceptualmente, quando colocamos frente a frente um sujeito (*tão-só*) descentrado e um cosmo produtor de múltiplos "cones de luz", nos quais, a cada momento, o mesmo sujeito se acha no ponto central de um processo, no espaçotempo de Einstein e Minkowski, quer tenha ou

[24] Cf, Jacques Lacan, "A instância da letra do inconsciente ou a razão desde Freud". In: Eduardo Prado Coelho, *Estruturalismo: antologia de textos teóricos*. Trad. Maria Eduarda Reis Colares et al. Lisboa: Portugália, 1968, 257-288, esp. 279-281.

não consciência de tal fato[25]. Esta derradeira consideração não nos soa mais esdrúxula do que continuar enxergando com olhos copernicanos (e newtonianos) o que requer uma visão de maior abrangência, menos óbvia ou convencional (ainda que à disposição dos interessados faz um tempo considerável, medível já em décadas).

Mais do que a já complexa de Lacan (da qual extraímos o mínimo aqui necessário), a reflexão de Derrida sobre o descentramento aparece carregada de aspectos que concernem à filosofia (não fosse o autor franco-argelino alguém de tal área!). Muito difícil de sintetizar, também, pois, enquanto produziu, o prolífico Derrida, em coerência com o seu percurso de "desconstrução", desenvolveu estratégias de estilo ou de escrita nada facilitadoras de sínteses do seu sofisticado pensamento, o que se tornou uma das suas heranças filosóficas marcantes, bem vistas por uns, desdenhadas por outros (e entre os segundos não nos postamos, o que não implica aceitarmos tudo o que ele assinou). Apesar de tal dificuldade, em 1976, quando a obra derridiana ainda contava com um número pequeno de títulos (embora já fundamentais no seu trajeto, como *Gramatologia* e *A escritura e a diferença*), Silviano Santiago publicou *Glossário de Derrida*, elaborado em colaboração com vários dos seus alunos de pós-graduação da época. Neste pequeno dicionário utilíssimo, os leitores encontram o verbete "Descentramento" e correlatos. Examinando-os em cotejo com o famoso artigo (de 1966) "A estrutura, o signo e o jogo no discurso das ciências humanas" (*final* de título *sintomático*, que retomaremos)[26], os leitores mencionados poderão escrever os seus próprios textos sobre o assunto, o que não tem deixado de ocorrer, dada a fortuna em torno de tal escrito de Derrida.

No volume supervisionado por Silviano Santiago, notamos que o descentramento, tal como aparece na obra do pensador franco-argelino,

[25] Pontos intrigantes a serem considerados, numa reflexão que trate da questão mente-matéria, para além de qualquer cartesianismo: a luz interage com a matéria – e desta somos feitos; o nosso cérebro é eletromagnético – e o eletromagnetismo tem como base a luz (os fótons); também as células dos nossos corações atuam por meio de atividade elétrica – de novo a problemática da luz. Tudo isto dá bastante o que pensar, em termos de física e subjetividade.

[26] Neste artigo, o termo "descentramento" aparece de modo explícito. Cf, Jacques Derrida, "A estrutura, o signo e o jogo no discurso das ciências humanas". In: Eduardo Prado Coelho, *Estruturalismo: antologia de textos teóricos*. Trad. Maria Eduarda Reis Colares et al. Lisboa: Portugália, 1968, 101-123, esp. 104 e 112. (Tal artigo foi primeiro apresentado como conferência, proferida por Derrida em Baltimore, em 1966.)

além de apresentar-se como uma prática de "leitura intertextual", opõe-se "aos conceitos clássicos de estrutura *centrada*, *origem* e *presença*" e que "a atividade interpretativa" faz-se "eliminando-se qualquer referência a um centro, a um sujeito", tudo isto no preciso verbete "Descentramento"[27], ao qual convém percorrer em paralelo com o relativo à "Desconstrução" (por coincidência – ou ordenação alfabética – na página seguinte do *Glossário*), onde se postula que a "leitura desconstrutora" (ou "leitura descentrada") tem como "proposição radical" a de "anulação do centro como lugar *fixo* e *imóvel*"[28].

Neste passo, alguns pontos decisivos se deixam reter, em termos da problemática ora abordada.

Um primeiro ponto: o descentramento diz respeito a uma explícita, forte prática de *leitura*, uma das marcas da atuação de Derrida no âmbito da filosofia. Leitura envia a *textos*, a *signos*, a *escritura*, aos quais o pensador, como se sabe, deu enorme atenção, analisando as obras alheias com uma consideração extrema não apenas aos seus significados, mas por igual aos seus detalhes gráficos, aos pés-de-página, às comas, até aos "brancos" do papel (admirador de Mallarmé que ele também era), os quais deveras significam – e às vezes bastante! Eis outro dos seus legados, das suas perícias (que foram também as de vários autores ligados ao estruturalismo e ao pós-estruturalismo, mas que na sua pessoa encontraram um dos praticantes mais espantosos). Se, por exemplo, lançássemos um olhar derridiano ao presente artigo, ressaltaríamos a sua "textualidade", mesmo nas páginas em que nos valemos de diagramas, como o do cone de luz – e nada garante que os mesmos não pudessem ser alvos de uma "leitura desconstrutora", signos gráfico-linguísticos que são... Do que duvidamos é que algum executor de tal leitura – entendida esta tão-só em termos de um "discurso" feito no interior das fronteiras das "ciências humanas" (cf. o sintomático final do título do artigo de Derrida citado mais acima) – conseguisse efetivamente "desconstruir" a ampla problemática espaçotemporal do cone de luz relativístico, oriundo de uma ciência como a física, a qual traz na sua bagagem, além da carga conceptual, tanto a matemática (uma forma de escrita, sim, ainda que mais do que apenas "alfabética") quanto a experimentação (reti- ou ratificadora). Levemos a sério que, de um ponto de vista estritamente filosófico, alguém possa perceber

[27] Silviano Santiago, *Glossário de Derrida*. Rio de Janeiro: Francisco Alves, 1976, 17, destaques nossos.
[28] Silviano Santiago, op. cit., 1976, 18, destaques nossos.

inconsistências importantes numa teoria científica, mas, para invalidá-la por completo, precisará de outro *corpus* científico (às vezes, inclusive, de um novo paradigma de ciência), e a relatividade restrita (para não mencionar a geral) de Einstein vem sobrevivendo quer às postulações concorrentes, quer aos experimentos que testam a sua validade[29]. Se tudo isto for visto como um "texto", um "constructo complexo", eis um que não se tem deixado desconstruir, com o seu cone de luz, o qual nos induz a pensar numa centralidade cósmica *plural* que diz respeito a muitas coisas do universo.

Um segundo aspecto do *Glossário de Derrida* citado concerne à "anulação do centro como lugar fixo e imóvel". Ora, lemos em Penrose que o ponto de propagação de um fóton *nada apresenta de especial*, em relação a outro ponto. Assim, por certo, tanto em termos do filósofo quanto da relatividade restrita, não haverá centro "*fixo e imóvel*" (pois a luz se propaga, comportando-se como um limite para as demais propagações ao longo do espaçotempo ou cone luminoso quadrimensional), embora pareça inegável que produza efeitos de *multicentramento*, "famílias de cones de luz" (nas palavras já referidas de Penrose). Este último é elemento que conduz à discordância nossa em relação a um item importante de Derrida (afinal, um dos criadores da concepção de descentramento, talvez até lançador do termo), tal como não conseguimos concordar com Freud e Lacan páginas acima, deslocando-nos para o interior da área denominada filosofia das ciências, para além do campo das humanidades (ainda que em diálogo com estas, claro).

"A estrutura, o signo e o jogo no discurso das ciências humanas", ou o texto assinado pelo próprio Derrida, substitui o trio "Nietzsche, Freud e Marx", de Foucault, por "Nietzsche, Freud e Heidegger", como "nomes próprios" de autores cujas obras produziram o descentramento que o filósofo da desconstrução veio explicitar: a "crítica nietzschiana da metafísica, dos conceitos de ser e de verdade [...]; a crítica freudiana da presença a si, [...] da consciência do sujeito, da identidade a si, da proximidade ou da propriedade de si; e, mais radicalmente, a destruição heideggeriana da metafísica, da onto-teologia, da determinação do ser como presença"[30]. O autor que mais impulsionará a reflexão derridiana de 1966 será, entretanto, um quarto: o

[29] O que não garante que ela siga incólume no futuro, até mesmo imediato. O século XX deixou claro que nenhuma teoria tem tal garantia, com as suas tremendas revoluções científicas, como as duas relatividades de Einstein (1905 e 1916), as quais poderão vir a ser revolucionadas, por sua vez. Se o forem, apostemos que o serão, necessariamente, por teorizações ainda mais espantosas.

[30] Jacques Derrida, op. cit. na nota 26, 1968, 104.

etnólogo Claude Lévy-Strauss, com as implicações estruturalistas da sua obra.

Lendo os trabalhos do antropólogo, Derrida distingue uma visão clássica da concepção de estrutura e uma visão nova, a do estruturalismo de então (novidade que, depois, irá tornar-se importante para os que se enxergarão como pós-estruturalistas). A estrutura clássica teria sempre um centro, e a história se encarregaria de substituir esse centro por outros (nada menos que "essência, existência, substância, sujeito" e ainda "transcendentalidade, consciência, Deus, homem, etc."[31]). Como lugar privilegiado, o centro furtar-se-ia, contudo, ao jogo combinatório e à permuta de elementos, típicos da sua estrutura maior: ele assim escaparia à estruturalidade que comandaria. Após as produções de Nietzsche, Freud e Heidegger, o que o estruturalismo lévy-straussiano questionava, segundo Derrida, era a necessidade de existência dessa espécie de centro. Um fator decisivo para tal questionamento foi o interesse de tal estruturalismo pela problemática dos signos, sobretudo por causa do impacto da reflexão de Ferdinand de Saussure a respeito dela. De fato, a linguagem, que, desde cedo, fez parte das preocupações do pensamento ocidental (e não só), passou a ser um dos temas filosóficos importantes da época – e um signo, ao estar no lugar de algo, não é exatamente uma presença, um centro do que quer que exista.

Lendo o trabalho de Derrida (que ora segue, ora critica Lévy-Strauss, como era o seu costume), é impossível não admitirmos, por nossa conta, que muito dos elementos do mundo em que vivemos não parecem mesmo apresentar centros: a história humana, os mitos, as línguas, os demais sistemas de signos, os signos que não formam sistemas, as culturas, a própria vida... – e assim por diante. Mas a lista não terá um limite, *por maior que ela seja*? Antes de retomar a questão do espaçotempo relativístico, com o seu cone de luz, vejamos a seguinte passagem do texto derridiano, concernente à "pesquisa" efetuada por Lévi-Strauss a respeito dos mitos: "Com efeito, o que parece mais sedutor nesta pesquisa crítica de um novo estatuto é o declarado abandono de toda referência a um *centro*, a um

[31] Jacques Derrida, op. cit., 1968, 103. Críticos de Derrida poderiam dizer que, nesses centramentos que ele critica, tudo – da "essência" ao "etc.", passando pelo "sujeito" e pelo "homem" (até "Deus"!) – foi, afinal, nivelado, sem consideração pelas diferenças históricas, histórico-filosóficas, envolvidas nos termos. Defensores dele contra-argumentariam lembrando que, apesar de tais diferenças, a noção de centro se manteve para esses termos na história (européia) da metafísica, o que geraria novos ataques dos detratores, etc.

sujeito, a uma *referência* privilegiada, a uma origem ou a uma arquia absoluta"[32].

Sabemos: a "pesquisa crítica" visada é mesmo a do autor de *O pensamento selvagem* e mais livros influentes. A passagem em causa tornou-se consideravelmente citada, não poucas vezes deixando-se em segundo plano, todavia, a *referência efetiva* que ela faz à produção de Lévi-Strauss (ironia involuntária dos seguidores do filósofo da desconstrução? ato falho?)[33]. Como se o importante no trecho (o seu sentido *privilegiado*) começasse no passo "o abandono declarado de...". Vimos M. Kraus referir-se ao "absolutismo da velocidade da luz", num contexto teórico em que tempo e espaço se tornam conceitos relativísticos, dos quais a significação do absoluto foi, pois, descartada, o que é algo compatível como uma "leitura desconstrutora"; todavia, há no referido contexto precisamente o desafiante "*absolutismo* [invariância] da velocidade da luz" (recaída na metafísica? ingenuidade do "cientificismo"?)... O inegável é que a luz possui várias características espantosas: mostra-se uma limitação para o que ocorre no espaçotempo; revela-se um fator decisivo na famosa equação relativística $E = mc^2$, que trata da "equivalência" entre massa (*m*) e energia (*E*); quiçá os seus fótons não envelheçam; apresenta um comportamento estranho, a dualidade onda-partícula, ou seja, exibe predicados *contraditórios* (onda é algo que se espalha, partícula é algo que se localiza com mais precisão[34]), o que nos conduz ao terreno da mecânica quântica, fora da alçada do artigo presente... Difícil, portanto, não notar a luz como *uma* das "referências privilegiadas" do universo, embora não a "origem" dele[35]. Conforme afirmamos em relação a Lacan, algumas coisas precisam ser *matizadas* aqui,

[32] Jacques Derrida, op. cit., 1968, 112.

[33] Aqui não esqueçamos: a pesquisa de Lévi-Strauss em tela dizia respeito à mitologia, ou seja, uma área em que, por causa da criação coletiva, anônima, dos seus produtos, potencialmente incessante, é mais fácil trabalhar com o abandono das referências citadas por Derrida. Não por acaso a mesma área foi uma das que mais serviram a C. G. Jung na elaboração do seu conceito de "inconsciente coletivo".

[34] A estranha dualidade onda-partícula (da luz como da matéria) é algo que faz lembrar a reflexão de Derrida. Em que aspecto, mais precisamente? Na sua postura de pôr em questão, a identidade pressuposta nos conceitos herdados da tradição, bem como a noção de presença (na qual não se encaixa bem a referida dualidade). Em tal sentido, Derrida é um filósofo tremendamente contra-intuitivo, adjetivo que somos obrigados a assimilar também no trato com as relatividades e o *quantum* da física moderna.

[35] Se existe uma "origem" do cosmo, ela é o que os cientistas chamam de singularidade, algo em que o espaçotempo não "funciona" mais, uma condição da qual praticamente nada se sabe.

nesta versão do descentramento. Quanto aos referidos "centro" e "sujeito", o que já dissemos implica que (apenas por ora?) não temos condição de *abandonar* tais noções de todo, ou ao menos não a primeira delas.

Baseada em parte no cone de luz relativístico, a trama do universo não deve ser a de "algo" com um centro, mas a de um "objeto" descomunal (talvez finito, porém ilimitado) com centros múltiplos e incessantes. Supomos que o modelo quadrimensional minkowskiano, que envolve a noção de centro(s) dessa maneira infindável e multiplicadora (enquanto a luz se propagar pelo cosmo), escapa à desconstrução derridiana da ideia de centro, sem prejuízo de tal desconstrução ser *válida para muitas coisas*. Nesse modelo, que se diagramatiza como o cone de luz, os centros *não* organizam a totalidade em movimento que é a estrutura do espaçotempo. Não a comandam. Eles é que são estabelecidos pelo referido cone.

3. Observação final

Efeito curioso, até mesmo perverso: repetidos como "mantras" os enunciados sobre o descentramento, ou apenas parafraseados pelos herdeiros ou partidários dos três célebres autores (o que não deixa de ser uma forma de repetição – ou iterabilidade, como diria Derrida), num processo de reificação linguística eles adquirem um valor "absoluto", alcance genérico ilusório, funcionando como uma falsa tautologia, com um potencial de aplicação irrestrita que, de fato, não possuem. Diluem-se e, em simultâneo, ganham um *status* de verdade superior, inquestionável, fazem-se "clássicos", por um imprevisível oximoro começam a funcionar como um novo "centramento", numa "neometafísica", conforme, aliás, pode ocorrer com quaisquer outros produtos sígnicos da tradição que herdamos – "cones de luz" aqui incluídos, é óbvio, embora estes careçam de maior divulgação, o que nos leva ao derradeiro item.

Ao término do presente artigo, gostaríamos de lembrar que, em 1959, C. P. Snow lançou o importante ensaio *As duas culturas*[36]. Nele, o autor (não por acaso físico e romancista) tratou da distância crescente entre as ciências naturais e as humanidades: intelectuais de um campo passaram a ignorar as conquistas dos intelectuais do outro e vice-versa, criando-se um abismo lamentável entre os dois setores.

[36] C. P. Snow, *As duas culturas e uma segunda leitura*. Trad. Geraldo Gerson de Souza. São Paulo: Editora da Universidade de São Paulo, 1995.

Parte considerável dos integrantes das ciências humanas aceita a concepção de descentramento sem questioná-la em nada, recitando-a ou apenas efetuando paráfrases ("traduções" boas ou ruins, não importa aqui) do que os seus criadores propuseram. A maioria esmagadora dos físicos aceita os cones de luz oriundos de Einstein e Minkowski, nos quais não se pode negar que a noção de centro tem importância, algo que se revela ainda mais relevante quando sabemos que tais estruturas físicas são "válidas, geralmente, em todos os pontos do espaço-tempo", para retomar outra vez as palavras de Penrose. Pertencendo o redator do texto presente ao terreno das humanidades, ele procurou contribuir, na medida das suas possibilidades, para o que o próprio Snow anteviu como o surgimento de uma "terceira cultura", apenas quatro anos depois da publicação do seu trabalho de 1959.

Em nosso terceiro milênio, não será desejo desmedido querer colaborar com os esforços que levem em conta as possíveis conexões do sujeito humano em particular (corpo e psique, tanto consciente quanto inconsciente), e da vida em geral, com o restante do universo: as ligações da nossa condição, e da dos demais seres vivos, com o que as ciências naturais e físicas vêm descobrindo a respeito da matéria e da estrutura da realidade, a qual requer ser entendida em sentidos micro e macrocósmico, e *não apenas* em termos de "mesodomínios" sociais, históricos, antropológicos, etc., newtoniamente localizados (ou com ênfase na separação física das coisas). As conquistas dessas ciências (teorias da relatividade e física quântica, sobretudo) precisam ser mais assimiladas pelas humanidades, sem desrespeito pela lógica das suas elaborações, nem tratamento delirante das suas descobertas[37].

[37] Cf. Alan Sokal, Jean Bricmont, *Imposturas intelectuais: o abuso das ciências pelos filósofos pós-modernos*. Trad. Max Altman. Rio de Janeiro/São Paulo: Record, 2001, passim.

Understanding Admissibility

George Masterton
(Philosophy Department, Lund University)
george.masterton@fil.lu.se

Introducing Admissibility

Chance is an enigmatic topic of philosophical interest; consequently, there are many strong opinions on the subject and very little agreement. That said, there is one point upon which all are agreed; whatever chance is, if the chance for *A* is known to be *x*, then it is *prima facie* reasonable to believe *A* to degree *x* and to act/bet accordingly. If you know the coin is fair, then you should be as prepared to bet heads as tails; if you know that it is biased so that the chance of heads is 2/3, then you should be prepared to bet on heads at odds of 2:1. There are almost as many notational variants of this principle as there are people who have written on the topic of chance, but for our purposes van Fraassen's is instructive.

> **Miller's Principle:** My subjective probability that *A* is the case, on the supposition that the objective chance [at *t*] of *A* equals *x*, equals *x*.
> Symbolically: $[C](A|ch_t(A) = x) = x$.[1]

van Fraassen's name for this principle is a little misleading; Miller's principle is actually a principle that relates probabilities at different linguistic levels. Miller's principle, roughly stated, is: let P_1 be a probability function defined on the object language, *A* be a sentence of that object language, *x* be a real on the unit interval and P_1 be a probability function defined on the meta-language; then $P_2(A|P_1(A) = x) = x$. Hence van Fraassen's Principle is a specific application of Miller's Principle – with $C = P_2$ and $ch_t = P_1$ – and not that principle in its full generality. In any case, van Fraassen took this principle to answer his 'how' question.

[1] Van Fraassen, 1989, 82.

> … I stated the fundamental question about objective chance: why and how should it constrain rational expectation? The 'how' is answered by Miller's Principle and its generalizations.[2]

Among those 'generalizations' that also answer his 'how' question, van Fraassen might have included Lewis' Principal Principle (PP).

> **Lewis' (PP):** Let *C* be any reasonable initial credence function. Let *t* be any time. Let *x* be any real number in the unit interval. Let $[X_t^A]$ be the proposition that the chance, at time *t*, of *A*'s holding equals *x*. Let *E* be any proposition compatible with $[X_t^A]$ that is admissible at time *t*.
> Then $C(A|[X_t^A]E) = x$. [3]

The two principles are more or less[4] identical save for the inclusion of admissible *E* in Lewis' PP. There are two justifications for admissible *E*'s inclusion: one that is generally accepted and another that is peculiar to those working within the Lewisian program on objective chance. The generally accepted justification is that the plausibility of reasonable credence tracking chances is thought to increase where such tracking is largely invariant to further conditionalization. For instance, if you know that the chance of heads on the next toss is 1/2, then your credence in the next toss landing heads should be 1/2 and should remain so even once you have found out that the coin is a 2 euro coin, that the temperature is 30 Celcius, etc. The other justification is that Lewis' RPP (see below) cannot be derived from the PP unless history and natural laws are (generally) admissible$_t$. Lewis, and those following his lead, rely on the derivation of the RPP from the PP to justify the former; thus admissibility is important in the justification of Lewis' RPP.

> **RPP:** Let *C* be any reasonable initial credence function. [Let $H_{tw}T_w$ be that proposition that holds at all and only those worlds historically and nomologically possible relative to *w* at *t*]. Then for any time *t*, world *w*, and proposition *A* in the domain of P_{tw}; $P_{tw}(A) = C(A|H_{tw}T_w)$.

Why is it important for Lewis that the RPP is justified? Lewis originally says of the RPP that it has the 'form of an analysis' of chance.[5] Later Lewis held this principle to state the definitive 'role' that something must satisfy if it is to count as objective chance.[6] I have argued elsewhere that this, together with

[2] Van Fraassen, 1989, 195.
[3] Lewis, 1980, 87.
[4] This qualification is needed as the Principal Principle is also restricted to initial reasonable credence functions, whereas "Miller's" Principle applies more generally to all reasonable credence functions
[5] Lewis, 1980.
[6] Lewis, 1994.

Lewis' Canberra Planer predilections, is enough to convince ourselves that Lewis held the RPP to be, or at least to motivate, an analysis of objective chance in terms of reasonable credence conditional on prevailing history and natural laws. [7] For Lewis, that such an analysis could be more or less derived from such an uncontroversial fact about chances as his PP was a boon. The same holds for many who are tempted by the *chance as ultimate belief* thesis[8]: the thesis that objective chances$_t$ are objective degrees of belief conditioned upon some ultimate evidence$_t$; in Lewis' case, prevailing laws and history$_t$. But this derivation[9] is only (generally) valid where $H_{tw}T_w$ is (generally) admissible $_{tw}$ [10]; one cannot derive the RPP from what van Fraassen refers to as Miller's Principle. Consequently, if a Lewisian wishes to justify an analysis of chance in terms of that belief that is reasonable given our ultimate evidence on the basis of credence's conformity to chances, then they need the concept of admissibility.

Unfortunately, Lewis explicitly failed to rigorously define admissibility; settling instead for the following rough and ready characterization:

> Admissible propositions are the sort of information whose impact on credence about outcomes comes entirely by way of credence about chances of those outcomes.[11]

Despite this inauspicious start, headway was made through the identification of two sufficiency conditions[12]:

1. *E* is 'as a rule'[13] admissible at time *t* (admissible$_t$) if *E* pertains entirely to times earlier than, or including, *t.*

2. E is admissible if it is a history-to-chance conditional.

He later allowed that the axioms and theorems of optimal systematizations of collections of history-to-chance conditionals (A.K.A. natural laws) are also

[7] Masterton, 2010.
[8] Williamson, 2008.
[9] Lewis,1980.
[10] I have found that writing on this and related topics is aided by a judicious use of subscripts for there is much indexing to worlds and times. Hence I use 'admissibletw' as an abbreviation of 'admissible at time t and world w' and likewise for other indexical concepts.
[11] Lewis, 1980, 94.
[12] Lewis, 1980, 92.
[13] This caveat covers such eventualities as the testimony of time travellers or infallible soothsayers. Their testimony before time t is part of our historyt, but as conditionalizing upon it must break the link between reasonable credence and chancet, so such historict occurrences must be inadmissiblet.

admissible.[14] With these sufficiency conditions he could rule $H_{tw}T_w$ (generally) admissible, and thereby, derive the RPP from the PP.

It took fourteen years before some of the many flaws in this initial characterization of admissibility were addressed. Lewis[15] – after prompting by Thau[16] – finally made two amendments to his concept of admissibility, together with one to his Principal Principle. The first amendment was to allow that admissibility admits of degrees: that a proposition can be more or less admissible.

> Admissibility admits of degree. A proposition *E* may be imperfectly admissible because it reveals something or other about future history; and yet it may be very nearly admissible, because it reveals so little as to make a negligible impact on rational credence.[17]

He then weakened the PP so it applies where *E* is admissible or 'nearly' so. Finally, his most important amendment was to agree with Thau[18] that admissibility$_t$ is relative: one proposition is admissible$_t$ for another, not admissible$_t$ *tout court*.

> [D]egrees of admissibility are a relative matter. The imperfectly admissible *E* may carry lots of inadmissible information that is relevant to whether *B*, but very little that is relevant to whether *A*.[19]

These are certainly substantial improvements, yet still there is no necessary and sufficient condition for admissibility. What we do now have is a fairly clear idea of some of the main features of the concept:

- Admissibility is indexical: there is an admissibility for every time, and possibly also for every world.
- Admissibility is relative: one proposition is admissible for another.
- Admissibility admits of degree: one proposition can be more or less admissible for another.
- A proposition is generally admissible$_t$ iff it is admissible$_t$ for every proposition for which a chance$_t$ is defined.
- Propositions that hold of the world prior to *t* are generally admissible$_t$ as a rule, and natural laws are generally admissible without exception.

Like a golden thread running through this list is Lewis' original characterization of admissible propositions as 'the sort of information whose

[14] Lewis, 1994.
[15] Lewis, 1994.
[16] Thau, 1994.
[17] Lewis, 1994, 486.
[18] Thau, 1994, 500.
[19] Lewis, 1994, 486.

impact on credence about outcomes comes entirely by way of credence about chances of those outcomes.

Other perspectives on admissibility

Other commentators have more or less followed Lewis' lead on admissibility. For instance, Bigelow[20] gave the following characterization of the concept immediately before Thau and Lewis' later amendments and one can see that they adhered closely to the then established view.

A proposition will be admissible [at *t*] iff it does not covertly smuggle in information about the future, information which, since it is about the future, might bear on the present$_t$ rational credence about outcomes in a way that short-circuits the normal route via the present rational credence about present chances of outcomes.

Then came Thau's insight that though Lewis' earlier characterization of admissibility was essentially correct, it missed the essential feature that admissibility is always relative to another proposition.

> A proposition is admissible [at *t*] if it doesn't provide direct information about the outcomes of chancy events that occur subsequently to *t*. […] A proposition is inadmissible with respect to another proposition if it provides direct evidence about it. [21]

As I have already stated, Lewis was so impressed with this insight that he immediately adopted it.

Around the same time, Hall[22] noted that a necessary condition for *E*'s *general* admissibility$_t$ is that either the chance$_t$ of *E* is undefined, or else it is 1. Halpin[23] concurred with this opinion shortly thereafter. The argument for this condition is simple: If the chance$_t$ of *E* is defined, then *E*'s general admissibility$_t$ requires *E* be admissible$_t$ for itself; and so $C(E|E, X_t^A) = x$. But this is only so for reasonable *C* if $x = 1$; consequently, either the chance$_t$ of generally admissible$_t$ *E* is undefined, or its chance$_t$ is 1. This is an important result as the general validity of the derivation of the RPP from Lewis' PP depends upon the general admissibility$_{tw}$ of prevailing$_{tw}$ history and natural law. Therefore, either historic$_{tw}$ propositions and natural laws$_w$ have a chance$_{tw}$ of 1, or their chance$_{tw}$ is undefined, if the RPP is to be derivable from

[20] Bigelow, 1993, 454.
[21] Thau, 1994, 493, 500.
[22] Hall, 1994.
[23] Halpin, 1998.

the PP. This is an important lesson in the context of the debate on Humean Supervenience; through the PP, reasonable credence constrains the chances of generally admissible propositions to trivial values. This places us on the horns of a dilemma: either we accept that some chances can be dictated by what reasonable credence allows, or we introduce chance gaps so that, at the very least, no chance is ever defined for generally admissible propositions. Often people have no problem accepting that historical propositions have trivial chances, but which choice one makes for natural laws is another matter entirely[24].

In the last decade or so there have been attempts to give a full definition of admissibility by Loewer, Hall and Hoefer. Hall's attempt is the most interesting and distinct of these. [25]

> *E* is admissible with respect to [initial reasonable] credence $C_{[\,]}$, proposition *A*, and time *t* iff $C_{[\,]}$ takes it as certain that the *t*-chances treat *A* and *E* as independent.
> Symbolically the "definiens" is:
>
> $$C\big(ch_t(A|E) = ch_t(A)\big) = 1.$$

To his credit, Hall's definition is formal and precise. Moreover, the definiens can be demonstrated to be a sufficient condition for the Principal Principle to apply; both from the premises Hall assumes and from the assumption that reasonable C conforms to $C(A|X_t^A) = x$. The latter demonstration builds on a rather tricky proof, originally by Skyrms (1988), where one establishes that any reasonable credence C that obeys $C(A|X_t^A) = C(A|ch_t(A) = x) = x$, must also obey $C\big(A\big|E, X_t^{A|E}\big) = C(A|E, ch_t(A|E) = x) = x$. But if C obeys the later and $C\big(ch_t(A|E) = ch_t(A)\big) = 1$, then it will obey $C(A|E, X_t^A) = C(A|E, ch_t(A) = x) = x.$, which is the PP. Hence, $C\big(ch_t(A|E) = ch_t(A)\big) = 1$ is a sufficient condition for the PP to apply for any C that respects $C(A|X_t^A) = x$. A final advantage of Hall's definition is that it is fairly obvious how it could be used to generate a definition of degree of admissibility.

[24] Though it is way beyond the scope of this paper, it turns out that the two solutions that have been offered to the greatest challenge facing Humean Supervenience, commonly known as The Bug, correspond to these alternatives. Lewis (1994)/Hall's (1994) response corresponds to restricting the chances for prevailing laws to unity and accepting the uncomfortable consequences for chances that follow from this, whilst Hoefer (2007) has offered a chance gap solution that corresponds to the latter alternative.

[25] Hall, 2004, 102.

> *E* is admissible to degree α with respect to [initial reasonable] credence $C_{[\,]}$, proposition *A*, and time t iff $C_{[\,]}$ takes it as certain that $1 - |ch_t(A|E) = ch_t(A)| = \alpha$.

However, that $C\big(ch_t(A|E) = ch_t(A)\big) = 1$ is sufficient for the PP to apply where C is reasonable is not enough to establish this condition as a definiens for admissibility. Moreover, admissibility defined in the manner outlined by Hall would be very different from how admissibility is typically conceived; Hall's definition has little to do with the manner in which one proposition informs on another via its chance. Both of these points indicate that $C\big(ch_t(A|E) = ch_t(A)\big) = 1$ may be unsuited to the task of defining admissibility. A further potential problem is that the degree of admissibility definition I extrapolated from Hall's definition of relative admissibility does not behave as we might expect. Consider a seer's testimonies on the results of a soon to be conducted toss of a fair die and a fair coin in a C that grants that the seer is infallible. According to that definition, the seer's testimony that heads will be the result of the coin toss is 3 times as admissible $\left(\alpha = \frac{1}{2}\right)$ as that same seer's testimony that the result of the die cast will be 5 $\left(\alpha = \frac{1}{6}\right)$ in such a C. But in both cases the reason for the seer's testimony's inadmissibility – namely, the granted entailment of the result in question by that testimony – is the same, so it is natural to expect the two testimonies to be equally inadmissible for their respective results. That they are not is a bit of an unwelcome surprise. True, the extrapolated degree of admissibility definition is not Hall's, and true, one might reconcile oneself to degrees of admissibility that depend on the chances involved, but still I take this as grist for the mill.

Hoefer's concern lay mainly with general admissibility, which he refers to simply as 'admissibility'. He offered two definitions of this concept:

> Any proposition (your "evidence") that does not contain information relevant to the outcomes of chance events except by containing information about their (objective) chances [is admissible]. [26]

> Propositions that are admissible with respect to outcome-specifying propositions A_i contain only the sort of information whose impact on reasonable credence about outcomes A_i, if any, comes entirely by way of impact on credence about the chances of those outcomes. [27]

[26] Hoefer, 1997, 324.
[27] Hoefer, 2007, 553.

The latter definition is of particular interest as it plainly contains within it a definition of relative admissibility; one that seems to be exactly like that proffered by Loewer. [28]

Loewer's definition, and the sufficiency condition for inadmissibility – which may as well have been given as a definition – that Loewer[29] draws from it, strike me as the best characterizations of the concept available to date.

> *Q* is admissible relative to *A* at time *t* iff *Q* provides information about *A* only by providing information about the chance of *A* at *t*. [30]
>
> Information about A is inadmissible if it is information about A over and above information about A's chance. [31]

Loewer's definitions of (in)admissibility capture the relative and indexical nature of the concept whilst at the same time incorporating Lewis' original specification in a succinct and tidy way. I particularly like the later 2004 condition for inadmissibility, which seems to me to capture everything currently understood about the concept in one concise sentence.

All that having been said, there is still substantial room for improvement. When is information about *A* information over and above information about *A*'s chance? How do we quantify the degree to which *A* informs over and above informing about *A*'s chance? At best, the culmination of 30 years of ruminating on this concept have provided us with a promissory note of a definition; one in which a great many details are still to be filled in and made precise. I now turn to the task of answering these outstanding concerns.

A framework for understanding admissibility

Our task here is to give a formal intensional definition of relative admissibility and its degrees. There are several desiderata against which such a definition might be judged and these can often be in tension. The following is a partial list of these desiderata:

Continuity: Ideally, a rigorous definition of an established concept should be as faithful as possible to its informal characterisations. Of course, if there is fundamental disagreement between these informal

[28] Loewer, 2001, endnote 5.
[29] Loewer, 2004, 1116.
[30] Loewer, 2001, endnote 5.
[31] Loewer, 2004, 1116.

characterisations, then this desideratum can only ever be partially satisfied.

Improvement: A new definition of an established concept should be an improvement on those that have been offered previously. The improvements sought herein are in terms of clarity, precision and quantifiability, though others might also be pertinent.

Informativity: An intensional definition that makes the extension of a term epistemically transparent is to be preferred to one that leaves such epistemically opaque.

Clarificatory: Where the extension of a term has been in dispute, it is a virtue of a definition if it reveals such disagreement to be rational or, otherwise, explicable.

Coherence: Any definition must be logically consistent.

Because these, and other, desiderata might be conflicting, definitions have to be judged in terms of the balance they strike between them. Unfortunately, commentators are likely to value the desiderata differently, differ in terms of how they measure definitions against them and even differ in terms of how they balance them. This all makes it very difficult, if not impossible, to propose a definition that will meet with every commentator's approval.

That having been said, it can be worthwhile to try and give a precise, intensional definition of a disputed concept in order to clarify the terms of the dispute. In this spirit I shall suggest a basic framework for constructing definitions of relative admissibility and its degrees based on the probabilistic notions of conditional independence and resiliency. Conditional independence is typically[32] defined as follows:

> If $P(A|B,C) = P(A|C)$ [where $P(B,C) > 0$], we say that *A* and *B* are *conditionally independent* given *C*; that is, once we know *C*, learning *B* would not change our belief in *A*. (Pearl, 2000, p.3).

Now consider the following substitution instance of the above.

> If $C(A|E,X_t^A) = C(A|X_t^A)$ [where $C(E,X_t^A) > 0$], we say that *A* and *E* are *conditionally independent* given X_t^A; that is, once we know X_t^A, learning *E* would not change our belief in *A*.

[32] The gloss given to the definition after the semi-colon is typical but might not be universally endorsed; some might argue that conditional independence is not equivalent to knowing C making credence in A invariant to learning B. For my purpose – conceptual analysis of Lewis' admissibility – it suffices that this gloss of screening off is common in the literature and that the connection between conditional independence and indirect informing is widely acknowledged.

Conditional independence is often known as *screening off*: i.e., X_t^A screens off *A* from *E* in *C* exactly where *A* and *E* are *conditionally independent* given X_t^A in *C*. Learning *E* after X_t^A would not change our belief in *A* iff *E* only informs on *A* via X_t^A, if at all. In the main (Lewis/Loewer) tradition on relative admissibility, *E* is admissible$_t$ for *A* iff *E* informs on *A* only via *A*'s chance$_t$. All this suggests that relative admissibility might be fruitfully defined in terms of screening off by chances in something like the following manner:

E* is admissible$_t$ for *A iff $C(A|E, X_t^A) = C(A|X_t^A)$ [where $C(E, X_t^A) > 0$].

We can simplify this proposal by noting that reasonable credence functions are regular according to Lewis (1980). Such regularity implies that $C(E, X_t^A) = 0$ if, and only if, *E* and X_t^A are inconsistent/incompatible. Where *E* and X_t^A are incompatible the Principal Principle does not apply (see earlier citation) and the question of *E*'s admissibility$_t$ for *A* is mute. It follows that the question of *E*'s admissibility$_t$ for *A* is only pertinent if $C(E, X_t^A) > 0$. Consequently, in any definition of admissibility in terms of conditional independence in initially reasonable C, the "where" clause above will be redundant and may be omitted giving:

E* is admissible$_t$ for *A iff $C(A|E, X_t^A) = C(A|X_t^A)$.

Recall that this is only a framework, and not a definition *per se*; many details still have to be resolved before the above can spawn a definition. Despite these flaws the above does already have some merits. Firstly, the definiens is precise and familiar to those with a working knowledge of probability theory. Secondly, there is a continuity between this framework and the definitions offered by Loewer, Lewis, Hoeffer, etc through the oft assumed association between 'C screening off *A* from *B*' and '*B* informing on *A* only via C'. Thirdly, if admissibility is defined in terms of screening off, then degrees of admissibility are naturally defined in terms of degrees of screening off. Skyrms (1977) introduced *resiliency* as a measure of the degree to which one proposition screens off another from a third, so all we need do in order to define a measure of admissibility$_t$ is to co-opt Skyrms' notion of resiliency:

***E* is admissible at *t* for *A* to degree α** iff

$$1 - |C(A|E, X_t^A) - C(A|X_t^A)| = \alpha,$$

equivalently:

E* is admissible at *t* for *A* to degree *α iff credence in *A*, given the chance$_t$ of *A*, has resiliency over *E* of α.

It naturally follows within this framework that *E* is admissible$_t$ for *A* iff *E* is admissible at *t* for *A* to degree 1; or, to paraphrase Skyrms (1977):

> To say that $[X_t^A]$ *shields-off* [*A*] from [*E*] is to say that $[C(A|X_t^A)]$ has resiliency 1 over *E.*

All of this is promising, but we have yet to produce a definition. This becomes clear when we attempt to formalize the above framework as a definition schema. Strictly speaking the "definition" and "principles" offered in this text, and more generally in the literature as a whole, are not definitions and principles *per se* but rather definition and principle *schemata*: i.e. representations of sets of related definitions and principles. Other notable examples of such schemata are Tarski's T-schema, the standard definition of conditional probability, $A \rightarrow (A \vee B)$ of propositional logic, etc. A schema, generally, is a system consisting of two parts: a *schema-template* and a *side note*. The former can be thought of as being comprised of three types of component: schematic constants, schematic variables and quantifier-bound variables. It is important to recognize that the schematic variables of a schema-template and its quantifier-bound variables are very different: schematic variables range over formulas whereas quantifier-bound variables range over objects in the universe of discourse. For instance, the universal instantiation schema – $\forall x[A(x) \rightarrow A_a^x]$, where *A* is a formula of a first-order propositional language, *x* is a variable, *a* is a term substitutable for *x* in *A,* and A_a^x is the formula obtained once *a* has been substituted for *x* in *A(x)* – would make little sense unless there was a distinction between the schematic variables *A* and *a*, the schematic constants ∀ and → and the quantifier-bound variable *x*.

While a schema-template is a purely syntactic object – a string-type with string-tokens for every permutation of the schematic variables – its attendant side note expresses a proposition. This proposition determines the appropriate interpretation of the instances of the schema in question; the proposition – definitions, principles, axioms, etc – expressed by the schema instances. To this end, the side note gives the domains for the schematic variables, tells us how to read the schematic constants, and gives the intended domains of any quantifier-bound variables present.

Returning to the proto-definition of relative admissibility offered earlier we have the proto-schema template:

E* is admissible$_t$ for *A iff

$$C(A|E, X_t^A) = C(A|X_t^A).$$

This proto-template is meaningless without an attendant side note identifying the schematic constants and variables and quantifier bound variables, as well as the relevant domains of the latter. Plainly, the language of the schema instances is English and the schematic constants are 'is', 'admissible', 'for', 'iff", '(', '|', ')' and '='. Equally plainly, *E*, *A* and *t* are schematic variables ranging over designations of propositions and times, respectively. But how are C and X_t^A to be read in the schema? Obviously, they are not constants of the schema, so they are either quantifier bound variables – where the quantifiers have yet to be added to the schema template – or else they are variables of the schema. If C is a schematic variable, then it will range over designations of reasonable initial credence functions, if it is quantifier bound in the schema then its intended domain will be the reasonable initial credence functions. If X_t^A is a schematic variable, then it will range over designations of chance$_t$ of *A* propositions; if quantifier bound, then its intended domain will be the chance of *A* at *t* propositions.

We can quickly exclude three definition schemata if we consider the four schemata templates where C and X_t^A are quantifier bound.

1. E* is admissible$_t$ for *A iff $\forall C \forall X_t^A [C(A|E, X_t^A) = C(A|X_t^A)]$.

2. E* is admissible$_t$ for *A iff $\forall C \exists X_t^A [C(A|E, X_t^A) = C(A|X_t^A)]$.

3. E* is admissible$_t$ for *A iff $\exists C \forall X_t^A [C(A|E, X_t^A) = C(A|X_t^A)]$.

4. E* is admissible$_t$ for *A iff $\exists C \exists X_t^A [C(A|E, X_t^A) = C(A|X_t^A)]$.

All these templates have the same side note; namely: The language of this schema's instances is English, the intended domain of C is the reasonable initial credence functions, the intended domain of X_t^A is the chance$_t$ of *A* propositions – $\{ch_t(A) = x : x \in [0,1]\}$ –, and the schematic variables are *E* and *A* – ranging over designations of propositions – and *t* – ranging over designations of times.

The third and fourth such definition schemata are obviously too weak, as it surely cannot suffice that *E* is screened off from *A* in a single credence function by a (all) chance$_t$ of *A* proposition(s) for it to be admissible$_t$ for *A*. We can also rule out the second. As $C(A|E, X_t^A)$ is either greater than, or equal to, $C(A|X_t^A)$ when the latter is equal to zero and less than, or equal to, $C(A|X_t^A)$

when the latter is equal to 1, so there must exist an X_t^A such that $C(A|E, X_t^A) = C(A|X_t^A)$ in every C. Hence the second schema implies that everything is admissible for everything else and can be rejected on this account. Indeed, all schemata conforming to the framework where X_t^A is existentially quantified can be ruled out on this account.

So far we have only one viable definition schema for relative admissibility, but what about all those schemata where C or X_t^A are schematic variables as opposed to quantifier bound variables? Here we encounter a problem: As a definition schema represents a set of definitions – one for every permutation of the values of schematic variables –, so it follows that, if a schematic variable occurs in the definiens of a schema template but not in the definiendum, then there will be multiple definitions for the same definiendum. This can cause problems for, unless the definiens of each of these multiple definitions of the same definiendum are equivalent, such a schema will be incoherent. As a general rule, it is best to avoid such problems by ensuring that any schematic variable occurring in the definiens of a template also occurs in the definiendum, and vice versa. This assures a one to one correspondence of definiendum to definiens in the instances of the schema. Where C is concerned, this is the appropriate route to take. Maintaining X_t^A as a quantifier bound variable this gives another definition schema for relative admissibility:

5. E* is admissible$_t$ for *A* in *C iff $\forall X_t^A\ [C(A|E, X_t^A) = C(A|X_t^A)]$.

> The language of this schema's instances is English, the intended domain of X_t^A is the chance$_t$ of *A* propositions – $\{ch_t(A) = x : x \in [0,1]\}$ –, and the schematic variables are *E* and *A* – ranging over designations of propositions –, C– ranging over designations of reasonable initial credence functions – and *t* – ranging over designations of times.

Applying the same method we can construct two further schemata where X_t^A is a schematic variable.

***6. E* is admissible for A with respect to X_t^A** iff $\forall C\ [C(A|E, X_t^A) = C(A|X_t^A)]$.

> The language of this schema's instances is English, the intended domain of C is the reasonable initial credence functions, the schematic variables are *E* and *A* – ranging over designations of propositions – and X_t^A–ranging over designations of chance$_t$ of *A* propositions.

***7. E* is admissible for *A* in *C* with respect to X_t^A** iff $C(A|E, X_t^A) = C(A|X_t^A)$.

The language of this schema's instances is English, the schematic variables are *E* and *A* – ranging over designations of propositions –, C – ranging over designations of reasonable initial credence functions – and X_t^A –ranging over designations of chance$_t$ of *A* propositions.

The problem with this approach is that relative admissibility so defined is chance relative: there being a relative admissibility of *E* for *A* for every logically possible chance$_t$ of *A*. To this author's mind, and contrary to Meacham (2010), such chance relative admissibility is not sufficiently continuous with the literature on the subject to be acceptable. However, simply deleting 'with respect to X_t^A' from the definiendum in the above templates leads to the aforementioned problem of multiple non-equivalent definitions for one and the same definiendum. For example, two instances of schema 6 with 'with respect to X_t^A' deleted from the definiendum would be.

E* is admissible$_t$ for *A iff $\forall C[C(A|E, ch_t(A) = 0) = C(A|ch_t(A) = 0)]$.

E* is admissible$_t$ for *A iff $\forall C[C(A|E, ch_t(A) = 1) = C(A|ch_t(A) = 1)]$.

This implies that $\forall C[C(A|E, ch_t(A) = 0) = C(A|ch_t(A) = 0)]$ iff $\forall C\ [C(A|E, ch_t(A) = 1) = C(A|ch_t(A) = 1)]$. While there are values of *E* and *A* that satisfy this equivalence, there are also plenty that do not.

To get around this problem one can add a condition to the side note to ensure that the schematic variable X_t^A ranges over designations of a single chance$_t$ of *A* proposition. This move will make admissibility implicitly relative to some particular chance$_t$ of *A*, but to which chance$_t$ of *A* should it be so relative? There are many choices one could make here and which one feels appropriate seems to be more a matter of taste than anything else; indeed, any choice seems arbitrary. In any case, examples of the creed include:

8. E* is admissible$_t$ for *A iff $\forall C[C(A|E, X_t^A) = C(A|X_t^A)]$.

The language of this schema's instances is English, the intended domain of C is the reasonable initial credence functions, the schematic variables are *E* and *A* – ranging over designations of propositions – and X_t^A – ranging over designations of the proposition giving the actual chance$_t$ of *A*.

9. E* is admissible$_t$ for *A* in *C iff $C(A|E, X_t^A) = C(A|X_t^A)$.

The language of this schema's instances is English, the schematic variables are *E* and *A* – ranging over designations of propositions –, C – ranging over designations of reasonable initial credence functions – and

X_t^A – ranging over designations of the proposition giving the expected (in C) chance$_t$ of *A*.

This gives us four candidate schemata: 1, 5, 8 and 9, with the latter two serving as exemplars for further schemata. For each of these there is an associated degree of relative admissibility definition schema in terms of resiliency over chances. Where X_t^A is a schematic variable – as in schemata 8 and 9 – it is easy to proceed by directly co-opting Skyrms' definition of resiliency:

8. *E* is admissible$_t$ for *A* to degree α iff $\forall C[1 - |C(A|E, X_t^A) - C(A|X_t^A)| = \alpha]$.

The language of this schema's instances is English, the intended domain of C is the reasonable initial credence functions, the schematic variables are *E* and *A* – ranging over designations of propositions –, α – ranging over designations of reals on the unit interval – and X_t^A – ranging over designations of the proposition giving the actual chance$_t$ of *A*.

9. *E* is admissible$_t$ for *A* in *C* to degree α iff $1 - |C(A|E, X_t^A) - C(A|X_t^A)| = \alpha$.

The language of this schema's instances is English, the schematic variables are *E* and *A* – ranging over designations of propositions –, C – ranging over designations of reasonable initial credence functions, α – ranging over designations of reals on the unit interval – and X_t^A – ranging over designations of the proposition giving the expected (in C) chance$_t$ of *A*.

Where X_t^A is bound by a universal quantifier the task is more difficult, for the value of $C(A|E, X_t^A) - C(A|X_t^A)$ may vary depending upon the value that X_t^A takes. My proposal is to define degree of admissibility in terms of minimal resiliency given chances for such schemata: i.e., *E*'s degree of admissibility$_t$ for *A* is the least degree to which an X_t^A screens off *A* from *E*.

***1.* *E* is admissible$_t$ for *A* to degree α** iff

$$\forall C \begin{bmatrix} \exists X_t^A [1 - |C(A|E, X_t^A) - C(A|X_t^A)| = \alpha] \wedge \\ \forall X_t^A [\, 1 - |C(A|E, X_t^A) - C(A|X_t^A)| \geq \alpha] \end{bmatrix}.$$

The language of this schema's instances is English, the intended domain of C is the reasonable initial credence functions, the intended domain of X_t^A is the chance$_t$ of *A* propositions – $\{cht(A) = x : x \in [0,1]\}$ –, and the schematic variables are *E* and *A* – ranging over designations of

propositions –, t – ranging over designations of times – and α – ranging over designations of reals on the unit interval.

5. *E* is admissible$_t$ for *A* in *C* to degree α iff

$$\exists X_t^A[1 - |C(A|E, X_t^A) - C(A|X_t^A)| = \alpha] \wedge$$
$$\forall X_t^A[1 - |C(A|E, X_t^A) - C(A|X_t^A)| \geq \alpha].$$

The language of this schema's instances is English, the intended domain of X_t^A is the chance$_t$ of *A* propositions – $\{ch_t(A) = x : x \in [0,1]\}$ –, and the schematic variables are *E* and *A* – ranging over designations of propositions –, t – ranging over designations of times –, C – ranging over designations of reasonable initial credence functions – and α – ranging over designations of reals on the unit interval.

It can be easily checked for each of these schemata that they satisfy the criterion that *E* is admissible$_t$ for *A* iff *E* is admissible$_t$ for *A* to degree 1. The schemata 8 and 9 for degree of relative admissibility have the peculiar property that how admissible a soothsayer's proclamation about the future is depends upon the chance picked out by the condition in the side note. For instance, 8 implies that a soothsayer's prophecy that a coin to be flipped will land heads is less admissible than that same soothsayer's prophecy that a dice to be rolled will come up 6. The prophecies entail the outcomes so $C(H|P_H, X_t^H) = C(six|P_6, X_t^6) = 1$, but $C(H|P_H, X_t^H) = \frac{1}{2}$ and $C(six|P_6, X_t^6) = \frac{1}{6}$, in all initially reasonable C; hence the soothsayer's prophecy that the result of the coin toss will be heads is admissible to degree 1/2, while their prophecy that the dice roll will result in a six is admissible to degree 1/6. This is a decidedly peculiar way for degree's of admissibility to behave and reflects badly on not only the definitions of degree of relative admissibility canvassed above, but also their associated definitions of relative admissibility 8 and 9. Indeed, were the coin double-headed $(ch_t(H) = 1)$, the soothsayer's prophecy would be admissible$_t$ according to 8 precisely because of this chance relativity, and this is arguably also contrary to what one expects. Together with the fairly arbitrary nature of the side note conditions specifying to which chances the definitions are to be relative, one can argue that there is sufficient reason to reject definition schemata for relative admissibility and its degrees where X_t^A is a schematic variable. Accepting this argument – as I do – leaves only two candidate definition schemata conforming to the framework developed herein:

The Objective Schema:

E* is admissible$_t$ for *A iff $\forall C \forall X_t^A [C(A|E, X_t^A) = C(A|X_t^A)]$.

***E* is admissible$_t$ for *A* to degree α** iff

$$\forall C \begin{bmatrix} \exists X_t^A [1 - |C(A|E, X_t^A) - C(A|X_t^A)| = \alpha] \wedge \\ \forall X_t^A [1 - |C(A|E, X_t^A) - C(A|X_t^A)| \geq \alpha] \end{bmatrix}.$$

The language of this schema's instances is English, the intended domain of C is the reasonable initial credence functions, the intended domain of X_t^A is the chance$_t$ of *A* propositions: $\{ch_t(A) = x : x \in [0,1]\}$. The schematic variables are *E* and *A* – ranging over designations of propositions –, *t* – ranging over designations of times – and α – ranging over designations of reals on the unit interval.

The Subjective Schema:

E* is admissible$_t$ for *A in C iff $\forall X_t^A [C(A|E, X_t^A) = C(A|X_t^A)]$.

E* is admissible$_t$ for *A* to degree α *in C iff

$$\exists X_t^A [1 - |C(A|E, X_t^A) - C(A|X_t^A)| = \alpha] \wedge$$
$$\forall X_t^A [1 - |C(A|E, X_t^A) - C(A|X_t^A)| \geq \alpha]$$

The language of this schema's instances is English, the intended domain of C is the reasonable initial credence functions, the intended domain of X_t^A is the chance$_t$ of *A* propositions: $\{ch_t(A) = x : x \in [0,1]\}$. The schematic variables are *E* and *A* – ranging over designations of propositions –, *t* – ranging over designations of times – and α – ranging over designations of reals on the unit interval.

Understanding Admissibility and its Degrees

Both these schemata, the Objective and the Subjective, score reasonably well in terms of their continuity with the characterizations of relative admissibility to be found in the literature. They also improve on these aforementioned characterizations in their precision and clarity. Best of all, they both facilitate a natural definition of degree of relative admissibility; thereby, filling a lacuna in the literature. Finally, both schemata are perfectly coherent. Unfortunately, each has a flaw: the Subjective schema allows for

highly peculiar extensions of the admissibility predicate, while the Objective schema is uninformative.

For any particular agent's reasonable initial credence function C, pair of propositions *A* and *E*, and time *t*, we can verify for a great many chances$_t$ of *A* that those chances screen off *A* from *E* in C. We can then make an induction to all such chances doing the same. This will give us a good, but defeasible, reason to believe *E* admissible$_t$ for *A* in C according to the definition and so we may ascertain which propositions are admissible$_t$ relative to each other for a particular subject. Alternatively, we can see whether *A* is independent of *E* in C and then make an *a fortiori* argument from such independence to independence conditional on the chances, and so to *E*'s admissibility$_t$ for *A* in C. Finally, it is a consequence of this definition schema that *E* is inadmissible$_t$ for *A* in any C that grants any credence to the entailment of *A* by *E*. In brief, we can use the definition to help sort the admissible$_t$ from the inadmissible$_t$ for any *A* in C, and this makes the definition informative. However, by the very nature of this definition what is admissible for one person may not be admissible for another. If a person is convinced that any coin they toss on Tuesday's is bound to land heads, then for that person, the proposition that it is Tuesday will be inadmissible for the proposition that the result of coin toss to be made is heads. Indeed, one can imagine any number of examples where an agent's peculiar, but rational, beliefs give rise to strange extensions of the admissibility predicate for them. So while the Subjective schema is informative, it is unacceptably subjective.

This leaves only the Objective definition schema for admissibility in terms of screening off by chances as viable; unfortunately, this schema is uninformative. Suppose we claim that *E* is admissible$_t$ for *A*, then according to the objective definition schema what we are claiming is that initial reasonable credence is such that $C(A|E, X_t^A) = C(A|X_t^A)$, for all X_t^A. But how are we to verify this? The above is not implied by other generally agreed principles of reasonable credence, is not supported by a dutch book argument and cannot be ascertained empirically. It seems that the only way to ascertain whether or not initial reasonable credence is like this is simply to stipulate that this is so. Consequently, this definition schema is, for the most part, uninformative. The qualification of the preceeding is there because it follows from the Objective definition schema that, if *E* entails *A*, then *E* is always inadmissible for *A*; hence knowledge of entailments implies knowledge of inadmissibilities by the objective definition making that definition conditionally informative to a limited extent.

But does a definition have to be informative for it to be of philosophical use? As there are other examples of uninformative definitions enjoying prominent positions in philosophy, the answer appears to be "No.". An example of such is the Platonic definition of knowledge as true, justified, belief. Famously, one cannot use this definition to identify what is known about the external world, for truth transcends any evidence one can have about the external world. I.e., there is no evidence for *P*, where *P* is about the external world, possessible by *X* such that *P* cannot be false. Consequently, Plato's definition of knowledge–which is often presented as a schema–is largely uninformative. Despite this shortcoming, philosopher's have found Plato's definition of knowledge to be illuminating even when applied to knowledge of the external world.

So it seems that whilst informativity is a virtue definitions should aspire to, uninformative definitions still have their uses in philosophy; particularly in the clarification of meaning. It is in this spirit that I offer the Objective definition schema for relative admissibility and its degrees. While it is admitted that this schema is largely useless at settling disputes over the extension of the admissibility predicate, it is hoped that the increase in our understanding of Lewis' admissibility gleaned from this definition schema is sufficient justification for its endorsement.

References

J. Bigelow, J. Colins and R. Pargetter, "The Big Bad Bug: What are Humean Chances". *Brit. J. Phil. Sci.* 44, 443-462, 1993.

N. Hall, "Correcting the Guide to Objective Chance". Mind 103(412), 505-517, 1994.
--- "Two Mistakes about Credence and Chance". *Australian Journal of Philosophy*, 82(1), 93-111. 2004.

J. Halpin, "Lewis, Thau, and Hall on Chance and the Best-System Account of Law". *Philosophy of Science* 65(2), 349-360, 1998.

C. Hoefer, "On Lewis' Objective Chance: Humean Supervenience Debugged". *Mind* 106(422), 321-334, 1997.
--- "The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance". *Mind* 116(463), 549-596, 2007.

D. Lewis, "A Subjectivist's Guide to Objective Chance". In: *Philosophical Papers Volume II*. New York, Oxford University Press, 83-133, 1986. Originally published in:

--- “Humean Supervenience Debugged”. *Mind* 103(412), 473-490, 1994.

R. Jeffrey (Ed.). *Studies in Inductive Logic and Probability Volume II*, Berkley, University of California Press, 1980.

B. Loewer, “Determinism and Chance”. *Studies in History and Philosophy of Science Part B* 32(4), 609{620 (2001).
--- “David Lewis’ Humean Theory of Objective Chance”. *Philosophy of Science* 71, 1115-1125, 2004.

G. Masterton, *Objective chance: A study in the Lewisian tradition.* Uppsala University Press. 2010.

C. J. G. Meacham, “Two Mistakes Regarding the Principal Principle”. *British Journal of Philosophy of Science* 61, 407-431, 2010.

J. Pearl, *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press, 2000.

B. Skyrms, “Propensities and Causal Necessity”. *The Journal of Philosophy* 74(11), 704-713, 1977.

B. Skyrms, “Conditional Chance”. In: J. Fetzer (Ed.). *Probabilistic Causation: Essays in honor of Wesley C. Salmon.* Dordrecht, 1988.

M. Thau, “Undermining and Admissibility”. *Mind* 103(412), 491-503. 1994.

B. van Fraassen, *Laws and Symmetry*. Oxford, Oxford University Press, 1989.

J. Williamson, “Philosophies of Probability”. In: *Handbook of the Philosophy of Mathematics*. Draft of July 21, 2008. To be published in: A. Irvine (ed.). *Handbook of the Philosophy of Mathematics, Volume 4 of the Handbook of the Philosophy of Science*, Elsevier. 2008.

Truth and Historicism in Kuhn's Thesis of Methodological Incommensurability

Marco Marletta
(University of Palermo)
marcom89@libero.it

Methodological incommensurability and incomparability

The thesis of incommensurability is a much discussed subject in Kuhn's philosophy of science, which, since it has been proposed by Thomas Kuhn and Paul Feyerabend in 1962, has given rise to very different interpretations. This is partially due to the multidimensional nature of the concept of "incommensurability" and sometimes to the lack of clarity of Kuhn himself. In *The Structure of Scientific Revolutions* he distinguishes three aspects of incommensurability, each of which could easily appear independent from the others.

To sum up, Kuhn says that the main features of incommensurability are as follows: a) first, the proponents of paradigms do not agree about methods, standards and aims of science[1]; b) second, and accordingly to the holistic nature of theory change, although the new paradigms holds many concepts of the old theory "within the new paradigm, old terms, concepts and experiments fall into new relationship one with the other"[2]; c) finally, the third aspect of incommensurability is that "the proponents of competing paradigms practice their trades in different worlds"[3]. We can call these aspects of incommensurability a) methodological, b) semantic, c) ontological[4]. In this

[1] Kuhn, 1970a, 148.
[2] Ibid., 149.
[3] Ibid., 150.
[4] Buzzoni, 1986, 111, partially Hoyningen-Huene, Sankey 2001b, ix.

paper I will focus especially on methodological aspect and on his relationship with semantic incommensurability[5].

Methodological incommensurability is a specifically Kuhnian theme. Though also Feyerabend is an opponent of scientific method's monism, he never talks about incommensurability in this context: he has always restricted incommensurability to its semantic dimension. Instead, in *The Structure of Scientific Revolutions*, Kuhn affirms that "the proponents of competing paradigms will often disagree about the list of problems that any candidate for paradigm must resolve. Their standards or their definitions of science are not the same."[6] According to this thesis there are not shared, objective methodological rules or neutral scientific standards for theory comparison and choice; and that is because every paradigm determines its own standards of evaluation and scientific propriety[7]. Incommensurability is due to the lack of external standards which do not depend on the paradigms themselves and can reduce theory choice to a neutral mechanical algorithm. In sum, two paradigms are incommensurable from a methodological point of view because: a) they focus on different problematic fields; b) they disagree on the priority to be given to these problems in the context of their research program; c) they define in different ways the most basic problems, which reflect the pragmatic, the research strategies and the specific interests of the same paradigm[8].

---

[5] Despite to Kuhn's exposition, Hoyningen-Huene and Sankey (2001b) distinguish only two aspects of incommensurability: methodological and semantic. And indeed it is probably right, since the third aspect, the "world changes" thesis or "ontological relativism" (Sankey, 1997) is a complex position which involves not only the thesis of incommensurability, but also the structure of paradigms and the refutation of the correspondence theory of truth (Bird, 2011).

[6] Kuhn, 1970a, 148.

[7] "To the extent, as significant as it is incomplete, that two scientific schools disagree about what is a problem and what a solution, they will inevitably talk trough each other when debating the relative merits of their respective paradigms. In the partially circular arguments that regularly result, each paradigm will be shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponent. There are other reasons, too, for the incompleteness of logical contact that consistently characterizes paradigm debates. For example, since no paradigm ever solves all the problems it defines and since no two paradigms leave all the same problems unsolved, paradigm debates only involve the question: Which problems is it more significant to have solved? Like the issue of competing standards, that question of values can be answered only in terms of criteria that lie outside normal science altogether, and it is that recourse to external criteria that most obviously makes paradigm debates revolutionary." (Kuhn, 1970a, 109-110)

[8] See Doppelt, 1978/1983, 121.

Many critics have interpreted this claim as something like radical incomparability between rival scientific theories[9]. Methodological incommensurability has been regarded as a source of epistemological relativism about theory comparison: if theories are incommensurable (or, according to this interpretation, incomparable), scientific changes are fundamentally irrational, since they cannot be explained by means of rational procedures. Scientific revolutions would merely be "conversions"[10]. But such an interpretation has been strongly refuted by Kuhn himself: he explicitly says that incommensurability does not imply incomparability[11].

> Remember briefly where the term 'incommensurability' came from. The hypotenuse of an isosceles right triangle is incommensurable with its side or the circumference of a circle with its radius in the sense that there is no unit of length contained without residue an integral number of times in each member of the pair. There is thus no common measure. But lack of a common measure does not make comparison impossible. On the contrary, incommensurable magnitudes can be compared to any required degree of approximation.[12]

In responding to his critics, Kuhn affirms that is aim was not to make theory choice an irrational process. He would only saying that, although theory choice is generally rational, it is not mechanical and regulated by only one scientific method; as he has written in *The Structure of Scientific Revolutions* "there is no neutral algorithm for theory choice, no systematic decision procedure which, properly applied, must lead each individual in the group to the same decision"[13]. The evaluation of scientific theories is necessary a practical process, which involves decisional, deliberative and subjective elements. Kuhn does not want to say that scientists do not use logic and experience[14]; but rather that logic and experience are not able to force theory choice; the evaluation of a scientific theory is very different than a mathematical proof:

> In a debate over choice of theory, neither party has access to an argument which resembles a proof in logic or formal mathematics. In the latter, both

[9] See among the others Lakatos, 1970, 179 n. 1; Newton-Smith, 1981, 9-10; Putnam, 1981, 118, Scheffler, 1967, 16-17; Shapere, 1966, 67-68.
[10] The term "conversion" is used by Kuhn sixteen times in Kuhn, 1970a, 144-159.
[11] Among the critics who have denied the identification between incommensurability and incomparability see Bernstein, 1983, 82, and Hoyningen-Huene, 1989/1993,. 218-221.
[12] Kuhn, 1983/2000, 35. See also Kuhn, 1970c/2000, 163; Kuhn, 1976b/2000, 189; Kuhn, 1979/2000, 204.
[13] Kuhn, 1970a, 200.
[14] Kuhn, 1970c/2000, 156.

> premises and rules of inference are stipulated in advance. If there is disagreement about conclusions, the parties to the debate can retrace their steps one by one, checking each against prior stipulation. At the end of that process, one or the other must concede that at an isolable point in the argument he has made a mistake, violated or misapplied a previously accepted rule. After that concession he has no recourse, and his opponent's proof is then compelling. Only if the two discover instead that they differ about the meaning or applicability of a stipulated rule, that their prior agreement does not provide a sufficient basis for proof, does the ensuing debate resemble what inevitably occurs in science.[15]

To replace the scientific standards based model for theory comparison, in the seventies Kuhn has provided a value based model[16]. He lists several values used by scientific communities[17]: a) accuracy (of the factual statements, both from a quantitative and qualitative point of view); b) consistency (absence of internal contradictions); c) scope (the domain of possible application); d) simplicity (the ability to unify apparently different group of phenomena); e) fruitfulness (the ability to predict and to apply to new phenomena). Scientists do not considered these values rules which determine choice, but rather "values, which influence it"[18]; moreover they can be interpreted in different ways and, in some situation, they can conflict with one other.

Without going further into the problem of Kuhn's theory of scientific method and his adequacy[19], we are probably faced with a reason which forced Kuhn, in his latest work, to break down the problem of incommensurability and the problem of scientific method in theory comparison. Indeed, defending his philosophy from the accusation of relativism, he said that

> Nothing […] implies either that there are no good reasons for being persuaded or that those reasons are not ultimately decisive for the group. Nor does it even imply that the reasons for choice are different from those usually listed by philosophers of science: accuracy simplicity, fruitfulness, and the like. What it should suggest, however, is that such reasons function as values and that they can thus be differently applied, individually and collectively, by men who concur in honoring them.[20]

[15] Ibid.
[16] Kuhn, 1977b/1977a.
[17] Ibid., 321-322.
[18] Ibid., 331.
[19] See Nola, Sankey, 2000b, 26-30.
[20] Kuhn, 1970b/1970a, 199.

As it has been remarked by Siegel[21], this argumentation for incommensurability already does not involve incommensurability, but only a theory of value based theory choice in scientific practice. Also Bird says that, in the kind of semantic incommensurability developed in his works of the eighteens, the question of relativism or absolutism about theory comparison criteria is simply not being asked[22]. It appears, at first sight, that Kuhn, by means of his discussion on scientific values, merely drops out of the problem of methodological incommensurability and relegates incommensurability to his semantic aspect. Kuhn himself seems to confirm this interpretation where he says that

> Both Feyerabend and I wrote of the impossibility of defining the terms of one theory on the basis of the terms of the other. But he restricted incommensurability to language; I spoke also of differences in "methods, problem-field, and standards of solution", something I would no longer do except to the considerable extent that the latter differences are necessary consequences of the language-learning process.[23]

Kuhn makes methodological incommensurability dependent from semantic incommensurability. But this assertion does not imply that methodological incommensurability is dissolved; rather, we have to look for the foundation of this kind of incommensurability in the semantic dimension of incommensurability itself. For this I will divide Kuhn's thesis of methodological incommensurability in two sub-theses:

1) *There is not a scientific method which constraints theory choice and assures his correctness: theory choice is a deliberative process*. We have just discussed this thesis; it does not necessary imply neither relativistic consequences nor incommensurability. Moreover it is not a particularly original or revolutionary thesis. Also Karl Popper and many others philosophers of science have said that scientific method cannot force scientist's choices and that theory choice entails practical decisions[24].

2) *Incommensurability does not mean incomparability*: *we can compare scientific theories' accuracy, fruitfulness, scope, consistency, simplicity. But we cannot compare them to discover which theory is closer to truth.* While the first sub-thesis has been shelved in the development of Kuhn's work, this second thesis constitutes the linkage between methodological and semantic

[21] Siegel, 1987, 57.
[22] Bird, 2000, 240-241.
[23] Kuhn, 1983/2000, 34 fn. 2.
[24] Popper, 1959, 61.

incommensurability and it has been supported by Kuhn in his whole scientific life. I will explain the reasons of this linkage in the next section.

Methodological incommensurability, truth, historicism

Discussing the critics on epistemological relativism, Kuhn himself relates methodological incommensurability with his critique of truth as the aim of science[25]. Referring to the above analysis of the role of proof in theory choice, he compares mathematical proof and truth, since they both suppose inter-theoretical applications, i.e. the applications in which incommensurability plays a role[26]. Proof and truth are meaningful concepts only in a shared practical context, which constitutes the basis of the agreement between scientists about the empirical assertions of a theory confirmed by experiments and then regarded as true (or false, or not tested). But, when we try to extend the use of terms like 'proof' and 'truth' above the intra-theoretical context, Kuhn affirms that "dealing with the comparison of theories designed to cover the same range of natural phenomena, I am more cautious"[27].

Incommensurability blocks the possibility of a neutral comparison between scientific theories. This statement does not mean that paradigms are incomparable, because we can always compare their accuracy, consistency and so on; instead, paradigms are incomparable referring to the evaluation of their respective likeness to truth. In his evolutionary account of the development of science, truth has no place[28]. At least, incommensurability, also in his methodological feature, does not involve relativism about the rationality of theory choice, but rather it is a form of relativism about truth. Kuhn has always refuted the accusations of irrationalism, but, about truth, he says that he can rightly be called a relativist: "one scientific theory is not as

[25] For Kuhn's critique of the idea of truth and especially of the theory of truth as correspondence, see Bird, 2000, 209-266; and Kuakkunen, 2007.
[26] Kuhn, 1970c/2000, 162.
[27] Ibid., 160.
[28] "It is now time to notice that until the last very few pages the term 'truth' had entered this essay only in a quotation from Francis Bacon. And even in those pages it entered only as a source for the scientist's conviction that incompatible rules for doing science cannot coexist except during revolutions when the profession's main task is to eliminate all sets but one. The developmental process described in this essay has been a process of evolution form primitive beginnings – a process whose successive stages are characterized by an increasingly detailed and refined understanding of nature. But nothing that has been or will be said makes it a process of evolution toward anything." (Kuhn, 1970a, 170-171).

good as another for doing what scientists normally do. In that sense I am not a relativist. But there are reasons why I get called one, and they relate to the contexts in which I am wary about applying the label 'truth'"[29]. Then methodological incommensurability does not imply that all the theories are equally good, but that all the theories are equally close (or far) to the truth.

Kuhn returns more explicitly and deeply on this argument in his latest works: the evaluation of change of belief is now embedded in the evolutionary dimension of scientific knowledge[30]. This evolutionary account does not try to explain the rationality and the correctness of our convictions, but rather the change of convictions itself. The non evolutionary point of view's aim is to evaluate scientific theories isolated, in order to calculate their truth or probability, where truth means "something like corresponding to the real, the mind-independent external world"[31]. But, Kuhn adds that

> Sticking therefore with the formulation that assumes truth to be the goal of evaluations, notice that it requires evaluation to be indirect. Seldom or never can one compare a newly proposed law or theory directly with reality. rather, for purposes of evaluation, one must embed it in a relevant body of currently accepted beliefs-for example, those governing the instruments with which the relevant observations have been made-and then apply to the whole a set of secondary criteria. Accuracy is one of these, consistency with other accepted beliefs is another, breadth of applicability a third, simplicity a fourth, and there are others besides. All these criteria are equivocal, and they are rarely all satisfied at once.[32]

Kuhn reiterates that the verification of truth and the validity of proof is not an inter-theoretical function; a theory cannot be tested by means of a direct clash with reality. Moreover scientific values are meaningless if they are not placed in the context of scientific community's shared practice. In such a context the application of scientific values is more fruitful, although it cannot serve to eliminate disagreement at all. The evaluation of the change of convictions is more ductile "especially since what must be compared are only sets of beliefs actually in place in the historical situation"[33]. As we have

[29] Kuhn, 1970c/2000, 160.
[30] "For the philosopher who adopts the historical perspective, the problem is the same: understanding small incremental changes of belief. When questions about rationality, objectivity, or evidence arise in that context, they are addressed not to the beliefs that were current either before or after the change, but simply to the change itself." (Kuhn, 1992/2000, 112)
[31] Ibid., 114.
[32] Ibid.
[33] Ibid., 115.

previously said, Kuhn admits the possibility of the evaluation of theory referring to scientific values: a paradigm can be more accurate, more consistent, can have a broader field of application and can be simpler than his rivals "without for those reasons being any truer"[34]. A clash between two rival theories is conceivable and it could be productive in an evolutionary perspective; but a direct clash between theory and reality, in a classical perspective, is just not an option. Theory evaluation is an historical process which can only be realized by a comparative point of view. And, as Kuhn himself says, incommensurability is "an essential component of any historical, developmental, or evolutionary view of scientific knowledge"[35].

According to our interpretation, a connection between methodological incommensurability, truth and history of science is emerging. This connection will become clear returning to Kuhn's early works. In *The Structure of Scientific Revolutions*, Kuhn introduces his first extended description of methodological incommensurability by means of a statement about the historical and evolutionary conception of science.

> *Paradigms differ in more than substance, for they are directed not only to nature but also back upon the science that produced them*. They are the source of the methods, problem-field, and standards of solution accepted by any mature scientific community at any given time. As a result, the reception of a new paradigm often necessitates a redefinition of the corresponding science. Some old problems may be relegated to another science or declared entirely "unscientific". Others that were previously non-existent or trivial may, with a new paradigm, become the very archetypes of significant scientific achievement. And as the problems change, so, often, does the standard that distinguishes a real scientific solution from a mere metaphysical speculation, word game, or mathematical play. The normal-scientific tradition that emerges from a scientific revolution is not only incompatible but often actually incommensurable with that which has gone before. (Italics mine)[36]

In this passage Kuhn describes the alteration of scientific standards, problem fields, and scientific aims after scientific revolutions. Merely, he resumes the features of methodological incommensurability that we have presented in the first section. But, in addiction to that, he relates methodological incommensurability to a consideration about the historical structure of paradigms: "they are directed not only to nature but also back upon the science that produced them". According to Kuhn, paradigms have a double directionality. From one hand they are connected to nature and, from

[34] Ibid.
[35] Kuhn, 1991/2000, 91.
[36] Kuhn, 1970a, 103.

the other hand, to their historical tradition and past science. This assertion summarizes Kuhn's historicism. He does not want to say only that every scientific paradigm is relative to the historical and social context in which it develops; rather the historical structure of paradigms is inextricably linked to the knowledge of nature embodied in the paradigms themselves.

In fact, as we have just seen, the confrontation between paradigm and nature is not immediate: a direct contact between theories and reality cannot exist. However Kuhn does not affirm simply that the contact between paradigm and nature is mediated by the paradigm itself. If this were the case, Kuhn would only say that observation is theory laden, which is an achievement accepted by nearly all of the philosophers of science. Instead Kuhn's claim is more radical. He states not only that the relationship between paradigm and nature is mediated by the paradigm itself, but that it is mediated also by the relationship between the current and the past paradigm. Anyway the relation between historically successive paradigm is incommensurability, and exactly semantic incommensurability, since every paradigm inherits his lexicon by the science which come first it. Roughly, incommensurability influences the connection between paradigms and nature. In fact, as we have seen, if two theories are incommensurable, we cannot determine which one is closer to truth. Summarizing, the historical nature of paradigms (their constitutive relation with paradigms which produce them) plays a fundamental role in the determination of the relationship between paradigm and world, which consequently cannot be a direct clash, but always a comparative evaluation between two theories. That is because the historical relation between current and past paradigm is expressed by incommensurability, which denies the possibility of an evaluation of the likeness to truth of a single theory. Incommensurability, truth and historicism (i.e. the evolutionary model of scientific progress) create a circle in which every element implies the others.

In Kuhn's philosophy of science, both incommensurability and truth are historical concepts. To be more exact, the fact that incommensurability is an historical concept does not mean that it is a concept gathered from the analysis of the history of science[37]. Kuhn tells us that it was not by reflecting on the history of science that he first thought about incommensurability, but on his very activity as an historian of science[38]. The historian experiences

[37] For Kuhn's conception of history of science see Hoyningen-Huene, 1989/1993, 3-27.
[38] "Feeling that way, I continued to puzzle over the text, and my suspicions ultimately proved well-founded. I was sitting at my desk with the text of Aristotle's Physics open

incommensurability when he is studying an ancient scientific text and he notices apparently nonsensical passages. While many researchers have considered these passages as signs of antique mistakes, Kuhn believes that they are the results of the incommensurability between successive paradigms. Kuhn denounces the impossibility of an Archimedean, external point of view from which we understand history of science as a cumulative development: "for the historian, in short, no Archimedean platform is available for the pursuit of science other than the historically situated one already in place"[39].

The connection between this kind of historiography in which incommensurability plays a constitutive role and the truth relativistic conception of methodological incommensurability is immediately observed by Kuhn: "though both rationality and relativism are somehow implicated, what is fundamentally at stake is rather the correspondence theory of truth"[40]. As we have seen regards to truth and proof, also the concept of an external Archimedean point of view on history of science presupposes inter-theoretical applications. But, again, knowledge cannot be evaluated in isolation, but only in a shared practical context: another time, only the change of belief can be justified, while all single theories are equally distant to truth:

> On the developmental view, scientific knowledge claims are necessarily evaluated from a moving, historically situated, Archimedean platform. What requires evaluation cannot be an individual proposition embodying a knowledge claim in isolation: embracing a new knowledge claim typically requires adjustment of other beliefs as well. Nor is it the entire body of knowledge claims

---

in front of me and with a four-colored pencil in my hand. Looking up, I gazed abstractedly out the window of my room-the visual image is one I still retain. Suddenly the fragments in my head sorted themselves out in a new way, and fell into place together. My jaw dropped, for all at once Aristotle seemed a very good physicist indeed, but of a sort I'd never dreamed possible. Now I could understand why he had said what he'd said, and what his authority had been. Statements that had previously seemed egregious mistakes now seemed at worst near misses within a powerful and generally successful tradition. That sort of experience – the pieces suddenly sorting themselves out and coming together in a new way-is the first general characteristic of revolutionary change that I shall be singling out after further consideration of examples. Though scientific revolutions leave much piecemeal mopping up to do, the central change cannot be experienced piecemeal, one step at a time. Instead, it involves some relatively sudden and unstructured transformation in which some part of the flux of experience sorts itself out differently and displays patterns that were not visible before." (Kuhn 1981/2000, pp. 16-17). See also Kuhn, 1989/2000, 59 fn. 1.

[39] Kuhn, 1991/2000, 95.

[40] Ibid.

> that would result if that proposition were accepted. Rather, what's to be evaluated is the desirability of a particular change-of-belief, a change which would alter the existing body of knowledge claims so as to incorporate, with minimum disruption, the new claim as well. Judgments of this sort are necessarily comparative: which of two bodies of knowledge-the original or the proposed alternative-is better for doing whatever it is that scientists do.[41]

Better, in the process of evaluation an external point of view seems to exist. But it is only a temporally, historical situated pseudo-Archimedean point of view: it is constituted by the same agreement of scientific community on the paradigm itself[42] (i.e. also on the scientific values previously presented): "the historical perspective, thus, also invokes an Archimedean platform, but it is not fixed. Rather, it moves with time and changes with community and sub-community, with culture and subculture"[43]. The conditions of theory comparison are paradigm-dependent. The traditional non evolutionary philosophy of science fails because it designates neutral language and observation as judge of scientific theories' likeness to truth[44]; or equally, as the Archimedean platform for theory choice. Instead Kuhn's opinion is that every evaluation is relative to a scientific community and his shared lexicon:

> From the historical perspective, however, where change of belief is what's at issue, the *rationality* of the conclusions requires only that the observations invoked be neutral for, or shared by, the members of the group making the decision, and for them only at the time the decision is being made. By the same token, the observations involved need no longer be independent of all prior beliefs, but only of those that would be modified as a result of the change. The

[41] Ibid., 95–96.
[42] Ibid., 96.
[43] Kuhn,1992/2000, 113.
[44] "The semantic conception of truth is regularly epitomized in the example: 'Snow is white' is true if and only if snow is white. To apply that conception in the comparison of two theories, one must therefore suppose that their proponents agree about technical equivalents of such matters of fact as whether snow is white. If that supposition were exclusively about objective observation of nature, it would present no insuperable problems, but it involves as well the assumption that the objective observers in question understand 'snow is white' in the same way, a matter which may not be obvious if the sentence reads 'elements combine in constant proportion by weight'. Sir Karl takes it for granted that the proponents of competing theories do share a neutral language adequate to the comparison of such observation reports. I am about to argue that they do not. If I am right, then 'truth' may, like 'proof', be a term with only intra-theoretical applications. Until this problem of a neutral observation language is resolved, confusion will only be perpetuated by those who point out (as Watkins does when responding to my closely parallel remarks about 'mistakes') that the term is regularly used as though the transfer from intra- to inter-theoretical contexts made no difference." (Kuhn, 1970c/2000, 161-162)

> very large body of beliefs unaffected by the change provides a basis on which discussion of the desirability of change can rest. It is simply irrelevant that some or all of those beliefs may be set aside at some future time. To provide a basis for rational discussion they, like the observations the discussion invokes, need only be shared by the discussants. There is no higher criterion of the rationality of discussion than that.[45]

Then, like proof, truth can be only an intra-theoretical concept and consequently an historical concept: truth is not correspondence with a mind-independent reality, but only the result of a rational evaluative process. The product of a successful theory comparison is internal to the historical situation which enables the evaluation itself: the problem of the truth or falsity (intended as a relation between a language and something external to it) simply is not the question being asked: "justification does not aim at a goal external to the historical situation but simply, in that situation, at improving the tools available for the job at hand"[46]. Or, referring to the lack of an Archimedean point of view: "Only a fixed, rigid Archimedean platform could supply a base from which to *measure* the distance between current belief and true belief. In the absence of that platform, it's hard to imagine what such a *measurement* would be, what the phrase 'closer to the truth' can mean" (italics mine)[47].

In this last passage I have stressed the words "measure" and "measurement" because they are strictly related to incommensurability. As Kuhn has repeated several times, incommensurability is a mathematical term which means "no common measure". But outside of its original context, its function is metaphorical: "the phrase 'no common measure' becomes 'no common language'. The claim that two theories are incommensurable is then the claim that there is no language, neutral or otherwise, into which both theories, conceived as sets of sentences, can be translated without residue or loss"[48]; obviously we must specify that the lack of common measure does not imply incomparability. But the measure metaphor does not stop there. As well as denouncing the absence of a common measure to explain inter-theoretical relations, Kuhn compares paradigms just to units of measurements or, better, to metric or coordinate systems[49].

---

[45] Kuhn, 1992/2000, 113.
[46] Kuhn, 1991/2000, 96.
[47] Kuhn, 1992/2000, 115.
[48] Kuhn, 1983/2000, 36.
[49] "Two people may use a set of interrelated terms in the same way but employ different sets (in principle, totally disjunct sets) of field coordinates in doing so. Examples will be found in the next section of this paper; meanwhile the following metaphor may prove suggestive. The United States can be mapped in many different

A metric system is a condition for the possibility, or a formal matrix, of justification and truth-value attribution and discussion in the domain of the system itself. Probably, in this conception Kuhn is debtor of Wittgenstein's discussion about the standard meter[50]. Wittgenstein says that if we want to know if is it true or false that something is a meter long, we can (ideally) compare this object with the standard meter in Paris. "The table is one meter long" is an empirical proposition verifiable or falsifiable relatively to the metric system of measurement. But a question such as "Is the standard meter in Paris a meter long?" is meaningless referring to the same system; the proposition "the standard meter in Paris is a meter long" is not an empirical proposition, but a grammatical proposition and consequently it is neither true nor false[51].

Kuhn's description of paradigms is very similar to this: truth, proof and justification are meaningful only in an intra-theoretical context, while it is impossible to evaluate the likeness to truth of a paradigm[52]. Every shared paradigm is a system of measurement which enables theory evaluation and justification by means of common scientific values such as accuracy,

---

coordinate systems. Individuals with different maps will specify the location of, say, Chicago by means of a different pair of coordinates. But all will nevertheless locate the same city provided that the maps are scaled to preserve the relative distances between the items mapped. The metric that accompanies each of the various sets of coordinates must, that is, be chosen to preserve the structural geometrical relations within the mapped area." (Kuhn, 1989/2000, 63).

[50] Kuhn's paradigm are just been compared with Wittgenstein's standard meter and color samples (Baltas, 2004, Malone, 1993, but also Glock, 1996). For a discussion of the relevance of standard meter in Wittgenstein's philosophy see Baker and Hacker, 2005, 189-199).

[51] "There is one thing of which one can say neither that it is one meter long, nor that it is not one meter long, and that is the standard meter in Paris.–But this is, of course, not to ascribe any extraordinary property to it, but only to mark its peculiar role in the language-game of measuring with a meter-rule.–Let us imagine samples of color being preserved in Paris like the standard meter. We define: "sepia" means the color of the standard sepia which is there kept hermetically sealed. Then it will make no sense to say of this sample either that it is of this color or that it is not." (Wittgenstein, 1958, § 50, 25).

[52] "A lexicon or lexical structure is the long – term product of tribal experience in the natural and social world, but its logical status, like that of world meaning in general, is that of convention. Each lexicon makes possible a corresponding form of life within which the truth or falsity of propositions may be both claimed and rationally justified, but the justification of lexicon or of lexical change can only be pragmatic. With the Aristotelian lexicon in place it does make sense to speak of the truth or falsity of the Aristotelian assertion in which terms like 'force' or 'void' play an essential role, but the truth values arrived at need have no bearing on the truth or falsity of apparently similar assertions made with the Newtonian lexicon." (Kuhn, 1993/2000, 244)

consistency and so on. Thanks to these values we can compare the respective merits of two rival theories in relation to their respective methods, standards, aims: the meter of comparison is not an absolute Archimedean platform, but the same scientific practice and the concrete historical situation. But, according to Kuhn, traditional epistemology just looks for an objective meter to evaluate isolated scientific theories' truth or probability. Kuhn denounces the impossibility of such an inter-theoretical meter: since every theory is a metric system which enables truth-value attributions, in order to attribute a truth-value to the metric system itself, we need for a meta-metric system (i.e. an Archimedean platform) able to map the different paradigms more or less close to truth. Kuhn refers to this meta-metric system by different expressions: an Archimedean platform, a common measure, a neutral observational language, truth, the world-in-itself. Every one of these concepts, attributed by Kuhn to the traditional non evolutionary epistemology, supposes the possibility of a non historical evaluation of theories: a direct clash between theories and reality which Kuhn considers absolutely impossible.

It remains that Kuhn, in his works of the eighties and nineties, puts the methodological thesis of incommensurability aside to examine in depth its semantic implications. The reason can now become clearer. We have seen that the discussion about the justification of conviction change can be meaningful only in an evolutionary perspective which does not aim to overstep the historical situation. The discussion about theory choice comes true in the light of a horizon of agreement within the scientific community; in other words a provisional Archimedean platform, i.e. a shared paradigm, or lexicon or language: "'no common measure' becomes 'no common language'"[53]. Only a neutral lexicon in which the statements of every theory are translatable could constitute a direct access to reality and a source of inter-theoretical truth evaluation. The transition from methodological to semantic incommensurability is due to Kuhn's analysis of the origin of the agreement within scientific communities about paradigm. Shortly, Kuhn discovers the roots of such an agreement in the constitutive role played by paradigm learning in scientific practice[54].

The applicability of scientific values in theory choice, although in a non inter-theoretical sense, presupposes a shared perspective enabled by scientific training: "I spoke also of differences in "methods, problem-field, and

[53] Kuhn, 1983/2000, 36.
[54] Kuhn, 1974/1977a.

standards of solution", something I would no longer do except to the considerable extent that the latter differences are necessary consequences of the language-learning process"[55]. Though I cannot analyze here the constitutive nature of learning process in science, I want only to remark that this is also a Wittgensteinian theme. Kuhn says that, like Wittgenstein's standard meter, a paradigm cannot be justified recurring to reality. The foundation of paradigms (or grammar) lies in scientific (or linguistic) practice itself, institutionalized by scientific (or simply linguistic) training: "How do I know that this color is red? – It would be answer to say: 'I have learnt English'"[56]. The priority of scientific learning process finds the foundation of incommensurability of standards, methods and problem-fields in the semantic question of the dependence of meaning (and then of meaning change) from scientific practice and uses[57]: "kind terms [the constituents of the structure of a lexicon] are learned in use: someone already adept in their use provides the learner with examples of their proper application"[58].

Anyway, the pragmatic (un)foundation of paradigms repurposes the main consequence of methodological incommensurability thesis: since a direct clash between a theory and reality is impossible, all the theories are equally close to truth.

Conclusions: two problems on falsification

We have seen that methodological incommensurability concerns with the impossibility of a direct access to reality (a meta metric system or Archimedean platform) which enables us to map theories in a range of increasing likeness to truth. Truth is a meta-meter which plays no role in

[55] Kuhn, 1983/2000, 34 fn. 2. Or equally, "my original discussion described nonlinguistic as well as linguistic forms of incommensurability. That I now take to have been an overextension resulting from my failure to recognize how large a part of the apparently non linguistic component was acquired with language during the learning process" (Kuhn, 1989/2000, 60 fn. 4).
[56] Wittgenstein, 1958, § 381, 176.
[57] With regards to Wittgenstein, about the constitutive nature of learning process in relation with the structure of grammar see Williams, 1999, in particular 58-59 and 206 and ff. This is a good exposition also referring to Kuhn's conception of scientific training and, again in accordance with Kuhn's philosophy of science, stresses the social nature of learning.
[58] Kuhn, 1993/2000, 230.

Kuhn's philosophy of science[59]. This observation about the always indirect relationship between theories and reality can help us to solve two problems regarding Kuhn's interpretation of falsificationism. Confirming the connection between these problems and incommensurability, Kuhn exposes them just before introducing the three aspects of incommensurability (methodological, semantic, and ontological) quoted at the beginning.

In *The Structure of Scientific Revolutions*, Kuhn says that verification and falsification are after all equivalent. Falsification cannot be identified with anomalous experience, but rather it is "a subsequent and separate process that might be equally called verification since it consists in the triumph of a new paradigm over the old one"[60]. Especially because this criticism is referred explicitly to Popper, it could sound bad because Popper has always stressed the asymmetry between verification and falsification. This apparent misunderstood is due to the fact that this criticism to the falsificationist method has been interpreted simply as a refutation of the concept of neutral observation: since observation cannot be the final and irrevocable judge of theories, scientists are not forced to abandon a theory after a falsification. This is true and supported by Kuhn, but also by Popper who have often reaffirmed the inexistence of neutral observations and ultimate falsifications. Then the source of the disagreement between Kuhn and Popper must be another. This source is the idea expressed by Popper that we can test theories by a match with reality. Kuhn criticizes Popper not only by a technical insight about the difficulties of falsificationism, but also from a more general epistemological point of view. For Kuhn verification and falsification are equivalent because they both presuppose the possibility of a direct clash between scientific language and reality.

[59] Hoyningen-Huene remarks the connection between the refutation of the theory of truth as correspondence and the impracticality of a direct access to reality. He demonstrates that the main argument presented by Kuhn against the correspondence theory (Kuhn, 1970b, 206) is rather an epistemological argument which "proceeds from the assumption that it's essentially meaningless to talk of what there really is, beyond (or outside) of all theories. If this insight is correct, it's impossible to see how talk of a 'match' between theories and absolute, or theory – free, purely object – sided reality could have any discernible meaning. How could the (qualitative) assertion of a match, or the (comparative) assertion of a better match, be assessed? The two pieces asserted to match each other more or less would have to be accessible independently of one another, when one of the pieces is absolute reality. But if we had access to absolute reality – and here we can only return to our initial premise – what interest would we have in theories about it?" (Hoyningen-Huene, 1989/1993, 263-264).

[60] Kuhn, 1970b, 147.

Again in *The Structure of Scientific Revolutions*, Kuhn affirms that the question of the agreement between theory and reality becomes meaningful only in a comparative perspective. While Kuhn has rejected the problem of theories' verisimilitude, "questions much like that can be asked when theories are taken collectively or even in pairs"[61]. This assertion has often been interpreted in connection with another previous thesis: since no theories is completely successful in his problem- field, if anomalies were falsifications, we would reject all theories at all time[62]. This connection seems to mean that scientists' dogmatism up against falsification is reasonable because it will be damaging for science if we drop out of our best theory without a better alternative[63]. But, again, Kuhn's critique is also more general. Kuhn says that theory comparison can only be a theory-theory match and not a theory-reality match because the latter kind of comparison is, in principle, impossible. We have just seen the historical and evolutionary reasons which have led Kuhn to such an intra-theoretical conception of truth. Anyway, the affinity between the always indirect match between paradigm and reality and incommensurability is now reaffirmed: in fact, after these considerations about theory comparison, Kuhn introduced the most detailed analysis of incommensurability. Again, methodological incommensurability is not a relativistic and irrationalist danger for theory comparison: we can, more or less easily, establish which theory is more accurate, consistent, simply and so on; but, without a "common measure", we cannot decide which theory is closer to truth.

[61] Ibid.
[62] Ibid., 146.
[63] Also this traditional interpretation is surely right and Kuhn supports them explicitly: "once it has achieved the status of paradigm, a scientific theory is declared invalid only if an alternate candidate is available to take its place. No process yet disclosed by the historical study of scientific development at all resembles the methodological stereotype of falsification by direct comparison with nature. That remark does not mean that scientists do not reject scientific theories, or that experience and experiment are not essential to the process in which they do so. But it does mean –what will ultimately be a central point - that the act of judgment that leads scientists to reject a previously accepted theory is always based upon more than a comparison of that theory with the based upon more a comparison of that theory with the world. The decision to reject one paradigm is always simultaneously the decision to accept another, and the judgment leading to that decision involves the comparison of both paradigms with nature and with each other." (Ibid., 77).

Bibliography

Baltas, A., (2004) "On the Grammar Aspects of Radical Scientific Discovery", Philosophia Scientiae 8, pp. 169 – 201.

Baker, G.P., Hacker, P.M.S., (2005), *Wittgenstein: Understanding and Meaning*. Part I: Essays. Volume 1 of An Analytical Commentary on the *Philosophical Investigations*. Second extensively revisited edition by P.M.S. Hacker, Blackwell Publishing, Oxford.

Bird, A., (2000), *Thomas Kuhn*, Acumen Publishing, Chesham.

Bird, A., (2011) "Thomas Kuhn's Relativistic Legacy", in S. Hales (ed.), *A Companion to Relativism*, Wiley – Blackwell, Oxford 2011, pp. 475 – 488.

Brown, H.I., (1983), ""Incommensurability", Inquiry 26, pp. 3 – 29.

Buzzoni, M., (1986), *Semantica, ontologia ed ermeneutica della conoscenza scientifica. Saggio su T.S. Kuhn*, Franco Angeli, Milano.

Doppelt, G., (1978), "Kuhn's Epistemological Relativism. An Interpretation and Defense", Inquiry, 21, pp. 33 – 86; reprinted in M. Krausz, J.W. Meiland (eds.), *Relativism: Cognitive and Moral*, University of Notre Dame Press, Notre Dame 1983, pp. 113 – 146.

Glock, H.J., (1996), "Necessity and Normativity", in H. Sluga, D.G. Stern, (eds.), *The Cambridge Companion to Wittgenstein*, Cambridge University Press, Cambridge 1996.

Hoyningen-Huene, P., (1989), *Die Wissenshaftphilosophie Thomas S. Kuhns. Rekonstruktion und Grundlagenprobleme*, Friedrich Vieweg & Sohn, Braunschweig; English translation by A.T. Levine, *Reconstructing Scientific Revolutions. Thomas S. Kuhn's Philosophy of Science*, The University of Chicago Press, Chicago and London 1993.

Hoyningen-Huene, P., Sankey, H. (eds.), (2001a), *Incommensurability and Related Matters*, Kluwer Academic Publishers, Dordrecht.

Hoyningen-Huene, P., Sankey, H., (2001b), "Introduction", in Hoyningen – Huene, Sankey (2001a), pp. vii – xxiv.

Kuukkanen, J., (2007), "Kuhn, the Correspondence Theory of Truth and Coherentist Epistemology", Studies in History and Philosophy of Science 38, pp. 555 – 566.

Kuhn, T., (1970a), *The Structure of Scientific Revolutions*, 2nd revisited edition, International Encyclopedia of Unified Sciences, vol. 2, no. 2, e University of Chicago Press, Chicago and London.
--- (1970b), "Postscript – 1969", in Kuhn (1970a), pp. 174 – 210.
--- (1970c), "Reflection on my Critics", in Lakatos, Musgrave (1970), pp. 231 – 278; reprinted in Kuhn (2000), pp. 123 – 175.
--- (1974), "Second Thoughts on Paradigms", in F. Suppe, (ed.), *The Structure of Scientific Theories*, University of Illinois Press, Urbana – Chicago – London 1974, pp. 459 – 482; reprinted in Kuhn (1977), pp. 293 – 319.
--- (1977a), *The Essential Tension. Selected Studies in Scientific Tradition and Change*, The University of Chicago Press, Chicago and London 1977.

--- (1977b), "Objectivity, Value Judgment, and Theory Choice", in Kuhn (1977a), pp. 320 – 339.
--- (1979), "Metaphor in Science", in A. Ortony, (ed.), *Metaphor and Thought*, Cambridge University Press, Cambridge 1979, pp. 409 – 419; reprinted in Kuhn (2000), pp. 196 – 207.
--- (1981), "What are Scientific Revolutions?", Occasional Paper #18, Center for Cognitive Science, MIT; reprinted in Kuhn (2000), pp. 13–32.
--- (1989), "Possible Worlds in History of Science", in S. Allén, (ed.), *Possible Words in Humanities, Arts and Sciences*, De Gruyter, Berlin 1989, pp. 9–32; reprinted in Kuhn (2000), pp. 58 – 89.
--- (1991), "The Road Since Structure", in A. Fine, M. Forbes, L. Wessels, (eds.), *PSA 1990: Proceedings of the 1990 Biennial Meeting of Philosophy of Science Association*, Philosophy of Science Association, East Lansing 1991, pp. 3–13; reprinted in Kuhn (2000), pp. 90 – 104.
--- (1992), "The Trouble with the Historical Philosophy of Science", Robert and Maurine Rothschild Distinguished Lecture, 19 November 1991, Occasional Publications of the Department of the History of Science, Harvard University, Cambridge, Massachusetts; reprinted in Kuhn (2000), pp. 105 – 120.
--- (1993), "Afterwords", in P. Horwich, (ed.), *World Changes. Thomas Kuhn and the Nature of Science*, MIT Press, Cambridge, Massachusetts 1993, pp. 311 – 341; reprinted in Kuhn (2000), pp. 224 – 252.
--- (2000), *The Road Since Structure*, edited by J. Conant, J. Haugeland, The University of Chicago Press, Chicago and London.

Lakatos, I., (1970), "Falsification and the Methodology of Scientific Research Programmes", in Lakatos, Musgrave (1970), pp. 91 – 195

Lakatos, I., Musgrave, A., (eds.), (1970), *Criticism and the Growth of Knowledge*, Cambridge University Press, Cambridge.

Malone, M.E., (1993), "Kuhn Reconstructed: Incommensurability Without Relativism", Studies in History and Philosophy of Science 24, pp. 63 – 93.

Newton – Smith, W.H., (1981), *The Rationality of Science*, Routlege and Kegan Paul, London.

Nola, R., Sankey, H. (eds.), (2000a), *After Popper, Kuhn and Feyerabend. Recent Issues in Theories of Scientific Method*, Kluwer Academic Publishers, Dordrecht.

Nola, R., Sankey, H. (2000b), "A selective Survey of Theories of Scientific Method", in Nola, Sankey (2000a), pp. 1 – 65.

Popper, K.R., (1959), *The Logic of Scientific Discovery*, Hutchinson & Co., London.

Putnam, H., (1981), *Reason, Truth and History*, Cambridge University Press, Cambridge.

Sankey, H., (1997), "Kuhn's Ontological Relativism", in D. Ginev, R.S. Cohen, (eds.), *Issues and Images in the Philosophy of Science*, Kluwer Academic Publishers, Dordrecht, 1997, pp. 305 – 320.

Scheffler, I., (1967), *Science and Subjectivity*, Hackett, Indianapolis.

Shapere, D., (1966), “Meaning and Scientific Change”, in R.G. Colodny, (ed.), *Mind and Cosmos. Essays in Contemporary Science and Philosophy*, University of Pittsburgh Press, Pittsburgh 1966, pp. 41 – 85.

Siegel, H., (1987), *Relativism Refuted. A Critique of Contemporary Epistemological Relativism*, Reidel, Dordrecht.

Williams, M., (1999), *Wittgenstein, Mind and Meaning. Towards a social conception of mind*, Routledge, London.

Wittgenstein, L., (1958), *Philosophical Investigations*, translated by G.E.M. Anscombe, Basil Blackwell, Oxford.

Are Colors Real?

Emiliano Boccardi
(Centre in Metaphysics of the University of Geneva)
emiliano.boccardi@gmail.com

Introduction

Do colors exist in the world as mind-independent properties or, as many have argued, are they *virtual properties*: properties the world *might* have instantiated, but in fact doesn't? I am here going to assume that this problem, also known as the *problem of color realism*, concerns the existence and nature of color properties as they are *represented* by visual experience. It is natural to think that, within this framework, the starting point for a discussion of color realism would be some theory of the representational content of perceptions. This is not supposed to be the result of a conceptual analysis of color concepts, or at least not only. It is something that we supposedly know by pure introspection on the content of our color experiences. But is it true? If our color experiences have a determinate content, and if they are (at least sometimes) veridical, this is a fact that falls beyond the grasp of our a-priori reason. It is possible, at least in an epistemic sense of the word, that color perceptions systematically fail to have determinate contents. It is also (epistemically) possible that, although they have determinate content, color perceptions systematically fail to be veridical, under a metaphysically thick notion of truth, or correctness. This, incidentally, opens the logical space for so called eliminativism about colors: the thesis that nothing in the world is really colored.

On the other hand, and for the same reasons, if at least some color perceptions do have veridical contents, this bounds the meaning of color perceptions at all possible worlds. In other words, if we discover that color properties are type-X properties of our world, this fixes the content of color

perceptions once and for all, in spite of the fact that there might be (counterfactually) possible worlds where color perceptions systematically fail to have determinate contents or to be veridical. I think the best way to describe this situation is through a two-dimensional theory of color perceptions. On one side is what I shall call the *character* of color perceptions. What are colors? If we are to be guided at all in answering this question, I argue, it must be the character of color perceptions that guides us. The character of perceptual experiences, as I see it, is a map from contexts to contents. It is the aspect of meaning that guides our enquiry into the reality of color properties.

You're looking at a red tomato on the table, and your perception seems to have the (propositional) content that there is a red tomato on the table. The event of your looking, and that particular tomato, (partly) constitute the token-context of your perception. The character of the perception fixes a content for that particular token color experience. The character trades with "representations" and their semantic properties, while the content, if it exists at all, only trades with material objects and their properties.

In a (Kripkean) sense, whatever the content of your experience happens to be, it must be a necessary intentional property of your particular experience. This property is necessary, but it is so a-posteriori: it is open to discovery what that particular content in fact is, if at all.This distinction between the character and the content of color perceptions, moreover, is what explains the significance of our enquiry. A philosophical theory of colors is cognitively revealing, it is informative, only if there is a difference between the character and the content of our color experiences. We have a-priori access to the character of color experiences, but not to their contents. Another way to express this thought, is to say that the character of color experiences, which is cognitively accessible for us, gives us (implicitly) a descriptive knowledge of the content of our experiences. This "description", if we are lucky (i.e. if at least some of our color perceptions are ever veridical, hence if they ever have determinate contents), must be sufficiently strict so as to fix a map from contexts to contents.

I think we are moderately lucky. I shall argue that the character of color perception, and the particular nature of our world, justifies some degree of optimism for the realist. There is at least one kind of properties that could constitute the content of veridical color perceptions. Such properties, however, I shall argue, are irreducibly extrinsic. So how does the character of color perceptions constrain the individuation of their contents? We said that

the character consists of a map from context to contents. How does this map work? How would we describe such map in a meta-language that contains both perceptual terms and terms for describing physical facts? Before answering this question, let me say something about the content of perceptual experiences in general. How are perceptual contents fixed, in general?

This question has occupied an entire sub-industry of philosophical enquiry for quite a long time now. So, if we are wondering how the character of color perceptions manage to fix their contents, it sounds like a good place to start would be a good theory of content. Moreover, as most scholars involved in the discussion share physicalistic intuitions (especially with regards to what fixes the content of *perceptual* experiences) a good place to start would be a *naturalistic* theory of content. In our case, a good naturalistic theory of content would provide us with a description of the mechanisms that underlie the character of color perceptions, viz. those mechanisms that constitute the mapping from perceptual contexts to perceptual contents. Many authors, however, have been discouraged from adopting this strategy because they have a poor opinion of the achievements of naturalistic theories of content so far. D. Hilbert, for example, concedes that "one way of settling the problem of color realism would be via some naturalistic theory of content". However, he goes on to argue, "none of these theories is well-enough developed to allow this sort of argument to be formulated in the required details".[1]

Now, while this is certainly true, I think that without some intuitions about what could help to fix perceptual content, we would be incapable of setting the whole enquiry about colors off the ground. What are we looking for, when we ask for the "real" content of color experiences? Moreover, suppose that we did succeed at individuating the "real" content of color experiences; how could we know that we have so succeeded, if we have no previous knowledge of what that content should be like to start with? As a matter of fact, I think that most debates assume, more or less implicitly, some restriction or other on the notion of perceptual content. These restrictions, moreover, function as tests for the adequacy of the various proposals. This goes relatively unnoticed because there is sufficient amount of agreement about what helps to fix contents in general.

If there is no consensus about the details of a naturalistic theory of content, in fact, there is wide agreement about a number of *necessary* conditions for something to be the content of a perceptual representation. I

[1] Byrne & Hilbert, 2003, 8.

think that these conditions are often implicitly at work in framing the debates and the various arguments for and against color realism. Among these presuppositions, for example, is the claim that the content of a *veridical* perceptual experience should be (at least in part) the *cause* of that experience, and that such causal relation contributes to individuating the content itself. In other words, many authors, in line with their physicalistic intuitions, assume a causal theory of content.

There is also wide agreement about the fact that the contents of perceptual experiences (e.g. color experiences), should mirror, at least to some extent, their phenomenal structure. Many arguments in the literature heavily depend on similar assumptions.[2] I propose to start by making these (purely semantic) presuppositions explicit (section 1.1). The peculiar character of color experiences is widely (and reasonably) believed to impose further, color-specific conditions on what fixes their contents. Among these, the intuition that physical objects should be the proper bearers of color properties, if anything is. In sections 1.2–1.7 I critically discuss a number of these further restrictions. With these restrictions in place, I go on to discuss the limits and scope of objectivist theories of color. I conclude that the content of color experiences must contain a relational property of objects. In particular, I argue that such relational property is the instantiation of the projection of the space of spectral reflectances (the distal stimuli) onto the 3-dimensional space of retinal (proximal) stimuli.

This has the somewhat unwanted consequence that part of the content of color experiences (on top of physical objects) are retinas. Retinas are part of the observers, if anything is, so, like most relationalist proposals, mine faces the challenge of mind-independence: under most understandings, if an object has the property of being (literally) colored, then whatever fact makes true the proposition that the object is colored must be a mind-independent fact, whatever this means.

The mind-independence restriction requires more then the mere objective nature of color properties. After all, the properties of our retinas are objective just as much as those of physical objects are. Retinas *are* physical objects! So, one may try to cheat, “although color properties are relational, and although the retinas of the observers are among the relata of color properties, such properties are nonetheless objective properties”. This would be cheating

[2] As I shall discuss in some details, for example, Hilbert’s account of metamers, or his contention that objects are represented as having proportions of hue magnitudes, implicitly draw on both of the above mentioned conditions.

because the rationale behind the mind-independence requirement is that it is to make room for faulty disagreements. What is red-relative-to-my-retina, may not be red-relative-to-yours, and this appears to block a-priori the possibility of error, or disagreement.

A second standard objection against relationalist theories of colors, is that part of what we perceive, when we perceive a colored object, is that color is a property that the object has monadically, not relationally. Monadic properties are typically conceived as necessarily intrinsic to their bearers, so that relationalist accounts appear to fly in the face of the very character of color experiences.

I believe my brand of relationalism has the resources to tackle both challenges. In section 1.5 I argue that the monadic character of color properties is relative to the mode of presentation of these properties in perception, and that it doesn't impose any restriction as to the extrinsic or intrinsic nature of the properties that are to be identified with colors. The character of color experiences, in other words, present color properties as monadic (it is the tomato that is red, not a system that includes something else, a part for the tomato), but this, we shall see, only entails that color properties must be *describable* as monadic, not that they must be intrinsic to their bearers.

My response to the observer-independence challenge is two-fold. According to my proposal, the character of color experiences only places second-order constraints on their contents.[3] This has the consequence that when I veridically perceive a red tomato on the table, what fixes the content of my representation is not some relation that only THAT tomato bears to MY retina. What fixes the content of the experience is the fact that THAT tomato and MY retina instantiate a second-order relational property: a property that could be instantiated by other (sufficiently similar) tomatos and other (sufficiently similar) retinas.[4] THAT particular tomato and MY particular retina right then, at most, make THAT experience veridical.

After presenting a formal toy model of color perception (section 2), I go on to consider a number of possible variants of my proposal (sections 3.1-3.3), testing them against standard anti-relationalist arguments. In particular, I test them against the threat of faultless disagreement, that hangs as a sword of

[3] In this respect, my proposal has a lot in common with functionalist accounts.
[4] My brand of relationalism, however, does have the consequence that if my retina was substantially different (as is the case with some non-human species), then the color properties of THAT tomato might turn out to be different.

Damocles over the heads of all non-physicalist accounts of colors. I conclude that a viable candidate is what I call a teleological relationalist theory of colors (section 3.3). According to this variant, what robustly fixes the content of color experiences, and makes room for genuine disagreement, is a teleological ingredient. Put crudely, according to this variant of my account, the character of any given color experience contains reference to *what would have had to have been the case, had the perceptual system actually harboring that experience instantiated it when functioning properly.*[5]

The account, then, rests on some naturalistic notion of proper function. I briefly mention a few alternative options as to how one may hope to naturalize functions (section 3.3), but ultimately I am interested in the viability of my account as a philosophical theory of colors.[6] I argue that color properties are objective mind-independent properties of physical objects. If I'm right, then, we can say that the world is populated by objectively (albeit relationally) colored objects. Some, however, will insist that objects are not *really* colored, if colors are not basic, intrinsic properties of them.

Here enters the second part of my response (section 4.1). I argue that all color experiences, independently of the particular makeup of the respective perceptual apparatuses, share the same character. Now, relational properties often possess "narrow correlates". The narrow correlates of a relational property are the properties of an object in virtue of which that object participate to the instantiation of the property (sec. 1.3-1.5). In the case of the relational property of weight, for example, the narrow correlate is mass. Mass is the basic intrinsic property in virtue of which material objects possess a weight, under suitable circumstances. According to my account, all color properties, independently on the observer, share the same narrow correlates, viz. the reflectance profiles of their bearers. This feature of my account, I shall argue, allows for as much room for disagreement as any other objectivist theory.

Because of the peculiar nature of the restrictions imposed on content by my account, moreover, all color properties, regardless of their observers, can be compared (metrically) with their common narrow correlates (reflectance profiles). Although reflectance profiles do not constitute, alone, the content of color experiences (they are not *the* colors), they have an essential role to play in any explanation of why we developed the capacity to perceive colors, or,

[5] I borrowed this way of expressing the teleological ingredient from Ruth Millikan.
[6] This will depend on the viability of some naturalistic theory of proper functions.

which is the same, in any explanation of why color perceptions can be so useful. The possibility to compare the contents of various color experiences as to how *accurately* they approximate the reflectance profiles of the bearers of color properties, therefore, provides us with a notion of relative "accuracy" of our perceptions. Once we know that a given color perception is veridical, according to my account, in fact, we can further ask how "accurate" it is. The perception is veridical iff its apparent bearer instantiates the relevant kind of relational properties. Such properties may approximate more or less accurately the reflectance profile of the object (i.e. the narrow correlate of the color property). My account, as we shall see (sections 3.2-3.3) allows for a quantitative notion of "accuracy". Depending on how much accurate the property in question is (in this technical sense), the correspondent perception will be said to be more or less accurate.

Because we can measure the distance of our color perceptions from "ideally accurate" color perceptions, moreover, we can judge how much our physical world is far from instantiating ideally accurate color contents. My verdict is: not much! We live in a quasi-colorful world (section 4.2). In the limit, as the degree of accuracy of various color experiences increases, I argue, my relationalist account conflates with standard realist accounts (such as Hilbert's), according to which colors are to be identified with reflectance properties of objects.. This, however, does not have the consequence that colors are, *really,* reflectance profiles. What colors *really* are depends solely on the character of color experiences in our world, and on how our physical world happens to be. Although what colors really are is a matter open for empirical discovery, I repeat, it is a *necessary* a-posteriori matter of fact. Whatever color properties turn out to be in this world (if anything does at all), those properties will be "the colors" at all other nomologically possible worlds.

Even if some creatures had retinas capable of discriminating and individuating single reflectance profiles of objects, this would not entail that what these creatures would perceive would be the "true" colors. Of course, the colors that these creatures would see would extensionally coincide with reflectance profiles. And of course the perceptions of these creatures would be much more accurate then ours (in the technical sense mentioned above). But this would not entail that the colors these creatures would perceive are the *true* colors. I consider my account to be a physicalist account of the nature of colors. My considerations, I hope, will allow us to avoid the consequence that if one rejects standard physicalist theories of colors, then one is

committed to think that nothing in the world is really colored: an admittedly embarrassing consequence.

Let us begin to make explicit the constraints that the character of color perceptions places upon their contents.

1. Constraints on the content of color experiences

1.1. Semantic desiderata

The following are widely accepted conditions that a property must satisfy for it to be (part of) the content of a perceptual experience.

1) Co-variation condition. Veridical color experiences form a domain whose (phenomenal) structure is (at least) omeomorphic to that of their contents.

This assumption derives from widely shared epistemological tenets. Both those who believe that sense-data mediate our experience of the external world, and those who believe that we have direct experience of the external world, will claim that we have (direct or indirect) experience also of the *structure* of the causes of our perceptions. Moreover, most philosophers find it plausible that such structure is (at least partly) captured by the phenomenal structure of our experiences.

So, for example, our auditory experiences of certain sounds can be arranged according to their pitches (e.g. Do, Re, Mi, Fa, Sol, La, Si). Call the structure determined by the relations of perceptual pitch similarities among these experiences: P_{phen}. According to the co-variation assumption, perceptual auditory experiences represent the world as instantiating (at least) the structure P_{phen}. It follows that P_{phen} consists of veridical auditory experiences only if a portion (D) of the world (W) is such that there exists relations defined on D such that their structure is homeomorphic to P_{phen}:

For some $D \subseteq W$ there are $R_1, R_2, ...R_n$ on D such that $(D; R_1, R_2, ...R_n) \cong P_{phen}$

Although this condition is the trademark of internalist theories of representational content (e.g. conceptual role theories, or Cummins' theory of content), most philosophers in the "causal camp" also sympathize with it. This is how Dretske expresses this requirement, for example.

> The fundamental idea is that a system, S, represents a property, F, if and only if S has the function of indicating (providing information about) the F of a certain domain of objects. The way S performs its function (when it performs it) is by occupying different states s1, s2,...,sn corresponding to the different determinate values f1, f2,...,fn,of F.[7]

Millikan is even more explicit on this point.

> [R]epresented conditions are conditions that vary, depending on the form of the representation, in accordance with specifiable correspondence rules that give the semantics for the relevant system of representation. [8]

2) Causality condition. The content of a veridical perceptual experience must be part of the cause of that experience, at least under some epistemically salient conditions.

As Hilbert points out, "any plausible version of physicalism will identify colors with physical properties implicated in the causal process that underlies the perception of colors". I would add that this is a desideratum of any non-eliminativist theory of colors that wishes to comply with physicalistic intuitions, and not only of the brand of physicalism advocated by Hilbert. The caveat on "epistemically salient conditions" is to avoid a vacuous notion of content, or, if you wish, it is to make room for epistemic error. More about this later (section 3.1-3.3).

3) Asymmetric dependence condition. If it is (nomologically) possible for a given non-veridical perceptual experience to be veridical, then its causes (qua causes of that experience) asymmetrically depend on the causes that the experience would have had, had it been veridical.

Fodor notoriously proposed a causal theory of content whose essential ingredient is the asymmetric dependence of the causes of non-veridical perceptual experiences on the causes of veridical ones. While it is still controversial whether this places sufficient (or substantial) constraints on the individuation of content, it seams to me safe to claim that any causal theory of content should be such as to have this condition come out true.[9]

4) Robustness condition. The content of a given perceptual experience must be robustly the same, regardless of whether the experience is veridical or not.

---

[7] Dretske, 1995, 2.
[8] Millikan, 1990, 224.
[9] As we shall see, however, nothing in my arguments hinges on the assumption that this condition holds.

We shall discuss this condition at length. For the moment, it suffices to say that the condition, among other things, is to make room for disagreement. If I say that a certain object is red, and you think I'm wrong, then we better mean the same thing by "red", otherwise our disagreement would be only apparent. More on this later.

1.2. Color experiences

Let us apply these general constraints to the case of the content of color perception. Let C_{phen} be the phenomenal structure of color experiences as of their hues. It consists, suppose, of the structure of similarities among them, plus the opponent structure. Let $C_{sim\text{-}phen}$ and $C_{op\text{-}phen}$ name respectively the similarity substructure and the opponent substructure. The above mentioned conditions on the individuation of perceptual content, then, allow us to say that color experiences are veridical if:

1. For some domain $D \subseteq W$, there are relations $(S_1, S_2, ...S_n)$ on D, such that: $(D; S_1, S_2, ...S_n) \cong C_{sim-phen}$ and $(D; S_1, S_2, ...S_n) \cong C_{op-phen}$
2. Under epistemically salient conditions, the instantiation of $(D; S_1, S_2, ...S_n)$ causes the instantiation of $C_{sim-phen}$ and $C_{op-phen}$
3. If an instance of C_{phen} is non veridical, its cause must depend asymmetrically on the relation that obtains between C_{phen} and its causes when C_{phen} is veridical.
4. The contents of C_{phen} would have been the same, even if the experiences in C_{phen} had not been veridical.

If we assume these conditions, then they provide us with constraints on what colors may be taken to be (if they exist at all): if the content of color perception is ever veridical, colors must (at least) be properties satisfying conditions 1-4. These constraints derive from the assumption that color perception is a representational phenomenon (i.e. that it involves tokening representational properties), plus the thesis that color properties are part of the content of color perceptions. It is easy to realize, however, that these conditions place very weak constraints, by themselves. In fact, without filling

in the details of their interpretation, the constraints are compatible with virtually every representational account of colors of which I'm aware of. My thesis, I anticipate, is that under the only sensible interpretation, these constraints are sufficient to rule out all but a relationalist accounts of colors.

The further constraints that need to be added, to individuate what kind of properties colors are (if they exist at all) come from conceptual considerations about the particular nature of color, as well as from our extensive knowledge of optics, colorimetry and the neurophysiology of color perception.

1.3. What are the bearers of color properties?

Notice, first, that the four conditions given above, short of further indications as to how one should interpret them, leave open the question of what portions of the world are to provide for the class of possible instantiations of the domain $D \subseteq W$. Should the portion D of the world include the brain of the perceiver? Should it also include the whole environment? Or should it only include the (supposedly) colored objects? One possible restriction can be justified by the following argument. If the account is to construe of colors as observer-independent properties, the domain D should be taken as excluding at least our brains (and our retinas). This, as we shall see, does not, by itself, commit us to say that the properties instantiated by D must not be ultimately related to the brain. It just means that the bearers (if at all) of the color properties represented by color experiences are to be found outside of the brain of the perceivers (if any). The intuitive argument for this thesis seams to be the following.

A minimal requirement for a property to be mind-independent (however one wants to construe this notion), is for it to be a property that is not necessarily co-instantiated with any mental property. Necessary co-instantiation, in fact, is a sign of "dependence", under all sensible understandings of the word "dependence". Since, presumably, the brain instantiates mental properties, the requirement that color and mental properties are never necessarily co-instantiated entails that the bearers of color properties must be found entirely outside of the brain. I will come back to mental independence later (section 4). We shall call this restriction the:

5) Externality condition. The bearers of color properties, if any, must be physically disjoint from the brain of their potential perceivers.

Another line of argument that places a-priori constraints on the suitable instantiating domain comes from our intuitive conceptual knowledge of colors. One may reason as follows. According to our pre-theoretical understanding of color concepts, colors, if they exist at all, must be properties *of the objects* that we perceive (or of their surfaces). Forget about what *kind* of properties colors are for the moment: whatever they are, they certainly must be properties *of the objects*! Should it turn out that, under closer scientific scrutiny, there are no properties *of the objects* that comply with conditions 1-5, then, too bad for real colors! In that case one should say that we perceive the world *as if* objects instantiated color properties, when in fact they don't.

The intuition that a mere introspective scrutiny of color perceptions will reveal something about the metaphysics that they presuppose, is very strong, and indeed very widely held. We could try to explain this intuition by saying that perceptual experiences have, among their properties, a formal, predicative structure. If I perceive a red object, I come to believe that I am in front a red object: "the proposition that there is a red bulgy object on the table is part of the content of the subject's experience", says Hilbert for example.[10] If perceptual experiences have (also) a propositional content, one cannot, supposedly, have a visual experience, without thereby coming to know its propositional content. Propositional contents, in turn, have a predicative structure,[11] whence the metaphysical presuppositions. Let us call this:

The Propositional Content Assumption. The content of perceptual experiences consists partly of structured propositions. By perceiving a visual scene, subjects also perceive the predicative structure of these propositions.

Setting aside the question of where these presuppositions come from, let us now turn to the consequences they would presumably have for a theory of color. Not only does perception present objects as colored, but perception also presents what these colors are like.

> When [a person] perceives a blue bead, not only does he perceive the bead to be blue, but he perceives what blue is like. The qualitative nature of the colors is manifest to us in our perception of them. Objects are perceived to instantiate color properties, and these color properties are perceived to instantiate higher-order properties that constitute their qualitative character. So, not only does

[10] Byrne & Hilbert, 2003, 5.
[11] By this I mean that grasping a proposition entails, eo ipso, grasping its surface logical structure, viz. grasping what is predicated of what.

> color perception present the existence and distribution of the colors, but it also presents their nature.[12]

Both eliminativists and realists about colors may sympathize with this line of argument.

The eliminativist master argument is that if colors cannot be thought of as properties that inhere in the objects and that cause our color experiences (in the counterfactually strong sense described above), then, we must conclude that nothing is really colored.

Most realists would also find this argument convincing. Hilbert, for example, argues that the representationalist theory of color perception entails that "the view that no physical objects are colored is equivalent to the view that the contents distinctive of color experiences (for example, that there is a red bulgy object on the table), are uniformly false".[13] On similar grounds, many typically discard as inadequate the idea the colors may be properties of light. Let us call this:

6) The proper subject condition. The proper subject of color ascriptions are physical objects

I think condition 6 is essentially correct, but that it hides a potential unwarrented presupposition: that if an object possesses a certain real property, it must possess it in and of itself. Before turning back to this important point, let me continue with our analysis of how we should fill in the details left open by the four semantic conditions on color content.

1.4. Are color properties relational?

Conditions 1-6 leave open what sort of properties are to constitute the domain of instantiation of the structure $(D; S_1, S_2, ...S_n)$. What kind of properties are colors? Are they extrinsic or monadic? Dispositional or non dispositional? And if they are dispositional, do they involve a relation to the (cognitive system of the) perceivers or not? Is there any a-priori argument that could help us to individuate the kind of property that colors are, if anything is a color property? As we have already seen, a part from the restrictions imposed by conditions 1-4, one can try to place constraints derived from our intuitive notion of color, or from the alleged metaphysical presuppositions of color

[12] Kalderon, 2007, 563.
[13] Byrne & Hilbert, 2003, 5.

experiences (conditions 5-6). In the previous section, for example, we argued that the proper subjects of color ascriptions must be physical objects. Does this place constraints on the kind of properties colors might be (if they exist at all)? Prima facie, I think, we would answer in the affirmative. This would be the argument.

If colors are properties of physical objects, and if they (or rather their instantiations) must be causally efficacious (condition 2), then colors must be *physical* properties of physical objects. "[I]t is of course *the object* that looks colored […]", says Hilbert for example, "and so the relevant physical property must be a property *of* objects".[14] Now, at a first glance, it seems that if this reasoning is sound, then we should restrict the domain of instantiation of color properties to *monadic*, intrinsic physical properties of material objects. But this is certainly wrong. No one thinks that this is what has been shown (not even Hilbert). But I think it is interesting to see what is wrong with this conclusion.

Consider the following example. The physical world appears populated by more or less heavy objects. When someone has a tactile experience, the tactile scene appears to the subject to be one way or another. Just like the proposition that there is a red bulgy object on the table is part of the content of a visual perceptual experience, the proposition that there is a heavy object in your hand, is part of the content of your tactile experience. Now, everything that we said about colors (conditions 1-6), also apply to this case. If your experience is to count as veridical (at least possibly veridical), then it must be taken as representing the world as populated by objects that possess the property of being heavy. A line of reasoning virtually identical to the one that we have seen above, lead us to conclude (correctly, I think) that the bearers of this property, if any, must be material objects. The co-variation condition on the content of representations, moreover, leads us to conclude that, if our perceptual experiences are ever to be veridical, the property in question must be a magnitude of some kind. The causality constraints, finally, entail that such property must be a physical magnitude instantiated by material objects.

It is rather straightforward, given our background knowledge of physics, to conclude that the property represented by this experience is weight. The property of having a certain weight, in fact, complies with the three desiderata on the content of representations (1-4), and with our pre-theoretical intuitions as to what kind of property it is, as well as to what entities could bear it (5-6). In this case, it is clear that these considerations, by themselves, do not entail

[14] Byrne & Hilbert, 2003, 9, my emphasis.

anything about the particular nature of weight. We know, on independent grounds, that weight is a relational property: it is a property that material objects have relative to the earth.[15] But nothing to this effect follows from a priori arguments.

It is worth to pause a moment to think about the representation of relational properties. First, does the fact that weight is not an intrinsic property mean that weight is not *really* a property possessed *by* material objects? Should we say that, strictly speaking, the property is *really* possessed, say, only by a system that comprises the object and the earth? If so, we should conclude that the proper subjects of weight ascriptions should be entire astronomical systems. But this is certainly wrong! The system that comprises the object that you're holding in your hand and the planet beneath your feet, is not the proper bearer of the property, as this is represented by the predicate *is heavy*! The object is the bearer of the property (relative to its mode of representation) regardless of whether the property is intrinsic or relational.

The property of being a hundred meters away from a plumber is clearly a relational property. But if you are a hundred meters away from a plumber, it is *you* who are a hundred meters away from a plumber! In this case, because of the predicative structure picked up by this particular representation of that property (viz. its conceptual content), *you* are the proper subject. Notice, however, that the same (relational) property can be presented in such a way that its proper bearer is, instead, the plumber. If the predicative structure intrinsic to a representation of that property had the plumber as its proper subject, then it would be *the plumber* that has the property of being a hundred meters away from you. Similarly, the same property can be seen as an intrinsic property of a pair constituted by you and the plumber. In this case, the proper subject of ascription of the property would be the pair constituted by you and the plumber.

Distinguishing intrinsic and extrinsic, or monadic and relational properties, is notoriously a tricky task. I do not wish to delve into the details of these distinctions here, but some clarification is in order. Let me introduce some useful concepts and distinctions. First, intuitively, whether a property is relational or not, seems to be a matter of objective fact, that can be subject to rational and empirical scrutiny. Given what we know about physics, for example, it seems that weight is unquestionably and objectively a relational property of material objects. This "fact" appears not to be relative to a

[15] Even Aristotle thought that, though for different reasons.

particular way of picking up (or of representing) the property. Yet one may reason as follows.

We think that weights are relational properties of material objects because their instantiations are always conditional on the presence, feature and distribution of other (astronomical) material objects. Why do we think that? Because we believe that the weight of an object is due to the gravitational force exerted on it by the presence of other massive objects. Change the distribution or the masses of these other objects, and weight changes accordingly; whence the idea that weight cannot be an intrinsic property. But why do we think that weight is *due to* gravitational forces, rather then thinking that it *consists of* gravitational forces? After all the weight of an object is nothing but one manifestation of the gravitational forces exerted upon it. So why not say that weight is the same property as (rather then being caused by) gravitational attraction, under certain circumstances?

But if weight is nothing but gravitational forces, then whether it is a monadic property or not depends on what portion of the world we take as its relevant bearer. Gravitational attraction, in fact, is a monadic property of the system that comprises the object and the planet. It is relational only if its bearer is taken to be the object alone. I think that it is safe to conclude from this example that whether a property is relational or not, is a matter that is relative to factors that do not depend on its intrinsic nature. One and the same property has different "modes of presentation", as it were, depending on how it is picked up by its representations. Presented as a property of the object, weight is relational, while presented as a property of a larger system, it is monadic.

We argued in the preceding section that the proper subject of color ascriptions must be material objects (condition 6). Now we can see that this condition, by itself, does not constrain the metaphysical nature of color properties. It constrains the nature of color properties only relative to our mode of representing them. Following these considerations, from now on, in stead of saying that a property *is* relational, we shall say it is *relationally fixed* (by a given representation). We should then better express condition 6 as follows:

6*: Proper subject condition. The proper bearers of color properties (as these are fixed by our color perceptions), are physical objects

1.5. Relational properties and their narrow correlates

Some relational properties can be thought of as relating a narrow correlate (relatum) with a wider correlate. The narrow correlate of a relational property R of an entity, is the intrinsic property (or properties) of that entity in virtue of which the entity contributes to the instantiation of R. In the case of weight the narrow correlate is mass. An object possesses the weight that it does in virtue of having a certain mass. Mass is (a) one of the relata that constitute the property of weight (the other relata being all the relevant celestial bodies in the surroundings); mass also happens to be (b) an intrinsic property of the proper subject of weight ascriptions (physical bodies).[16] These two features of mass make of it the "narrow correlate" of weight. I am going to argue that color properties must be relational, and that the reflectance profiles of their bearers are their narrow correlates.

A more precise definition of narrow correlate requires that we distinguish basic from non-basic properties. Intuitively, a property is non-basic if it is instantiated (when it is instantiated), in virtue of the instantiation of some other property. It is basic otherwise. To pin down this notion, I shall introduce the following:

Substitutivity Test. For any property P and a pair of objects x and y, it is true that, when x is in a nomologically possible context that fixes that x is P, had y been in that context instead of x, y would also have had P. Basic properties, intuitively, are those such that two objects sharing them pass the substitutivity test:

A *basic property* is a property that belongs to the minimal subset B of the properties of the world that satisfies the following requirement: every two objects that share all their B properties pass the substitutivity test. The notion of basic property can be used to define narrow correlates in a more precise fashion. The *narrow correlate* of a relationally fixed property R of an object Q, is the single minimal set of basic properties of Q by virtue of which it has the ability to contribute to the tokening of R. It is interesting for our discussion of color properties, I think, to make a few further remarks about narrow correlates. As I have already anticipated, I am going to argue that colors are relational properties, whose narrow correlates are reflectance profiles. What

[16] For the sake of the example I am pretending that our physical world is non relativistic.

we represent in our color perceptions, I shall argue, are these relational properties, and not their narrow correlates (the reflectances).

Some might have the (fallacious) intuition that the real content of a veridical perceptual experience of a property is always a narrow correlate. If colors are objective, mind independent properties of objects, one may think, they better be intrinsic properties of objects! Narrow correlates are intrinsic properties of the bearers of color properties, and they must be (at least partly) causally responsible for our perceptions. So why not think that it is the narrow correlates (the reflectances) that we represent? I think that this intuition derives from the predicative structure of our perceptions. We ascribe color properties to physical objects, by attaching to their names/descriptions monadic predicates such as "is red". It could be argued that this predicative structure (subject/monadic-predicate) is part of the implicit content of color perceptions. In other words, we instinctively think of monadic properties as inherent to their bearers, whence the intuition. It is interesting, for the purpose of exposing my thesis, to see how far this intuition can get. It can be spelled out as follows.

Suppose that a phenomenal structure W_{phen} represents the physical structure $(D;R_1,R_2,...R_n)$. Suppose further that a candidate for the instantiation of $(D;R_1,R_2,...R_n)$ is a certain class of relationally fixed, physical properties. As we said these relational properties must have narrow correlates. If this is so, should we not conclude (a priori) that the properties that are *really* represented by W_{phen} are these narrow correlates? Consider again our example. From the fact that weight has a narrow correlate, does it follow that what you are representing when you experience a heavy object in your hand, is, *really*, its mass? This is a tricky question. Notice, in fact, that the magnitude mass appears to comply with all the relevant desiderata, just as well as weight does. As I shall argue, however, the magnitude mass fails to comply with the robustness condition, hence the existence of narrow correlates will not affect, by itself, a given metaphysical account of perceptual representation. First, let us try to push the case for narrow correlates as far as it can get.

Co-variation condition. The instantiations of the magnitude mass can be arranged so as to have a structure that mirrors perfectly well those of the magnitude weight. If weights instantiate $(D;R_1,R_2,...R_n)$, then so do masses. Hence mass complies with condition 1 on the individuation of content.

Causality condition. If instantiating certain weight properties causes a perceiver to instantiate W_{phen}, then, a fortiori, so does instantiating their respective narrow correlates. After all, it is the instantiation of a certain masses that cause the instantiation of a certain weights, which in turn cause the instantiation of W_{phen}. Hence mass complies also with condition 2.

Asymmetric dependence condition. Remind that the asymmetric dependence condition states that: if it is (nomologically) possible for a given non-veridical perceptual experience to be veridical, then its causes (qua causes of that experience) asymmetrically depend on the causes that the experience would have had, had it been veridical.

Thus, suppose that you are hallucinating holding various heavy objects in your hand. This means that you instantiate the phenomenal structure W_{phen}, although there is nothing heavy in your hand. The cause of this instantiation is, say, that some scientist stimulates your neurons in the appropriate way. Strictly speaking, the proximal cause of the instantiation is a certain pattern of stimulation.The rationale behind the asymmetric dependence condition is that we would like the following hypothetical conditional to come out true. Had not the presence of heavy objects caused the instantiation of W_{phen} in the past, then the same pattern of stimulation that now causes the instantiation of W_{phen}, would not be causing it. This is the essence of the "dependence" condition in question. Of course, the above conditional may turn out to be vacuously true in the case that there exist no heavy objects in reality. So, if the condition is to cut some ice, it must be understood under the assumption that weight perceptions are, some times at least, literally veridical. Let us turn back to our question: do narrow correlates always also comply with the asymmetric dependence condition? It appears that they do. Suppose we take the magnitude mass (and not weight) to be part of the content of the proposition that you are holding a heavy object. The asymmetric dependence condition, then, would take the following form: had not the instantiation of mass caused the instantiation of W_{phen} in your past, then the same pattern of stimulation that now causes the instantiation of W_{phen}, would not be causing it. It is easy to realize that if weight complies with this condition, then so will the magnitude mass.

Robustness condition. I shall argue that narrow correlates sometimes fail to comply with the robustness condition. This is the case, for example, I argue, of colors. The robustness condition states that the content of a perceptual experience must be the same, regardless of whether the experience is veridical or not. So, if the content of a veridical experience to

the effect that you're holding a heavy object, is that there is an object with a given mass in your hand, then this should be the content of your experience, also in cases in which the experience is non veridical.

Now, imagine holding the same object in outer space. If, under these circumstances, you were nevertheless to experience the presence of a heavy object in your hand, your experience would not be veridical.[17] But the robustness condition imposes that the content in the two circumstances be the same. Hence, also now that you are hallucinating weight in outer space, the content of your experience is that there is an object with a given mass in your hand. But it is true that there is an object with a given mass in your hand! So your experience must be veridical, contrary to the hypothesis. This is enough, I believe, to conclude that the narrow correlates of relational properties that comply with the relevant desiderata for being the content of a given perceptual experience, are not necessarily the "true" contents of that experience.

A second remark on narrow correlates is in order. Which relational properties possess narrow correlates, and which don't? I don't have a full answer to this question, but it seems reasonable to assume that if a (relational) property is to have autonomous casual powers, as is the case with color properties, then it must have a narrow correlate. If this proves to be correct, then the causal condition on the individuation of the content of color experiences entails that, if colors are relational properties, they must have a narrow correlate. Before applying all that was said to the problem of color realism, let me make some further remarks about the conditions for identifying perceptual content.

1.6. The causality condition

There is an ambiguity in the expression of the causality condition, as expressed above. If the instantiation of the structure $(D; S_1, S_2, ...S_n)$ is the content of veridical color experiences, we said, it must *cause* the instantiation of structure C_{phen}. Now, there are infinitely many ways in which the world may instantiate both structures. So, to say that the instantiations of token-

[17] Imagine, for example, that while (really) holding the object in outer space, you find yourself in the same mad scientist scenario as before.

structures of type $(D; S_1, S_2, ...S_n)$ cause the instantiation of token-structures of type C_{phen} can be taken to mean either of the following:

> 1.It can be (minimally) taken to mean that each token of the structure $(D; S_1, S_2, ...S_n)$ causes a token of the structure C_{phen}, but no counterfactual causal conditional holds between the former and the latter. Or, maximally,
> 2.It can be taken to mean that the causal relations among the tokens of the two structures hold *because* there exist a law that causally connects the instantiations of $(D; S_1, S_2, ...S_n)$ with the instantiations of C_{phen}.

The essential difference between these two interpretations is that according to the second the causal relation between the two structures supports counterfactual conditionals, while according to the first it consists of mere material conditionals. Which of the two interpretations is most sensible? Remind that the purpose of the causal requirement is to participate in the individuation of content. Now, content (if color experiences have contents at all) must be robustly the same at different times and under different circumstances (robustness condition). The causal requirement, then, must be interpreted as supporting counterfactual claims. Not only must be the case that tokens of type $(D; S_1, S_2, ...S_n)$ accidentally happen to cause tokens of type C_{phen} under given circumstances. In a case where tokens of type $(D; S_1, S_2, ...S_n)$ are not instantiated, it must still be true that any token of type $(D; S_1, S_2, ...S_n)$ *would have* caused a token of type C_{phen}, had the former been instantiated.

In short, if we take the first interpretation of the causality condition to be the correct one, then the condition would merely suffice to say that certain properties cause our *allucinating* color properties, while we want the casual condition to help us grounding veridical color *representations*. This requires, as we said, that the members of the instantiation basis for a given color property share some relevant second-order properties, over and above the accidental fact of causing the same perceptual experience. In other words, the instantiation bases of color properties must carve nature at its joints. This leads us to opt for the second interpretation. I shall argue that, under this

interpretation, standard versions of color physicalism are not compatible with the causal condition.

1.7. A posteriori constraints on color properties

Finally, a number of constraints on what colors may reasonably taken to be come from our impressive body of knowledge about color processing in the visual system, from psychological data about color perception, from the optical properties of physical objects, and from how these may be recovered by our perceptual systems. Psychological data, for example, show a certain degree of constancy in the perception of colors. Objects appear to retain their color properties under very different environmental conditions. In particular, they appear to retain their color properties in spite of significant changes in illumination conditions. This suggests that colors, whatever they may be, should be properties that do not depend (to a too great extent) on illumination conditions:

6. Color properties must be retained under significant changes in the spectral power distribution and wave-length composition of the illuminant.

Finally, as noted by various authors, colors must be properties that can (at least under certain ideal conditions) be recovered by our perceptual apparatuses. We know that all the information about color properties is processed in the human brain from the patterns of stimulation of three types of photoreceptors in the retina: the L-, M- and S- cones. Light of various wavelengths stimulate these types of cells to varying degrees. Red light, for example, stimulates the L-cones much more than the M-cones, and it hardly has any effect on S-cones. This suggests that colors satisfy also the following desideratum:

8) Recoverabilty condition. Color properties, whatever they are, must be such as to be (at least approximately) discernable and identifiable by processing information that consists solely of patterns of stimulation of the three types of cones in the retina.

2. The geometry of color perceptions

In this section I introduce some formal properties of color perception. My aim is to provide a toy (formal) model of color perception. It should not be

taken as a realistic model: its purpose is simply that of clarifying my relationalist proposal.

2.1. Geometry of color stimuli

Physical color stimuli can be represented by functions *C*(*w*) from the range of visible wavelengths (represented by the interval of real numbers $I = [\mathrm{W}_{\min}, \mathrm{W}_{\max}]$) to the real numbers. In the intended interpretation, these functions assign to each wavelength $w \in I$ its intensity *C*(*w*). Each of these functions is a (linear) combination of pure "spectral color stimuli", i.e. stimuli whose intensity is non zero only for one wavelength value $\overline{w} \in \mathrm{I}$. Physical color stimuli, thus represented, are elements of a Hilbert space of square-integrable functions: *H(I)*.

As we said, stimuli of various wavelengths stimulate the three types of photoreceptive cells in the retina to varying degrees. Such "degrees" can be represented by three functions: s(w), m(w) and l(w). The "extent" to which a given physical color C(w) stimulates each of these receptors, can thus be calculated, respectively, as:

$$\int_{W_{\min}}^{W_{\max}} C(w)s(w)dw, \quad \int_{W_{\min}}^{W_{\max}} C(w)m(w)dw, \text{ and } \int_{W_{\min}}^{W_{\max}} C(w)l(w)dw.$$

Perceived colors can then be represented as points in a three-dimensional space: R^3_{color}. The relations between these points and the functions in *H (I)* will be crucial for our proposal. Let me spell them out in some more details. For reasons to be discussed later, the "human" case of a 3-dimensional perceptual space will be generalized to an N-dimensional space. Given the Hilbert space of physical color stimuli *H (I)*, we select an N-dimensional subspace: $H_N(I)$. We introduce, for $H_N(I)$, a basis: $\mathrm{b_n(w), n = 0, \ldots N}$. Each element $C \in H(I)$, can be approximated by its orthogonal projection onto $H_N(I)$. If we indicate the projection operator with O:

$$C(w) \approx O \cdot C(w) = C_N(w) = \sum_{\mathrm{n=0...N}} \beta_n b_n(w)$$

The coefficients are calculated as follows: $\beta_n = \langle C, b_n \rangle$, where $\langle .,. \rangle$ is the scalar product of *H(I)*.[18] Let S be the subset of *H(I)* that represents the color stimuli. Among these, the monochromatic stimuli, $m_{w_0}(w)$, are defined as those stimuli that are concentrated at some wavelength $w_0 \in I$. Call "black" the function $B \in S : B(w) = 0$, for all $w \in I$. And call "white" the function $W \in S : W(w) = 1$ for all $w \in I$.The line connecting the monochromatic stimulus $m_{w_0}(w)$ with the white point *W(w)* crosses the boundary of **S** at $m_{w_0}(w)$, hence the half-line $\{c \cdot m_{w_0}(w_0), c \geq 0\}$ lies at the boundary of S.

It follows that, given any two stimuli $C_i(w) \in S : i \in \{1,2\}$, their linear combinations $c \cdot C_1(w) + (1-c) \cdot C_2(w)$ are also stimuli, for all $0 \leq c \leq 1$. Thus S is convex. More precisely, it is the convex closure of the set of monochromatic stimuli.[19]

The image $S_N = \{O \cdot C : C \in S\}$ of S is a subset of $H_N(I)$ (also known as the "spectral locus"). The projection operator (O) can be chosen so that the spectral locus lies at the boundary of S_N. This matches the fact that monochromatic spectra lies at the boundary of *S*. The line in $H_N(I)$ connecting the projections of the limit points $m_{w_{\min}}(w)$ and $m_{w_{\max}}(w)$, viz. the line connecting $O \cdot m_{w_{\min}}(w)$ and $O \cdot m_{w_{\max}}(w)$, is called the "purple line". Both S and its image S_N consist of half-lines departing from the black point. They both form a (mathematical) cone, whose vertices are the spectral colors and whose apex is the black point. Each half-line in the S-cone represents a given color stimulus. Receding from the apex (the black point), the stimulus retains its chromaticity, while increasing its intensity.

As we said, human color space is three-dimensional because our eyes contain three types of receptors, each with its own type of spectral response. The projector operator, in the human case, maps the set of stimuli $S \subseteq H(I)$ onto a subset of a 3-dimensional space $S_3 \subseteq H_3(I)$. The choice of a basis for this space is rather arbitrary. At the beginning of this section, we suggested that a basis could match the fundamental response functions of the receptors in the eye. This, however, is not imposed upon us. Any three linearly

[18] This ensures that the basis is ortonormal
[19] From Grassmann's laws.

independent combinations of these bases will constitute a suitable basis for the same color space.

A concrete manifestation of this arbitrariness is the fact that color-matching data from normal individuals underdetermine the eye's primary response functions. Indeed, "[a]ll the colors of the spectrum […] can be mimicked by combinations of different intensities of […] blue, green, and red".[20] The amounts of the three primaries required to match a given color are called its "tristimulus values". Because any three linearly independent combinations of these color-matching functions is also a triplet of color-matching functions, the choice of "primaries" is arbitrary, so long as their vectors in color space are not coplanar. The projector vector O, can only be determined by empirical investigations performed on human (or other) observers. The characteristics of vector O depend on what sets of spectral stimuli are visually identical to a given subject. Any two stimuli belonging to such a set, are called metamers. Mathematically, metamers are stimuli mapped onto each other by functions whose projection under O is null.

2.2. The structure of phenomenal colors

In the previous paragraph we described some geometric properties of color stimuli. These stimuli are processed and modified by our perceptual apparatus shortly after being input to the cognitive system. Whatever the processes involved in this information processing, the result of them is the phenomenal structure of color properties as we experience them. There are several ways in which we can investigate empirically this structure. Probably the best known is the "Munsell color system". It is a 3-dimensional color space based on the phenomenal dimensions of hue, value (lightness), and chroma (color purity). It was introduced by Albert H. Munsell at the beginning of the twentieth Century, and was improved in the following decades through extensive (psychological) experimental studies. For the purposes of this paper, the details of Munsell color space are not relevant. It suffices to note the following.

Munsell color space can be represented cylindrically in a 3-dimensional space as an irregular color solid. For the purposes of my argument, as I said, it is irrelevant whether the details of this particular solid accurately represent

[20] Malin and Murdin,1984, 35, 60-61.

phenomenal color space. It will be here taken to represent the structure of phenomenal color experiences, whatever they are. What I mean, by this, is that minor changes in the detailed structure of Munsell cylinder won't affect the strength of my argument. In the notation introduced at the beginning of this paper (§ 1.2), Munsell color cylinder will be taken to be the structure C_{phen}. As noted in paragraph 1.2, the contents of color experiences, if they are ever veridical, must be such that:

For some portion of the world (subdomain $D \subseteq W$), there are relations ($C_1, C_2, ...C_n$) on D, such that: $(D; C_1, C_2, ...C_n) \cong C_{phen}$.

I shall argue that, contrary to most physicalist proposals, such structure must be homeomorphic to the N-dimensional subspace of the Hilbert space of color stimuli introduced in the previous section. I shall further argue that such structure can only be instantiated if colors are taken to be relational properties. According to my proposal, I anticipate, the reflectance profiles of the surfaces of material bodies are the narrow correlates of these relational properties.

3. Colors as instantiations of orthogonal projectors

Given the restrictions placed on color properties by the character of color perceptions (conditions 1-8), it follows that the properties represented by color experiences cannot be the spectral reflectances of the surfaces of objects. Spectral reflectances, in fact, instantiate at best the structure $S \subseteq H(I)$ described in section 2.1. The entities in this subset do not naturally instantiate the structure $C_{phen,}$ as required by the co-variation condition (condition 1). Intuitively, this means that spectral reflectances do not stand to each other in the right similarity relationships. If we identify colors with spectral reflectances, for example, then two reflectances belonging to a metameric pair should count as two different colors, while they appear to be exactly the same to all normal observers.

While, as we shall see, some authors are prepared to bite the bullet on this point, I think there are reasons to think that this is a drawback of standard physicalist accounts. More strikingly still, phenomenological colors that correspond to monochromatic stimuli ($m_{w_0}(w)$) vary continuously as w_0 takes up increasing or decreasing values within the visible spectrum, but tend toward the same color (puple/magenta) at both opposite extremes:

respectively in correspondence with w_0 = .40μm and w_0 = .70μm. There is no property of the vectors $m_{w_0}(w)$ in H(I) that correspond to this fact.

Now, while the structure S of distal stimuli is not homeomorphic to the space of phenomenal colors C_{phen}, the physical causal properties that instantiate the projection operator can be argued to be. In the case of humans, for example, there is a homeomorphic mapping from the Munsell color cylinder (the human C_{phen}) to the cone represented by S_N. My proposal is to identify colors with the relational properties that instantiate the projector operators. This ensures that the content of color experiences satisfies the co-variation condition.[21] As we shall see, the proposal can be argued to be immune to standard objections to relationalism.

Before exposing my proposal, it is interesting to consider Hilbert's response to the objection raised above. Hilbert proposes to identify colors with specific reflectances of physical surfaces. He is well aware of the above mentioned potential objection: "[d]eterminate colors", he writes, "cannot be identified with specific reflectances because there will typically be (indefinitely) many reflectances that result in the appearance of a given determinate color, and no motivation for choosing between them." (Byrne & Hilbert, 2003, 13) Here is how Hilbert proposes to amend his theory to meet this objection:

> The solution to this problem is clear: we can identify the determinable colors with reflectance types (or sets of reflectances) rather than with the specific reflectances themselves. For example, the property purple, on this modified account, is a type of reflectance rather than a specific reflectance. As a bonus, this proposal also solves the problem of metamers (and so it is not really an additional problem): both determinable and determinate colors are reflectance-types. Metameric surfaces are, according to the revised theory, the same in

[21] It may be objected that such mapping is not complete, or that it is not a "perfect" homeomorphism. Topographically, the two structures match pretty well. They are both 3-dimensional, they agree on conflating metameric pairs, and, finally, most phenomenical similarity relations are preserved. But not all! Human phenomenal color space is metrically distorted in ways that are not always matched by the cone S3. There are some qualitative properties expressed by the Munsell color system that have no match in the triple-reflectance color space. Similarity relations along the dimensions of brightness and saturation, for example, have a different character from changes of hue from unique green to unique yellow to red to blue. Such distortion of the color cylinder have no correspondence in SN. These differences, however, are minor, and do not play any significant role in standard color perception. Those who sympathize with my proposal, will have to bite the bullet. They will have to accept that there are (few) properties of color experiences that have no correspondence in reality, although most of them do.

determinate color in spite of their physical differences (Byrne & Hilbert, 1997a; Hilbert, 1987).

The resulting account is known as “Anthropocentric Realism”. Real colors, according to this view, are spectral reflectances. Then there are anthropocentric colors, identified with *groups* of spectral reflectances. Falk talk of colors, according to this view, refer to anthropocentric colors, while what is really represented in color experiences, are real colors. Now, if the considerations exposed in section 1.6 are sound, then neither “real” nor “anthropocentric” colors could be the content of color experiences. According to the causality condition, if the instantiation of a structure $(D; S_1, S_2, ...S_n)$ is the content of veridical color experiences, it must *cause* the instantiation of structure C_{phen}. As noted in section 1.6, this condition must be taken to entail that the causal relations among the tokens of the two structures must hold *because* there exist a law that causally connects the instantiations of $(D; S_1, S_2, ...S_n)$ with the instantiations of C_{phen}.

Hilbert concedes that “the reflectance-types that we identify with the colors will be quite uninteresting from the point of view of physics or any other branch of science unconcerned with the reactions of human perceivers”. However, he continues, “[t]his fact does not […] imply that these categories are unreal or somehow subjective (Hilbert, 1987, 11)”. I agree that the fact that these properties are “uninteresting” does not entail that they are unreal. But, given our causality condition, this is not enough. If they are to constitute the content of veridical color experiences, these properties must be projectible, i.e. they must be (jointly) capable of supporting inductive reasoning, quite apart from inductions related to the response of perceivers. So, if by saying that they are “uninteresting” Hilbert means that the only inductions that these properties support are related to color perceivers, then the causality condition rules them out as candidates for the content of color experiences. More on this point later. If spectral reflectances (or classes thereby) cannot be identified with colors, however, they certainly have a lot to do with them. For example, it is unquestionable that we would not perceive any colors, if it wasn’t for them. It could be argued even that we could not even hallucinate colors, if it wasn’t for them (asymmetric dependence condition). So what’s the role of spectral reflectances in color perception? As I have already anticipated, I argue that spectral reflectances are the narrow correlates of color properties.

3.1. Teleological relationism

3.1.1. The character of veridical color experiences

Let me briefly summarize what we said about color perceptions. The distal stimuli that cause color perceptions form a structure that can be represented by $S \subseteq H(I)$, as described in section 2.1 above. The stimuli undergo two transformations.

The first formally consists of an (orthogonal) projection that "squeezes" the space of spectral stimuli into a N-dimensional subspace of *H(I)*: $H_N(I)$. The resulting structure is the structure of proximal stimuli: a convex subset S_N of $H_N(I)$. The dimensionality of $H_N(I)$ depends on the number of types of photoreceptive cells in the perceptual apparatus of the perceiver.[22]

Such projection is (formally) realized by the projection operator O, so that the space of proximal stimuli is the image of S under O: $S_N = \{O \cdot C : C \in S\}$. The causal chain that links the reflectance properties of objects to color perceptions, must therefore instantiate the projector O. This causal chain, in humans, is realized by the reflectance properties of objects and by the three types of photoreceptive cells in the retina, resulting in a 3-dimensional space of proximal stimuli.

The second transformation is realized by the brain alone, and it leads to the instantiation of a structure that we called "phenomenal color space": C_{phen}. Here is a sketch of my proposal. I propose that color properties, i.e. the properties represented by veridical color experiences, should be identified with the physical properties that instantiate the projection operator O, whatever they are. As we said in the introduction, the character of color experiences is a map from perceptual contexts to perceptual contents. Now we can say how the character map works, i.e. how it assigns contents to various contexts.

[22] I deliberately leave open this dimensionality, to allow for color experiences in creatures whose visual apparatus is different from that of normal humans.

3.1.2. The context of perceptual token-experiences

Perceptual token-contexts are constituted by (1) an individual object (or surface), (2) an environment and (3) an individual perceptual apparatus. For the purpose of the individuation of content (as specified below), token-contexts belong to non-conventional types. Such types are individuated by the formal properties of their instantiations, and do not depend on the properties of the phenomenal color space. More precisely, two token-contexts belong to the same type (equivalence class) iff they instantiate all and only the same projection operators.[23] The character of color perceptions (the map) works according to the following instructions:

Rule 1. The reflectance profile of the perceived bearer of the color (the object) is always one of the relata of the color property represented by a veridical color perception. More precisely, it is the narrow correlate of the color property.

Before the distal stimuli are processed by the perceptual apparatus, the space of reflectances is projected onto a finite-dimensional space: the space of proximal stimuli.[24] Such projection can be represented by an orthogonal projection operator, O.

Rule 2. The content of a veridical color experience, given the context (as specified in step 1), is the relational property of the colored object in virtue of which the object and the perceptual apparatus co-participate in the instantiation of the projection operator O.

Notice that not all the physical details of a given token-context are relevant for applying rule 2. What a given color property is, is insensitive to changes to the properties of a context that leave unaltered the formal properties required to instantiate the projection operator. As we said, two token-contexts belong to the same type iff they instantiate all and only the same projection operators. Rule 2, then, is only sensible to the *types* to which a given context belongs. My proposal entails that individual colors are identified with relational properties. In the case of humans, for example, each color will be identified with a physical relational property whose *relata* are at least (1) the physical object (the bearer of the color property) and (2) the retina. Should we conclude that it is part of the essence of colors to be relational properties? If

[23] The idea of generalizing the relevant normative contexts to solve the problem of faultless disagreement has been defended in Cohen, 2004.

[24] Subsequent processing consists in further transforming these stimuli so as to construct the phenomenal color space

so, why isn't this transparently part of their characters? Why, that is, did we have to look at how our world is, to figure out that colors are relational properties, when supposedly this is a consequence of the *character* of their representation, and character is the cognitively accessible semantic dimension of representation? Could it have turned out that colors are intrinsic properties of their bearers, or that they are their reflectance profiles?

Yes and no. In other counterfactually possible worlds, the answer to the last question is yes: colors could have turned out to be reflectance profiles. But under a Krepkean notion of possibility the answer is no. The character of color perceptions is whatever allows perceivers to go from perceptual contexts to perceptual contents. As it happens, this map, as we described it, is "world specific", i.e. it allows to *successfully* individuate content (if at all) only at worlds sufficiently similar to ours. Once the character of color experiences is individuated (in our world), however, it remains robustly the same at all other possible worlds, like water remains robustly identical with H_2O at all possible worlds. A consequence of this is that the same character that individuates what colors are in our world, might not be successful at individuating contents at all in worlds nomologically very different from ours.

3.1.3. The character of non-veridical color experiences

Notice that rules 1 and 2 only provides us with means for fixing the content of color representations when (and if) they are veridical. To complete the identification of the character of color perceptions, then, we need to add another rule that fixes the content of non-veridical experiences in a robust way. Intuitively, such content will be individuated by those properties that *would* instantiate the relevant projection operator, in *that* context, if the experience *were* veridical.

This, however, is highly problematic, for it threatens my account to beg the relevant question. Suppose in fact that your retina starts to dysfunction (or to function differently), so that the same tomato that appeared red to you this morning, now appears to be blue. Remind that, for our purposes, the functioning of the retina is completely captured by the "degrees" to which a given physical color C(w) stimulates each of the three receptors. So the assumption that your retina functions differently this evening effectively means that (at least) one of the three response functions (s(w), m(w) and l(w)) has changed. This would lead inevitably to three different coordinates:

$$\int_{W_{min}}^{W_{max}} C(w)s(w)dw,\ \int_{W_{min}}^{W_{max}} C(w)m(w)dw,\ \text{and}\ \int_{W_{min}}^{W_{max}} C(w)l(w)dw.$$

In sum, the 3-dimensional projection of the Hilbert space of distal stimuli will be different. It follows that also the projection operator would be different from the one you and the tomato instantiated this morning! If I don't add anything to the account, this would have the consequence that this evening the content of your experience is different from the content that your experience of the same tomato had this morning, in spite of the fact that the tomato hasn't changed at all. Worst still, the two experiences will be (necessarily) equally veridical! This is a typical drawback of relationalist accounts. I believe that my framework has the resources to tackle this problem, but I will have to make relevant concessions to ecological theories of color. The secret, I believe, is in the relation between the manifest bearers of color properties and their narrow correlates. Let us resume our discussion of narrow correlates (sec. 1.5).

Consider again the example of weights. Weights, we said, are relational properties of familiar material objects and other astronomical objects. Our representations present the familiar objects as the proper bearers of weight properties. I (usually) weigh 75 kilos. It is *I* who weigh 75 kilos: not a system that includes the earth! We noticed, however, that this is not an irreducible feature of weight. Weights, in fact, are irreducibly extrinsic properties. What happens is that our representations pick up these properties in a monadic mode, as it were. This is why we ascribe weights to people and objects, and not to astronomical systems. Now, because the property is to be causally efficacious, we expect weights to have a narrow correlate. As it happens, this is mass. I have argued that, in spite of this, our weight representations do not have the magnitude mass as their sole content. However, the mass of an object, being a narrow correlate, plays a (Krepkean-) necessary role in the individuation of the content of veridical weight perceptions. Indeed, it could be argued that mass is what interests us, in making weight judgments, although it is not the content of weight perceptions.

What to make of false (or incorrect) weight perceptions? Suppose you wanted to buy a 1 kilo beefsteak. And suppose that the shopper tricks you in the following way. When he weighs the beefsteak in front of you, he activates an elevator that accelerates upwards the whole shop. As a result, you will get less meet then you expect. Yet both the scale and your perceptions would

agree that you're in front of 1 kilo beefsteak! What to make of this? Is your perception non veridical, in the elevator? If we specify the character of weight perceptions only making references to familiar and astronomical objects, like we have done, then there is no way to say that the shopper is wrong. According to the only available notion of weight, he's absolutely right: the beefsteak weighs 1 kilo!

Yet something must be missing from our specification of the character of weights, such that, if we took it into account, we could explain why the shopper is cheating, and why the weight perception is non-veridical. I think that part of the character of weight experiences, in fact, is that their contents correlate with mass (intuitively: quantity of matter). Mind it, I said that the contents *correlate* with mass, not that they *are* masses. When we experience a given weight, I expect to be experiencing a given quantity of matter. This is why you would be surprised if you were still hungry after eating your elevator beefsteak. I think that a reference to masses should therefore be inbuilt in the character of weight experiences. A weight experience is veridical (among other things), if it gives us an optimal idea of the mass of its proper subject.

Something analogous, I believe, happens to the character of colors. We said that reflectance profiles are the narrow correlates of color properties. This explains why color perceptions exist at all. It is by latching to reflectance profiles, that color properties convey information about the physical characteristics of objects. If it is true that people's hair tends to turn grey with age, and if ripening bananas and pears tend to turn yellow, and if it is true that red striped spiders are venomous, this is because color properties latch onto reflectance profiles. This explains why color properties are projectible, to some degree, and why we expect them to be found out there in the external world, independent from our perceptions.

The adequacy of such latching, I submit, must then be inbuilt into the character of color experiences. In particular, I propose that it should be relevant in fixing the content of false (or incorrect) color experiences. But how can we do that, without concluding that nothing is really objectively colored? Notice in fact that this notion of "adequacy", is relative to token-contexts: my perception of the tomato is "adequate" only relative to the present conditions of my perceptual apparatus. So, if we in-build the notion of proper functioning into the character of color perceptions, we seem to be confronted with the following dilemma.

3.1.4. The problem of error

In saying that my apparatus "dysfunctions", or that it "doesn't perform at its best", we appear to be saying either of these things:

1. Either we are saying that it "dysfunctions" in the sense that it fails to capture the exact reflectance of the tomato, in which case ALL possible perceptual apparatuses dysfunction.
2. Or we say that it dysfunctions in the sense that it is not "performing at its best", whatever this means.

In the first case, if proper functioning is in-build into the notion of veridical color perceptions as I have suggested, we will conclude that there are no veridical color perceptions after all, i.e. that nothing is really colored in the relevant sense (eliminativism). However, if we opt for the second interpretation, we fall into the relativist horn of the dilemma. It seems that a retina can only be "performing at its best" (or fail to do so) relative to itself. In fact, if we said that a retina is not performing at its best relative to a "healthy" retina, we must be referring to the first interpretation of "dysfunction". A "healthy retina" can only be (1) a statistically typical retina, in which case the epistemically normative character of the notion is lost; or (2) a retina that optimally approximates reflectance properties of objects. But "optimally" relative to what other possible retinas? This has the absurd consequence that, under the second interpretation, no retina could possibly dysfunction. A retina, in fact, can do nothing but follow the laws of physics. How could it possibly go wrong about that? How could you blame a retina for following the laws of physics?

This is the good old problem of error. Where are we to find room for epistemic error in a world that submissively obeys to the laws of physics? In a nutshell, this is the problem. If there is a sense in which a given color perception is non veridical, there must be a sense in which, in that context, that perception *could* have been veridical. Hence there must be a sense in which, in that context, the perception *could* have been different from what it is. Nothing empirical can be false, if it *could not* have been true!

Now, these modal notions must be understood in a nomological sense. This is because, if we strip the physical details from the context of a given perceptual experience, it is not clear anymore that it is THAT perceptual experience that COULD have been true. Let me be more precise about the problem of error. Remember that each element $C \in H(I)$ from the distal

stimuli (the reflectance profiles), can be approximated by its orthogonal projection onto $H_N(I)$. This, we have seen (section 2.1), can be expressed by:

$$C(w) \approx O \cdot C(w) = C_N(w) = \sum_{n=0...N} \beta_n b_n(w)$$

How "good" is this approximation? Could a different choice of response functions make this approximation better? Is there any other function in $H_N(I)$ that approximates $C \in H(I)$ better then $O \cdot C(w) = C_N(w)$ does? The following are standard mathematical notions that will help us to answer these questions.Given any two functions $C^1(w), C^2(w) \in H(I)$, define their inner product as:

$$\left\langle C^1(w) \mid C^2(w) \right\rangle = \int_{W_{\min}}^{W_{\max}} C^1(w) C^2(w) dw$$

This allows us to define a positive definite norm for each vector $C(w) \in H(I)$ in the Hilbert space of stimuli:

$$\left\| C \right\| =_{def} \sqrt{\left[\left\langle C \mid C \right\rangle \right]}$$

With this norm we can define a "distance" between any two functions of the space. Such distance turns our space into a metric space.

$$d(C^1, C^2) =_{def} \left\| C^1 - C^2 \right\| = \sqrt{\int_{W_{\min}}^{W_{\max}} \left[C^1(w) - C^2(w) \right]^2 dw}$$

Now we can give a precise definition of what it means to say that a given projection $O \cdot C(w) = C_N(w)$ "approximates" the stimulus *C(w)*. We shall say that the projected vector $O \cdot C(w) = C_N(w)$ approximates *C(w)* to a degree of accuracy that is measured by the distance:

$$d(C(w), C_N(w)) =_{def} \left\| C(w) - C_N(w) \right\| = \sqrt{\int_{W_{\min}}^{W_{\max}} \left[C(w) - C_N(w) \right]^2 dw}$$

Now, given a Hilbert space $H(I)$ and a point (vector/function) in it, $C \in H(I)$, and given a non-empty closed convex subset, such as $S_N \subseteq H(I)$, there exists a unique point $C_N^{Best} \in S_N$ which minimizes the distance between *C* and the points in tri-stimulus space S_N:[25]

$$C_N^{Best} \in S_N \,,\ \left\| C - C_N^{Best} \right\| = d(C, S_N) = \min\left\{ d(C, C_N^i) : C_N^i \in S_N \right\}$$

The existence of vector $C_N^{Best} \in S_N$ suggests that we may use it to ground the normative character of color perceptions. We could, for example, in-build a reference to it among the features that individuate the content of color experiences (relative to a given context), along the following lines:

The content of veridical color perceptions (proposal 1)

A given perceptual token-context fixes a tri-stimulus space $S_N \subseteq H(I)$ and a projection operator O. The content of color experiences are the properties that instantiate O. The character of these experiences determines the conditions under which their content is veridical:

> *The properties that instantiate a given projection operator O_i are the content of a veridical color experience only if O_i is such that, for any possible stimulus $C \in H(I)$, the image of C under O_i, $C_{N_i}(w) = O_i \cdot C(w)$, is the best approximation of C, relative to O_i: $C_{N_i}(w) = C_{N_i}^{Best}$*

Now the problem expressed above is quite clear. Call the projection operators that the tomato and your retina instantiated this morning and this evening, respectively, $O_{morning}$ and $O_{evening}$. The same reflectance profile of the tomato has two images in the two different tristimulus spaces: $C_{N_{morning}}(w) = O_{morning} \cdot C(w)$ and $C_{N_{evening}}(w) = O_{evening} \cdot C(w)$. The condition we placed above consequently splits into the following two conditions:

$$C_{N_{morning}}(w) = C_{N_{morning}}^{Best} \text{ and } C_{N_{evening}}(w) = C_{N_{evening}}^{Best}$$

Suppose that these conditions apply to our case. They express the fact that your retina performed “at its best” both this morning and this evening. The

[25] Rudin, 1987, theorem 4.10

retina performed "at its best" relatively to what it could (nomologically) have done, given its current properties at the time of assessment. If these are the conditions for a given color representation to be true, then we will have to say that the tomato *was* red this morning and *blue* this evening. If you and I instantiated respectively $O_{morning}$ and $O_{evening}$ in front of the same tomato, moreover, according to my proposal the tomato would then be red for me and blue for you. And that's that: no possible disagreement! This is the second horn of the dilemma.

On the other hand, if we required that the projection operator be such as to capture *exactly* the reflectance properties of the tomato, that is if O is required to be an identity operator, then neither $O_{morning}$ nor $O_{evening}$ could be the content of a true color experience: hence, strictly speaking, the tomato would be neither red nor blue. This is the first horn of the dilemma.

3.2. True colors relative to the dimensionality of color space?

Another option comes to mind. Perceptual contexts, as defined above, fix the relevant color spaces (hence also the projection operators) in two ways. First, they determine the *dimensionality* of the projection. For us trichromats, for example, this dimensionality is 3. Other perceptual contexts (in non-human animals or in anormal humans), will fix color spaces and projection operators differently.

Secondly, perceptual contexts fix the detailed shape of proximal perceptual spaces. This is determined, in the case of humans, by the response functions s(w), m(w) and l(w). So far we have proposed to make optimal performance of visual experiences relative to a given triplet of response functions. This created the problem of error as explained above. Could we not have fixed the normative notion of optimal performance relative to a given dimensionality, rather then to a specific triplet of response functions? To say that the tomato is red because this is the best I could do, given the current conditions, makes color properties relative to idiosyncratic visual conditions and to the current physical properties of the retina. But what if we define red relative to the best I could do *as a trichromat*, rather then relative to the best I could do as Emiliano (my name) this morning?

Technically, this is what the proposal would look like. Given a certain dimensionality (N), there are uncountably many N-tuplets of response functions, corresponding to as many projection operators. Let us confine

ourselves to the case of trichromats, for simplicity. Given a distal stimulus $C \in H(I)$, to each triplet $s_i(w)$, $m_i(w)$ and $l_i(w)$ in the space of possible projections, there corresponds an optimal approximation vector $C_i^{Best} \in S_3$. We can partition the class of triplets into those subclasses that share the same optimal approximation vector, so that now there is a one-one correspondence between 3-D projection operators belonging to the same equivalence class, and optimal approximation vectors. The proposal sketched above is then to select the normatively relevant (equivalence class of) projection operators so as to optimize the distance from the distal stimulus C.

Let T_i be the triplet $\langle s_i(w), m_i(w) \text{ and } l_i(w) \rangle$, and let T be the class of all these triplets. $C_i^{Best} \in S_3$ is the best approximation vector relative to T_i.

Now, there exists a vector $C^{Best} \in H(I)$ such that

$$\left\| C - C^{Best} \right\| = \min \left\{ d(C, C_i^{Best}) : C_i^{Best} \in T \right\}.$$

Notice that the identity of $C^{Best} \in H(I)$, hence of the correlated projection operator O^{Best} depends solely on the *dimensionality* of the projection, and on no other idiosyncratic feature of the perceptual context. Let us call O^{Best} the "3-best projection operator" (generalizing dimensionality: the N-best operator). The amended proposal, then, would be the following:

The content of veridical color perceptions (proposal 2)

> *The properties that instantiate a given projection operator O_i are the content of a veridical color experience only if O_i is such that, for any possible stimulus $C \in H(I)$, the image of C under O_i, $C_{N_i}(w) = O_i \cdot C(w)$, is the N_i-best approximation of C.*

The projection operator that wins this game will be called the N-best operator. Analogously, we define the "N-best tri-stimulus-space" and the "N-best N-tuple of response functions". The predictable complication with this version of my account is that it is very likely that, within the same dimensionality, there will be N-tuples that "win the game" relative to certain distal stimuli, while others win it relative to other stimuli. If this is the case, as I think it is, then my restriction won't suffice to establish a relation of total order among N-tuples/spaces/operators. One needs a *total* order because the

normative character of color perceptions now hinges upon the possibility to compare N-tuples/spaces/operators with N-best N-tuples/ N-best spaces/ N-best operators.[26]

What to do? One could of course try to further restrict the condition, for example by averaging for accuracy among the various possible projectors. The epistemically relevant order relation in the space of projectors would be then established by comparison with "average best performance", rather then with "best performance" *simpliciter*. I won't pursue this solution, however, because I think there is a general serious problem with my strategy that has to be tackled first. It is possible (if not likely), that under this restriction most ordinary color perceptions would come out false. Who guarantees that that the typical human retina is averagely the N-Best receptive apparatus? Indeed, one may have reasonable doubts about whether the typical human retina is an N-best retina (averagely or not averagely) with respect to any stimulus.

So eliminativists will probably sympathize with this relationalist proposal. Nothing, or nearly nothing, would be colored, in our world, if this theory of color were correct. Here goes my last proposal, as far as this paper is concerned. Essentially, the idea is to replace the restriction exposed in this section with a *teleological* restriction, thus weakening it substantially.

3.3. Teleological relationalism

My favorite version of the relationalist account advocated in this paper, is a teleological version. It is "teleological" because the epistemologically normative ingredient is a naturalized notion of purpose, or function, rather then the technical notion of best performance introduced above.

In a nutshell, this is the proposal:

The content of color perceptions (teleological proposal)

> *The content of color experiences are the projections that would have had to have been instantiated, had their respective perceptual systems instantiated that experience when* functioning properly.

What makes this proposal "teleological" is the fact that the epistemologically normative ingredient is a naturalized notion of purpose, or

[26] The instantiation of a projection operator, under this proposal, is the content of a veridical N-dimensional color experience only if it is an N-best operator.

function. The notion of "functioning", for example, could be borrowed from biology:

> The concept of a biological function is defined in terms of natural selection (Wright [[92]], Neander [[58]]) along the following lines: it is the function of biological system S in members of species Sp to F iff S was selected by natural selection because it Fs. S was selected by natural selection because it Fs just in case S would not have been present (to the extent it is) among members of Sp had it not increased fitness (i.e. the capacity to produce progeny) in the ancestors of members of Sp. [27]

We are not forced, however, to adopt this particular reductionist strategy. In fact, maybe we shouldn't. As etiological accounts of function cannot be cashed out in terms of the present state of the instantiating system, some might worry (with reason, I think) that these are causally epiphenomenal, i.e., causally inert. Remind that we are after *real* causal constraints on representations, so if this difficulty cannot be amended, this fact could threaten our proposal. Some authors suggest that biological function could be cashed out in non-etiological and non-teleological terms. Here it suffices to say that, while teleological functions are often considered as selected effects, they can also be considered as selected *dispositions*: certain traits are selected because they produce certain effects in response to certain causes.[28] Moreover, there is hope that one could define *proper functions* in non-etiological, and non-biological terms.[29] What's good about teleological solutions, is that they can be adapted to various theories of content to block the problem of error.

> An appeal to teleological functions can be combined with various ideas to form hybrid theories. [...] it's worth mentioning that such an appeal can also be combined with isomorphism theories (e.g. Cummings 1996). If we combine the idea that representations are isomorphic with their representeds with idea that psychosemantic norms depends on the norms of proper functioning, we can generate several proposals: for example, the proposal that the relevant mappings are those that the systems were designed to exploit... [30]

Here I do not wish to argue in favor or against of any particular teleological theory of content. For our purposes, what counts is that if any of these theories proves to be sound, it would allow us to induce an externalist

[27] Loewe, 127.
[28] This, of course, does not make teleological functions a set of current dispositions, but a set of selected dispositions.
[29] See for example Bickhard, 1991.
[30] Neander, 2004.

restriction to the contents of color perceptions, thus bypassing the relationalist dilemma. Notice that this restriction is weaker then that imposed by requiring that the contents of veridical perceptions be N-Best operators. A given projection can be the one *that would have had to have been instantiated, had the respective perceptual systems instantiated that experience when functioning properly* (I bet you can't say it without breathing), even if it is not an N-Best operator. Natural selection is very clever at designing solutions, but it is not perfect!

This proposal, however, is similar to the previous one in a relevant respect. Under any understanding of proper functioning, a perceptual apparatus functions properly only if it exploits all (and only) the photoreceptors that natural selection has designed for it. So, implicitely, this proposal also makes color perceptions true relative to the dimensionality of phenomenal color space. In fact, it does more: it makes them relative to specific *kinds* of perceptual systems.[31]

4. It's a quasi-colorful world

According to teleological relationalism, there will be as many classes of color properties as there are *kinds* of perceptual systems. So much the worse, I say, for the intuition that we and the bees, for example, represent (exactly) the same properties of a flower when we're looking at one. One advantage of my account, I think, is this. As I shall argue, although two different creatures might be representing different properties of a flower, when looking at it, it is still possible to say that these properties belong to a common natural kind. This is exactly as it should be, if the different properties in question are to deserve the name of colors.

Colors, according to teleological relationism, are relational properties of physical objects and perceptual apparatuses. These properties are represented in such a way that their proper bearers, relative to these representations, are the physical objects. Empirical discoveries allow us to say that color properties have a narrow correlate (reflectance profiles). Very different color representations (for example harbored by very different creatures), may represent different color properties of the same physical object, under the same environmental conditions.

[31] Not necessarily, as I have said, these kinds must be biological kinds.

Very likely, the subjective experience of these representations will also be very different. This is, I think, what should be expected. Any realist theory of colors also has the consequence that there are color properties that we humans cannot represent. Therefore, I don't take this to be a peculiar drawback of my account. Nor I think that this is a serious objection for anyone.

Up to a certain extent, whether a certain property is a *color* property, is a terminological issue. In my account, what all color properties have in common, is the *character* of their representations. My representation of a flower and a bee's representation of the same flower share the same character. This is to say that both I and the bee fix the content of color representations in the same way. Character, remember, is the map from context to content. My context is different from that of a bee's, whence the fact that we represent different properties. The content of my representation and that of the bee also (necessarily) share a common relatum: the colored object. Moreover, the two perceptions represent color properties that also have their narrow correlates in common: the reflectance profiles. All color properties (regardless of their class of provenience), are arranged in a metrical space that allows us to say which color property is the more accurate approximation to its correspondent narrow correlate.

Finally, most cases of disagreement that one may want to accommodate, e.g. the case of the red vs/ blue tomato of our example, can be easily accommodated by teleological relationalism. If the tomato looks blue to me, then my retina is not functioning properly, hence the content of my experience is (robustly) that the tomato is blue, while in fact it is red. This is as much room for error as teleological relationalism can afford. I think it is enough room. This is as far as the substantial, non-terminological dispute can go, I believe. Whether we want to call properties and perceptions that have that much in common "colors", I submit, is now a terminological issue.

4.1. So, is the world colored?

Suppose I'm right about what properties we represent when we have color experiences. What should we make of the claim that the world is objectively colored? Retinas necessarily contribute to instantiating color properties. Just as one can pick up color properties so that physical objects are their proper bearers (our representations do), one can also pick them up so that retinas

are the proper bearers. I am doing it now while writing, and you are doing it while reading these words. Doesn't this make color properties mind-dependent? Doesn't it violate the externality condition?

I think not. It is important to distinguish mere environmental differentiators from mental representations. Any metal bar, for example, implicitly categorizes environments that have the same temperature, because its length co-varies with temperature in a lawful way. However, we would not say that any metal bar *represents* the temperature of the environment. I will say that metal bars are environmental differentiators. Whether a metal bar also represents temperature, depends on whether an organism (or cognitive system) *uses it* to represent temperature. It is undeniable that certain metals are particularly apt to be so used. This is why we can build thermometers exploiting this property. However, thermometers only represent temperatures relative to our using them as representations. A thermometer, in and of itself, is a mere environmental differentiator.

While being a representation is certainly a "mental property", being an environmental differentiator is not. Now, according to my proposal, retinas (and similar perceptual apparatuses) are necessarily among the relata of color properties. Seen "from the side of the retina", so to speak, color properties are properties of the retinas. The narrow correlate of color properties, when these are viewed "from the side of the retina", are all those physical properties in virtue of which retinas act as environmental differentiators. As I noted above, however, these properties are not mental properties. Plausibly, a property is mind-dependent only if it is necessarily co-instantiated with some mental properties (whatever these are). The instantiations of the projector operators in no ways entail the (co-)instantiation of mental properties. In fact, we have seen, they only entail the (co-)instantiation of *environmental differentiators.* It follows that color properties, under my account, are not mind-dependent. Having said that, we can conclude that if I'm right, the world is indeed a colorful place, for color perceptions are often veridical!

Yet someone might still be perplexed at this solution. "All right", my detractor could concede, "the properties you call colors are not mind-dependent, but they are certainly different from the brain-independent properties that we were expecting!" I have already noted that the idea that objects should be the proper subjects of color ascriptions is due to the particular mode of presentation of color properties in our phenomenal world. We have seen how the ultimately extrinsic or intrinsic nature of color

properties is an empirical question, and not one that could be accessible to phenomenological introspection.This should be enough to dispel the impression that my account entails some form of eliminativism. However, I think that one can say more to diffuse this worry.

5. Conclusions

Color properties, according to the view put forward here, are objective properties that we use to gather information about distal stimuli. The properties that we represent in color perceptions (i.e. colors) are very similar to their narrow correlates (i.e. reflectances). Such similarity, we have seen, can be measured. The similarity explains why our color perceptions can be used in (approximately) sound inductive reasoning about properties of the objects that are not themselves relative to retinas. Being a ripe banana, or a venomous spider, for example, are certainly not properties that depend in any way on our retinas, let alone on some mental properties. What explains our capacity to infer retina-independent properties of bananas and spiders, I submit, is the (measurable) degree of similarity between the contents of our perceptions and their narrow correlates.

Summing up, colors are not basic properties of the world (see the definition of basic property in section 1.5), but they are extremely close to some basic properties of the world. Those readers who insist that the world can only be said to be *really* colored if colors are basic properties of objects, will have to content themselves with the claim that the world is quasi-colored: colors are quasi-basic properties. I have argued on a number of grounds that the properties that we represent in our color experiences should best be thought of as relational properties of physical objects and perceptual apparatuses. In particular, I have argued that color properties are those that instantiate the operators that projects the infinite-dimensional space of spectral reflectances onto the finite color spaces that organisms perceive.

Colors, under this account, are objective, mind-independent properties of the world. Teleological relationalism, that is, allows us to claim that the world is populated by objectively colored objects[32], and that most of our color perceptions are veridical. The account has been shown to be immune from

[32] As I said, those particularly picky about real colors being basic properties (see section 1.5), will have to content themselves with saying that the world is populated by quasi-colored objects.

standard objections to relationalism. In particular, it has been argued to resist standard faultless disagreement counterarguments.

References

Bickhard, M. H., 1991, How to build a machine with emergent representational content, *CogSci News* 4(1), 1–8.

Campbell, J., 1994, A Simple View of Color, in Haldane, John, and Wright, Crispin (eds.) (1994), *Reality, Representation and Projection*, Oxford: Clarendon Press: 257-69.

Cohen, J., 2004, Color Properties and Color Ascriptions: a Relationalist Manifesto, Philosophical Review, 113(4):451-506

Dretske, F., 1995, *Naturalizing the mind*, Mit Press, Cambridge, Ma.

Hardin, C. L., 1988/1993, *Color for Philosophers*, Indianapolis, Ind.: Hackett.
--- 2003, A Reflectance Doth Not a Color Make, *The Journal of Philosophy*, 100: 191-202.

Hilbert, D. R., 1987, *Color and Color Perception*, Stanford, Calif.: C.S.L.I.

Byrne, A. and Hilbert D., 2003, Color Realism and Color Science, *Behavioural and Brain Sciences*, 26: 3-21.

Millikan, R., 1990, *Biosemantics, Mind and Cognition* (W. G. Lycan, ed.), Blackwell: 221–230.

Rudin, W., 1987, *Real and Complex Analysis*, McGraw-Hill

Dossier

Introspection and Intuition in Mathematics

Introduction
Where experience matters

Alexandra Van-Quynh
(CFCUL)
aquynh@cii.fc.ul.pt

The present dossier gathers the proceedings of the conference organized by the Centre for Philosophy of Science of the University of Lisbon, "*Mathematics and Intuition: Epistemology and Experience*", that took place in Lisbon on September 25th and 26th, 2012. This conference aimed to give the proper place to experience in the process of mathematical concept construction. A significant part of the seminars was devoted to discussions on introspection. The opportunity was also given to mathematicians to speak about their own practice and to share thoughts on the role of intuition (some would speak rather about creativity) in the development of mathematics.

When Matthieu Haumesser discusses the possibility of the experience in light of Kant's transcendental philosophy, he proposes an enlarged reading of Kant's *Critic of Pure Reason* in which the empirical is not to be considered as inferior to the transcendental, and suggests that a key feature of Kant's philosophy resides in an irreducible *va-et-vient* between the *a priori* and the empirical.

Michel Bitbol and Claire Petitmengin discuss then the possibility and reality of introspection, as in recent years a strong movement of renewal and redefinition of introspection has been witnessed. The authors raise several questions of epistemological relevance about this renewal. They show that the conditions for a successful study of first-person experience are now fulfilled by the use of the method of the *elicitation interview* developed by Pierre Vermersch. Recalling Kant's redefinition of objectivity (objectivity is not something to be found ready-made *out there*, but a project of operational extraction of invariant structures out of a cluster of appearances), the authors

insist on a procedure for introspection, mediated by the interview of elicitation, employed as a descent and ascent investigating method that leads to generic structures of intersubjective value beyond individual reports.

The method of *interview of elicitation* is introduced by Maryse Maurel. Starting from the heritage of Husserl and the pre-reflexive consciousness coined by the philosopher, the author details how, following the insights given by Husserl in his model of passivity and by more recent works, the subject – expertly guided – can access the pre-reflexive gestures of an action he or she made.

The word is then given to two mathematicians, Pedro J. Freitas and António Machiavelo who, as Gian-Carlo Rota and William Thurston did a few decades earlier, give us a phenomenological viewpoint on mathematical practice. Each of them discuss the role and the origin of intuition in the development of mathematics and, despite the authors' claim of not being themselves philosophers, the reader will appreciate how full of philosophical insights those two texts are.

Les possibilités de l'expérience. Mathématiques, aperception pure et aperception empirique dans la *Critique de la raison pure* de Kant

Matthieu Haumesser
(Lycée A. Kastler, Cergy-Pontoise - France)
haumesser@free.fr

Dans le cadre d'une réflexion générale sur l'introspection, je me propose d'interroger ici le rapport qui, dans la philosophie de Kant, peut être établi entre deux caractérisations essentielles et complémentaires de la faculté humaine de représentation : l'aperception pure et l'aperception empirique.

L'aperception en général, c'est l'acte d'apercevoir ou de s'apercevoir, c'est l'acte condensé dans la formule « je pense » héritée du fameux *cogito* de Descartes. C'est un acte qui véhicule donc une essentielle réflexivité, par rapport à la simple perception. Et c'est aussi sur cet acte que se fonde toute une conception moderne de l'expérience, depuis Descartes précisément, au sens où c'est dans cette activité du « je pense », dans la mise en examen des préjugés ou dans la mise en équation des lois de la nature, que littéralement l'expérience se *constitue*. Bref, l'aperception est ce par quoi le sujet se trouve au centre du monde. Mais que signifie alors la distinction opérée par Kant entre aperception pure et aperception empirique ?

L'aperception empirique est un fait d'expérience, et comme telle elle est soumise à une radicale contingence. Il se trouve qu'à tel ou tel moment, je pense à ceci ou à cela, notamment à partir des sollicitations des sens (couleurs, sons, odeurs, plaisir, douleur, etc.), et plus généralement pour des raisons qui bien souvent m'échappent et qui ne semblent obéir à aucun nécessité. Cette aperception empirique fonctionne largement par des voies associatives – comme dans la rêverie par exemple. C'est d'abord ainsi que Kant comprend le concept central de « synthèse » dans la *Critique de la*

raison pure : l'opération qui consiste pour la pensée à « parcourir le divers des représentations », avec toute l'indétermination que ce divers peut impliquer[1]. Par exemple, en considérant un arbre, je prêterai attention à ses feuilles, puis aux branches, puis au tronc, en suivant le parcours de mes sensations. Mais évidemment, ces représentations peuvent éveiller dans mon attention d'autres représentations plus détachées de mon expérience immédiate et de la sensation : cet arbre m'en rappellera d'autres, ou il suscitera en moi des émotions plus ou moins définissables, etc. C'est pourquoi la faculté qui pour Kant est à l'œuvre dans la synthèse du divers est « l'imagination », à laquelle il prête toujours deux fonctions complémentaires : opérer la synthèse du divers des représentations (passer d'une représentation à l'une des multiples représentations virtuelles qui pourraient la suivre) et représenter, au-delà de la sensation immédiate, des objets absents[2]. Enfin, l'aperception empirique est aussi déterminée évidemment par l'activité de jugement, orientée par la recherche de règles pour ordonner l'expérience y compris lorsqu'il s'agit de simples préjugés.

Tout cela fait de l'aperception empirique une activité non seulement contingente, mais aussi largement mélangée, qui charrie tout ensemble des concepts, des intuitions, des représentations associatives de l'imagination, des émotions, etc. Les commentateurs de Kant négligent parfois l'importance pourtant décisive de cet entrelacs de multiples éléments qui est constitutif de l'expérience en tant que telle et plus précisément de cette aperception empirique dans laquelle ils trouvent à venir s'articuler.

*

A partir de là, on comprend tout de même ce qui conduit Kant, dans la *Critique de la raison pure*, à dégager cette autre forme d'aperception qu'il qualifie comme « pure »[3]. Il faut d'abord entendre par là : pure de tout élément empirique ou contingent. Mais cela signifie aussi : pure en comparaison de tout ce que l'aperception empirique a de mélangé. Avec l'aperception pure, Kant entend dégager une activité originaire de l'entendement qui précède l'expérience, et ce faisant qui ne se laisse pas

[1] Critique de la raison pure (dorénavant : CRP), B 104. Dorénavant, nous indiquerons, dans le corps du texte, la pagination de la seconde édition originale de 1787 (notée B), reproduite dans l'édition de la Preussischen Akademie der Wissenschaften (29 tomes , Berlin, G. Reimer, 1902-1983).
[2] CRP, B 151.
[3] CRP, B 131.

déterminer par ses éléments contingents, mais qui est capable de la constituer *a priori*, en la soumettant à des règles nécessaires et universelles issues directement de la pensée et anticipées dans les phénomènes : ce sont notamment les lois de la causalité, ou encore celles des mathématiques, auxquelles le réel doit se soumettre *a priori*. L'aperception pure ainsi comprise constitue l'élément d'identité qui unifie toute l'expérience :

> Le : *je pense* doit nécessairement *pouvoir* accompagner toutes mes représentations ; car, si tel n'était pas le cas, quelque chose serait représenté en moi qui ne pourrait aucunement être pensé – ce qui équivaut à dire que la représentation ou bien serait impossible, ou bien ne serait rien pour moi. La représentation qui peut être donnée avant toute pensée s'appelle intuition. Donc, tout le divers de l'intuition entretient une relation au : *je pense*, dans le même sujet où ce divers se rencontre. Mais cette représentation est un acte de la *spontanéité*, c'est-à-dire qu'elle ne peut pas être considérée comme appartenant à la sensibilité. Je l'appelle l'*aperception pure* pour la distinguer de l'aperception empirique, ou encore l'*aperception originaire*, parce qu'elle est cette conscience de soi qui, en produisant la représentation : *je pense*, laquelle doit pouvoir accompagner toutes les autres et est une et identique dans toute conscience, ne peut être accompagnée d'aucune autre.[4]

Kant distingue ici soigneusement l'aperception pure de l'aperception empirique : alors que celle-ci est un fait d'expérience, celle-là précède toute expérience et la rend possible, par la recherche, en toute représentation et dans la synthèse du divers, de la forme logique de l'universalité. Cela se comprend d'abord dans le cadre de l'expérimentation scientifique : ainsi de Galilée qui faisait rouler des boules sur des plans inclinés en faisant varier la pesanteur et en anticipant les résultats de ces expériences ; ou encore, de Toricelli qui « fit supporter à l'air un poids qu'il avait d'avance conçu comme égal à celui d'une colonne d'eau connue de lui ». L'un comme l'autre avaient compris « que la raison ne voit que ce qu'elle produit elle-même selon son propre plan, qu'elle devrait prendre les devants avec les principes qui régissent ses jugements d'après des lois constantes et forcer la nature à répondre à ses questions »[5]. Cela vaut à plus forte raison des mathématiques : ici l'aperception pure trouve un terrain d'exercice particulièrement fécond dans la mesure où elle anticipe des règles dans une intuition qui est elle-même *a priori* : celle de l'espace et du temps, dans les schèmes de la géométrie ou du nombre. Mais cette pure spontanéité de l'aperception vaut aussi pour un travail apparemment plus trivial de

[4] CRP, B 131-132.
[5] CRP, BXIII (préface de la seconde édition).

conceptualisation : par exemple,

> je vois un pin, un saule et un tilleul. En comparant tout d'abord ces objets entre eux, je remarque qu'ils diffèrent les uns des autres au point de vue du tronc, des branches, des feuilles, etc. ; mais si ensuite je réfléchis uniquement à ce qu'ils ont de commun entre eux, le tronc, les branches et les feuilles mêmes, et si je fais abstraction de leur taille, de leur configuration, etc., j'obtiens un concept d'arbre.[6]

Cet exemple de formation d'un concept peut paraître un peu étrange, dans la mesure où le travail de comparaison, de réflexion et d'abstraction ici décrit ne semble jamais avoir lieu tel quel, tant nous savons déjà ce qu'est un arbre lorsque nous en rencontrons un. C'est d'ailleurs ce que suggère le texte : d'emblée nous comparons les arbres « au point de vue du tronc, des branches, des feuilles ». C'est donc que nous disposons déjà, au moins implicitement, du concept recherché, et que nous l'anticipons dans les phénomènes. Cette anticipation de l'universalité qui précède toute expérience et par laquelle le « je pense » projette son identité dans le divers de représentations est, en son fond, l'aperception pure. L'étrangeté de ce texte vient de ce que Kant en rend compte du point de vue de l'aperception empirique, alors que l'activité de l'aperception pure est à la fois plus fondamentale et plus souterraine.

*

Pour Kant, les formes logiques du jugement et les formes de l'intuition (l'espace et le temps), dans la mesure où elles précèdent l'expérience et conditionnent *a priori* toute aperception empirique, sont le corrélat de l'aperception pure. Des formes de l'intuition, Kant dit qu'elles se « tiennent prêtes *a priori* dans l'esprit »[7]. Cette disponibilité est liée à leur statut de conditions de possibilité de l'expérience : elle signifie que toute aperception empirique effective est sous-tendue par un ensemble de possibilités inscrites originairement dans la texture de ces formes, et que l'entendement peut explorer dans les jugements synthétiques *a priori*. C'est ce qui se passe notamment dans les mathématiques, dans la géométrie évidemment, mais aussi dans les opérations arithmétiques les plus élémentaires. Kant donne ainsi le célèbre exemple de la somme de 7 et de 5 :

> je prends d'abord le nombre 7, et en me servant, pour le concept de 5, des

[6] *Logique*, AK IX, 94, tr. fr. J. Vuillemin, Paris, Vrin, pp. 102-103.
[7] CRP, B 34.

> doigts de ma main comme d'une intuition, j'ajoute alors, à la faveur de cette image que j'en ai, peu à peu au nombre 7 les unités qu'auparavant je prenais ensemble pour constituer le nombre 5, et je vois ainsi surgir le nombre 12. C'est dire que la proposition arithmétique est toujours synthétique. (CRP, B 15-16)

L'enjeu de cet exemple est d'abord de montrer que ce n'est pas dans les concepts de 5 et de 7 que l'entendement peut trouver, par simple analyse logique, le nombre 12 : il faut passer par l'intuition. Cet aspect a évidemment fait l'objet de nombreuses discussions. Mais, au-delà du rapport ainsi établi entre concept et intuition, on peut aussi interroger le rôle ici donné à l'aperception. De manière assez étonnante, Kant décrit ici ces opérations arithmétiques du point de vue de l'aperception empirique. C'est de ce point de vue que « je vois surgir » le nombre 12, et cela suppose même de donner à mon intuition un support matériel, en comptant sur mes doigts. Mais bien évidemment, la nécessité et l'universalité de ce résultat se jouent au niveau de l'*a priori*, de l'aperception pure qui conditionne la possibilité même de cette expérience. Du point de vue de l'aperception empirique, cette opération de l'aperception pure ne peut effectivement être vécue que comme le surgissement d'une vérité qui, ouvrant la possibilité même de l'expérience, doit être considérée comme ayant déjà été toujours là, au-moins de façon sous-jacente.

*****

Le fait que Kant ne puisse décrire ainsi les opérations de l'aperception pure qu'en adoptant le point de vue de l'aperception empirique peut cependant mener à une autre ligne de questionnement. Pourquoi en effet ce « je pense » originaire autour duquel se constitue toute connaissance (en particulier celle de la logique et des mathématiques), et que Kant situe rigoureusement au niveau transcendantal de *l'a priori,* est-il si facilement rapproché d'opérations apparemment triviales de l'attention empirique ?

Il y a là un problème traditionnellement négligé par les commentaires. Ceux-ci partent souvent d'un présupposé qui est loin d'aller de soi : une certaine dévalorisation plus ou moins subreptice de l'expérience et des « faits », par rapport à l'*a priori* et au « droit ». L'*a priori* peut en effet être considéré comme le soubassement légal, universel et nécessaire de notre connaissance, rendu possible par ces éléments purs de toute expérience que sont les concepts de l'entendement, les formes logiques du jugement, ou encore l'espace et le temps. La géométrie, par exemple, tirerait sa valeur de

ne rien devoir à des expériences, mais de procéder simplement par « construction de concepts » dans une intuition pure. Les mathématiques seraient ainsi l'élément le plus central d'une valorisation de la connaissance pure et *a priori*, par rapport à une connaissance empirique plus instable et contingente.

Cependant, une telle lecture peut conduire à sous-estimer la thèse, inlassablement répétée dans la *Critique de la raison pure*, selon laquelle l'*a priori* n'a de véritable signification pour la connaissance que dans la mesure où il *rend possible l'expérience*. Et cela vaut aussi pour les concepts des mathématiques, comme le nombre, qui ne peut trouver son « sens » qu'en étant appliqué en fin de compte à des objets empiriques : les « doigts », les « grains de la tablette à calculer », ou des traits que l'on peut avoir « devant les yeux »[8]. Ces affirmations apparemment un peu simplistes et déroutantes, concernant la connaissance la plus pure qui soit, peuvent inviter à mieux considérer la place et le rôle de l'*a priori* par rapport à ce qui reste le véritable point d'ancrage de la réflexion transcendantale et son seul horizon : l'expérience.

Il faut, dans cette perspective, prêter attention au détail de la formule par laquelle Kant rend compte de la constitution de l'expérience autour de l'aperception pure : « le 'je pense' *doit* nécessairement *pouvoir* accompagner toutes mes représentations ». En toute rigueur, l'aperception pure est définie comme une aperception *potentielle.* C'est aussi en ce sens qu'elle est une condition de *possibilité* de l'expérience. Mais cela signifie qu'elle n'a de sens que de se réaliser en une aperception effective, qui sera nécessairement empirique. C'est ce que Kant dit rigoureusement plus loin dans la *Critique* : « sans quelque représentation empirique, qui fournit la matière à la pensée, l'*acte* 'je pense' ne pourrait pas du tout avoir lieu »[9]. Il faut donc considérer l'aperception pure comme un 'je pense' potentiel ; seule l'aperception empirique, liée à la sensation et au mélange des éléments dont nous avons rendu compte plus haut, est un 'je pense' effectif ou, mieux encore, en acte.

Le 'je pense' qui « doit pouvoir » accompagner les représentations est, pour la même raison, un 'je pense' *virtuel.* C'est pourquoi il opère essentiellement au niveau des simples formes (logiques ou sensibles) des représentations. C'est aussi pourquoi Kant l'associe si étroitement au travail de l'imagination. On peut interpréter ainsi la fameuse distinction – que Kant,

[8] CRP, B 299.
[9] CRP, B 422, note.

certes, n'explicite pas clairement – entre « forme de l'intuition » et « intuition formelle », notamment au sujet de l'espace[10]. Dans la mesure où l'espace se prête, indépendamment de toute sensation et abstraction faite de toute expérience sensible effective, à un travail de l'entendement et donc à une conceptualisation qui donne lieu aux jugements synthétiques *a priori* de la géométrie, il est une « intuition formelle ». Mais cela veut sans doute dire : une intuition *simplement formelle*, sans matière et donc sans objet, liée comme telle à l'imagination, et dont le statut reste simplement virtuel. Comme « forme de l'intuition » au contraire, l'espace doit être compris comme la forme qui rend possibles des intuitions effectives, dans lesquelles intervient la sensation et avec elle le rapport à des objets – ce qui ne peut avoir lieu, cette fois, que dans l'aperception empirique.

Pour Kant, il est absolument crucial de maintenir fermement ce lien entre aperception pure et aperception empirique. D'un côté, parce que l'expérience doit être fondée sur la nécessité et l'universalité des jugements synthétiques *a priori* ; mais d'un autre côté, parce que c'est seulement dans une aperception empirique effective que viennent s'articuler en fin de compte les différentes facultés engagées dans la connaissance, en même temps qu'elles y passent de la puissance à l'acte. La pensée est ainsi constamment tiraillé entre l'idéal de l'aperception pure, fondement virtuel de la légalité de l'expérience, et la réalité de l'aperception empirique. L'oscillation entre ces deux sens de l'aperception est selon nous aussi décisive que l'oscillation entre concept et intuition. C'est en général en adoptant ce dernier point de vue que l'on interprète l'« intuitionnisme » kantien, notamment s'agissant des mathématiques. Mais Kant va au-delà de la nécessité de sortir des concepts pour opérer une synthèse dans l'intuition ; car même en tant qu'elles combinent concepts et intuition pure, les mathématiques manquent encore de la *réalité* qui ne peut se rencontrer que dans une intuition empirique :

> L'objet ne peut être donné à un concept que dans l'intuition et, même lorsqu'une intuition précédant l'objet est possible *a priori*, elle ne peut pourtant recevoir son objet, et par suite la validité objective, que par l'intuition empirique, dont elle est la simple forme. Tous les concepts et, avec eux, tous les principes, quelque possibles *a priori* qu'ils soient, se rapportent cependant à des intuitions empiriques, c'est-à-dire à des données pour une expérience possible. Sans cela, ils n'ont absolument aucune validité objective, mais sont plutôt un simple jeu, que ce soit de l'imagination ou de l'entendement, relativement à leurs

[10] Sur ce point, on lira avec profit l'article de M. Fichant, « Espace esthétique et espace géométrique chez Kant », Revue de métaphysique et de morale, octobre-décembre 1999, pp. 525-538.

> représentations. Que l'on prenne pour exemple les concepts de la mathématique, et plus précisément d'abord dans leurs intuitions pures. L'espace a trois dimensions, entre deux points il ne peut y avoir qu'une ligne droite, etc. Bien que tous ces principes, et la représentation de l'objet, dont cette science s'occupe, soient produits entièrement *a priori* dans l'esprit, ils ne signifieraient pourtant rien, si nous ne pouvions pas toujours présenter leur signification dans des phénomènes (dans des objets empiriques). Bien que tous [les principes de la mathématique pure] et la représentation de l'objet auquel cette science a affaire soient produits entièrement *a priori* dans l'esprit, ils ne signifieraient pourtant rien, si nous ne pouvions toujours présenter leur signification dans les phénomènes (dans des objets *empiriques*)[11].

Ce texte permet de comprendre pourquoi Kant, à chaque fois qu'il donne des exemples d'intuitions mathématiques, non seulement donne des exemples simples, voire simplistes, mais prend toujours soin d'insister, s'agissant de la mise en œuvre de ces opérations élémentaires, sur la nécessité qu'y intervienne un support matériel empirique : les doigts de la mains par exemple. Sans cet ancrage dans l'intuition empirique, les mathématiques resteraient un simple jeu de l'imagination : il y a là un débat qui va au-delà de la question de savoir comment les mathématiques se font (par intuitions et/ou par concepts) ; ce qui est en jeu ici, c'est le lien entre les virtualités mathématiques, produites par le jeu de l'imagination et de l'aperception pure, et l'aperception empirique, qui seule garantit l'ancrage de la pensée dans une réalité effective.

*

En quel sens alors faut-il comprendre cette exigence de réalité ? D'abord, elle correspond à la nécessité de donner à la pensée de véritables objets, au-delà d'un jeu simplement formel sur les représentations ; or aucun objet n'est encore donné par de simples concepts ou dans la simple intuition pure (qui n'est que formelle). Mais ensuite, du point de vue du sujet lui-même, l'ancrage de la pensée dans l'aperception empirique est aussi ce dans quoi s'atteste le travail effectif des facultés de l'esprit, dans sa globalité (sentiments de plaisir et de peine, sensations, désirs, etc.).

En tant qu'il fait valoir cette exigence, Kant pourrait presque ressembler à un empiriste. Et, de fait, il est en dialogue étroit, sur ce point avec la pensée de Locke. Dans l'*Essai sur l'entendement humain* (1790), celui-ci avait placé au cœur de sa démarche une définition de l'existence des idées fondée sur

[11] CRP, B 299.

leur perception effective et consciente dans l'entendement, bref, sur ce que Kant appellerait l'aperception empirique :

> For if these words "to be in the understanding" have any propriety, they signify to be understood. So that to be in the understanding, and not to be understood; to be in the mind and never to be perceived, is all one as to say anything is and is not in the mind or understanding.[12]

Kant, pour sa part, ne peut évidemment admettre une telle réduction de la pensée à ce qui est effectivement conscient ; à ses yeux la pensée consiste aussi, comme nous l'avons vu, en des représentations potentiellement conscientes. Il s'oppose clairement à Locke sur ce point dans l'*Anthropologie* :

> Avoir des représentations sans pour autant en être conscient, cela semble contenir une contradiction. Car comment pouvons-nous savoir que nous les avons si nous n'en sommes pas conscients ? Cette objection, *Locke la faisait déjà* et, pour cette raison, refusait l'existence même d'une telle sorte de représentations. Et pourtant, il se trouve que nous pouvons posséder une conscience médiate d'une représentation, bien que nous n'en soyons pas immédiatement conscients.[13]

La fin de ce texte correspond au partage entre aperception empirique et aperception pure que nous avons dégagé : au-delà ou en deçà du 'je pense' effectif, il y a aussi nécessairement ce 'je pense' potentiel qui « doit pouvoir » accompagner mes représentations.

Cela étant, il n'en reste pas moins que c'est dans l'aperception empirique que se réalise en fin de compte le travail des facultés. Or de ce point de vue, la position de Locke conserve une force considérable, dans la mesure où elle impose de ramener résolument toutes les idées, même les plus abstraites, à la façon dont elles se forment dans une perception effective, et par là, à l'exercice le plus concret des facultés intellectuelles. Or cela implique une dualité irréductible dans la considération de nos idées, et plus particulièrement des idées mathématiques. En effet, celles-ci, organisées déductivement comme elles doivent l'être, et correspondant à des idéalités valables universellement, ne sont jamais réductibles aux événements psychologiques par lesquels elles se manifestent concrètement à l'esprit, ou aux supports matériels (symboles, figures, courbes, etc.) grâce auxquels on les considère ; mais elles doivent cependant être pensées en relation avec

[12] *An Essay Concerning Human Understanding*, livre I, ch.1, §5 (dorénavant : 1.1.5), cité d'après l'édition de P. H. Nidditch, New York, Oxford UP, 1975.
[13] *Anthropologie d'un point de vue pragmatique*, AK VII 135.

eux. Cette dualité, qui caractérise le rapport entre idées et réalité, apparaît bien dans les deux textes suivants :

> […] we make take notice that universal propositions of whose truth or falsehood we can have certain knowledge *concern not existence* : and further, that all particular affirmations or negations that would not be certain if they were made general, are only concerning existence ; they declaring only *the accidental union or separation of ideas in things existing*, which, in their abstract natures, have no known necessary union or repugnancy. […] All the discourses of the mathematicians about the squaring of a circle, conic sections, or any other part of mathematics, concern not the existence of any of those figures : but their demonstrations, which depend on their ideas, are the same, whether there be any square or circle existing in the world or no. (*Essay*, 4.9.1)

> Every man's reasoning and knowledge is only about the ideas existing in his own mind ; which are truly, every one of them, particular existences : and our knowledge and reason about other things is only as they correspond with those particular ideas. So that the perception of the agreement or disagreement of our particular ideas is the whole and utmost of all our knowledge. (*Essay*, 4.17.8)

Considérées dans leur idéalité, les propositions mathématiques sont universelles. Mais cela les coupe de toute référence à l'existence réelle. A l'inverse, si on les considère en tant qu'elles sont perçues dans l'esprit, alors elles deviennent des « existences particulières ». Locke appelle « connaissance intuitive » la perception effective de l'accord entre deux idées. Empiriquement, toute connaissance, y compris celle des mathématiques, doit se ramener à des intuitions ainsi comprises. Mais il faudra toujours distinguer l'inscription des idées dans l'ordre virtuel de leurs connexions déductives (par exemple les propositions de la géométrie d'Euclide), et leur inscription dans l'ordre empirique de leur perception par l'esprit. C'est en adoptant ce second point de vue qu'il faudra tenir compte des habitudes intellectuelles de celui qui les pense, de ses dispositions, de son caractère, de ses souvenirs, de ses émotions, etc., bref, de tout ce qui fait l'épaisseur d'un sujet humain, en tant qu'il fait un usage effectif (donc empirique) de ses facultés. Vis-à-vis de cette réalité-là, l'édifice des mathématiques, en lui-même, n'est pour Locke qu'une virtualité.

Lorsque Kant, dans la *Critique de la raison pure*, distingue aperception pure et aperception empirique, l'enjeu est comparable. Et c'est bien pourquoi il reconnaît à Locke d'avoir bien rendu compte des exigences liées à « l'exercice » de notre faculté de connaître[14]. Même si Locke n'a pas été assez loin dans la reconnaissance du soubassement légal qui sous-tend *a*

[14] CRP, B 118-119.

priori la possibilité de l'expérience, il a posé avec force que l'idéalité de propositions nécessaires et universelles comme celles des mathématiques ne pouvait rester qu'une simple virtualité si elle ne s'ancrait dans les détours d'une perception effective. On peut peut-être mieux comprendre à la lumière de ce rapprochement pourquoi pour Kant également, l'aperception empirique doit toujours rester au centre de la réflexion, y compris s'agissant des mathématiques, et cela, sans que cela ne nuise le moins du monde à la cohérence et à la radicalité du discours transcendantal, en tant qu'il est et reste fondé sur la référence à l'aperception pure.

*****

Ainsi, l'originalité de la position de Kant s'agissant de l'aperception est d'avoir saisi avec une singulière radicalité à quel point l'acte effectif du 'je pense' est doublé en permanence, au niveau de l'*a priori* par cette aperception pure qui en conditionne la possibilité même. Mais inversement, Kant pose avec tout autant de force que l'aperception pure est un 'je pense' potentiel qui ne s'actualise que dans l'expérience. Le vrai problème kantien de l'aperception réside dans ce va-et-vient irréductible entre l'*a priori* et l'empirique.

C'est aussi là un des ressorts profonds de son fameux intuitionnisme en ce qui concerne la connaissance mathématique. Au-delà de l'ancrage des concepts dans l'intuition pure, s'y joue également la question de la réalité de la pensée mathématique, non seulement vis-à-vis des objets, mais aussi en tant qu'elle doit se ramener en fin de compte à une conduite empirique des représentations dans l'esprit : c'est cela qui la rend réelle, autant que la nécessité et l'universalité des propositions qu'elle formule, quoique d'une autre manière et à un autre niveau. Ce problème pourrait s'inscrire dans une lecture plus large de la philosophie critique, dans laquelle l'empirique ne doit pas être par principe considéré comme 'inférieur' au transcendantal pour la

On the possibility and reality of introspection

Michel Bitbol
(Archives Husserl, UMR CNRS, Paris, France)
&
Claire Petitmengin
(Institut Mines-Télécom, Paris, France)
michel.bitbol@ens.fr
claire.petitmengin@polytechnique.edu

Introduction

The reliability and accuracy of introspective research has been and is still a topic for hot debate (Hurlburt & Schwitzgebel, 2007). In the history of philosophy and psychology, conflicting claims have been made about whether this exploration of the so-called "inner" realm can be made reliable at all. According to the Cartesian, empiricist, and phenomenological lineage, consciousness is necessarily infallible about itself. Husserl (1913) thus replaced the standard psychological division between inner and outer perception he had inherited from Brentano, with a division between *certain* (immediate and complete) and *uncertain* (mediate and incomplete) perception within the flux of lived experience. Perception of immediate lived experience is certain because the *way it appears* coincides with *the way it is*, whereas perception of spatial objects is uncertain because at each moment they present themselves through partial profiles (or "adumbrations": *abschattungen*) whose spontaneous ontological interpretation can later be disconfirmed. The opposite view, however, has gained prominence during the past century. From the behaviorist rejection of introspection to the thorough doubts expressed by Schwitzgebel (2011), the common view has been that as soon as we try to report our experience, we fall into confusion, we gain no true knowledge, and we even tend to confabulate (Nisbett & Wilson, 1977).

But are these seemingly opposite positions really incompatible? There might in fact be no true contradiction between them, provided one realizes they rely on very different definitions of knowledge, and different conceptions of what is to be *expected* from introspection. Introspective (or rather first-person) reports by single individuals may indeed be flawed when they are taken at face value, as exhaustive descriptions of, and objective knowledge *about,* the cognitive processes taking place in these individuals. What they do is nothing more than reflectively expressing knowledge *by acquaintance* of elementary (pleasure, pain, fear, joy), or elaborated (temporally sequential, spatially distributed, proprioceptive or emotive) experience. But as such, they have a crucial epistemic role to play. Although first-person reports may fail to be self-sufficient pieces of knowledge (beyond acquaintance), they remain the unique and inescapable basis of any further empirical knowledge of ourselves and of our environment. First-person access is the testimony of our being-in-the-world, and the source of every claim of the availability of a surrounding world. This universal inescapableness and fundamental importance of first-person access should be no surprise, but it is often underrated in current epistemology.

One too often forgets that first-person reports are indispensable to ascribe functional meaning to most neurophysiological patterns (Lachaux, 2011; Kriegel, 2013), and to guide research in such field. One also too often loses sight of the fact that even the "objective experimental data" of natural sciences are nothing else than convergent *first-person reports* of a certain type. Actually, these data identify with specific first-person reports about having witnessed that a certain controlled phenomenon falls into one or another category defined by a preliminary intellectual framework (blue or red, positive or negative, On or Off, spin projection +1/2 or -1/2, etc.). In particular, measuring is tantamount to *reporting* that some meter-reading is seen to be included in, or excluded from, a given numerical interval. What gives objective data or measurements their reliability is nothing else than the coarseness of the categories in which first-person reports are constrained to fall, assisted by instrumental amplification of coarseness. Indeed, this coarseness makes final mistakes and disagreements virtually impossible: everybody can agree that this meter-reading falls in a certain numerical interval, even if there is persistent disagreement about associated nuances of color, emotive content, or interpretation.

It is now clear that reliability by no way requires the complete *elimination* of first-person reports. First person reports remain the *de facto* starting point and

ultimate warrant of the whole system of knowledge. The only question that remains open at this point is the following: is it possible to extend the domain of reliability of first-person access beyond the very coarse framework that is sufficient for perceiving *properties of public objects?* Can one extrapolate this domain of reliability towards more subtle aspects of experience that would afford information about the very process of perception, valuation, mental strategy, self-monitoring (and more generally cognition), though without claiming to disclose immediately cognitive processes as they are?

Even about the latter question, there are pessimistic and optimistic views that rely on different theories of mind and consciousness. The pessimistic view derives from a "scarce" view of mind and consciousness, according to which most mental processes being unconscious, they are doomed to remain forever inaccessible to first-person access. The optimistic view, instead, derives from an "abundant" view of consciousness, according to which most (or all) mental processes are *experienced* yet not always *attended* and *reflected upon* (Marcel, 2003; Block, 2011). In the latter case, one must only find a way to unfold the unattended experienced material, and bring it to full reflection[1]. Then, once a large field of experience is thus reflected and expressed (beyond the narrow circle of the objectifying coarse categories), the following task is to find renewed criteria of reliability and intersubjective agreement that would turn this extended reflection and expression into an acceptable source of knowledge. Is the latter program feasible? Lots of in-principle objections have been formulated against it in classical and modern literature. But since these objections target an abstract image of introspection rather than introspection *per se*, we want to quickly overcome them and see if a concrete project of rebirth of introspection can meet them *in practice*. We will thus list these objections in turn (Petitmengin & Bitbol, 2009; Vermersch, 1999) and outline some replies new introspection has in store for them, in addition to some theoretical rebuttals based on contemporary philosophy of science. We will focus on one of the currently available methods that we ourselves practice: the elicitation interview method[2] (Vermersch, 1994; Petitmengin, 2006). Our aim is to show that, irrespective of its alleged theoretical "impossibility", introspection is a living reality.

[1] The word "reflection" has to be used with caution, in view of its spurious connotations of detachment and look. This point will motivate an extensive discussion below.
[2] The expression « elicitation interview » translates the French original name of the method: « entretien d'explicitation ».

1. Is it necessary to transform a subject into an object?

The most archetypal objection against introspection is that it is impossible to observe one's own experience, because this presupposes a split between subject and object while in this case the object is nothing else than the subject itself. A very early form of this objection was formulated by Socrates himself, in the *Charmides* (167 c-d), in order to challenge a widespread conception of wisdom as self-knowledge: "Suppose that there is a kind of vision … which in seeing sees no colour, but only itself and other sorts of vision: Do you think that there is such a kind of vision? Certainly not!" (Roustang, 2009, p. 78). According to the Platonician dialogues that are most likely to express Socrates' position, then, there is no such thing as self-vision, self-hearing, and by extension self-knowledge. But the most well known version of the objection was stated by Auguste Comte (the creator of positivism): "As for observing … intellectual phenomena in their process of execution, there is an obvious impossibility. The thinking individual cannot split himself in two parts, one who reasons and the other one who looks at the reasoning. The observed organ and the observing organ being in this case identical, how could observation take place?" (Comte, 1830/2001).

We must point out from the outset that this kind of objection is directed against introspection as prejudice says it *should be*, rather than against introspection as it *is in fact* practiced. The prejudice is that part of the subject engages in second-order observing or monitoring of first-order mental processes. But, against this prejudice, many results, including from neurophysiology (Overgaard et al., 2006), are consistent with the idea that introspection merely involves a modified version of those very first-order mental processes. However, we do not want to discard the Comte-like objection too quickly. Instead, we will develop this objection and this prejudice one step further, and then compare it with a similar problem in the history of the interpretation of quantum mechanics. Such lateral strategy will substantiate our reply.

An important correlate of the alleged splitting of subject and object in introspection was stated repeatedly in the history of psychology : "suppose a particularly persistent introspectionist should desire to introspect the reporting or secondary series, would he not have to assume a third series, and so on, *ad infinitum* and *ad nauseam*?" (Ten Hoor, 1932). This threat of infinite regress pertaining to "inner observation" had been identified and discussed

much earlier by Harald Høffding (1905), a Danish philosopher who was a major inspiration of Niels Bohr, one of the most important creators of quantum mechanics. As a consequence, Niels Bohr (1934) tended to make a strong analogy between: (i) the situation of an introspector who wishes to observe herself by splitting into a subject part and an object part, and (ii) the situation of an experimenter in quantum mechanics who is (instrumentally and interpretationally) intermingled with microscopic phenomena, yet wants to observe them. In both cases, said Bohr, one witnesses a kind of dialectic between (a) the actual inseparability and (b) the alleged necessity of separation between subject and object. *De facto* inseparability imposes strong constraints on any attempt at enforcing some sort of artificial distinction between subject and object for the sake of knowledge. Indeed, as soon as some divide between object and subject is conventionally imposed *despite* their actual inseparability, part of the object to be known happens to be cut off (because it is conventionally retained on the subject-side of the divide).

However, this dialectical strategy advocated by Bohr is very disputable. Isn't it possible to do without any artificial separation of subject and object, yet approaching microphysical and experiential phenomena in a scientific way? As we argued in previous work (Bitbol, 1996, 2000, 2002), this can perfectly be done provided one does not attempt to objectify a putative property behind each *singular* phenomenon, but only the structure that enable us to anticipate phenomena of each *class,* and under each *type* of circumstance[3]. Such an alternative approach will be developed in section 4, as part of our discussion of the kind of objectivity that can be reached by introspective inquiry. Meanwhile, we have to probe further into the claim that the standard Comte's objection to introspection misses it target. To that purpose, we must be more accurate about the very definition of introspection, and show that once it is appropriately characterized, it automatically escapes the objection.

Are we really doomed to the dualist picture of inner and outer realms that would fully justify using the term "intro-spection" about a certain mental act of meta-awareness or "reflection"? Is this picture that makes it so easy to formulate Comte's objection doing justice to the real work of introspection? As a preliminary move, we wish to point out that few philosophers of the turn of

[3] In quantum mechanics, it is well-known (to the dismay of realist philosophers of science) that the project of objectifiying "properties" behind phenomena can hardly be worked out. Yet, one objectifies a universal anticipative structure which is nothing else than the state vector, that generates probabilistic predictions by means of the Born's rule.

the nineteenth and twentieth century, who determined the cultural background of the first wave of introspectionist psychology, took seriously this picture.

Thus, instead of taking the dualist picture for granted, the German Neo-Kantian philosopher P. Natorp (1912) gave a detailed account of how the dual organization of knowledge (object and subject, outer and inner) may arise from the undifferentiated continuum of experience. According to him, this occurs by way of a double-faced process in which objectivation comes first, and subjectivation arises as the by-product of the former. Objectifying means picking out the component of experience that remains invariable across personal, spatial or temporal situations; or at least the component of experience that vary in the same way (i.e. in a law-like way) irrespective of the personal, spatial or temporal situations. The "subjective" domain is then marked off by contrast and difference with the objectified part of experience. It includes whatever is left in experience after the objective domain has been delineated. Accordingly, the subjective domain evolves with the process of objectification, and it receives as many characterizations as there are delineations of objectivity. This means that accessing the domain of subjectivity is not just a gift, but a *discipline* symmetrical to the discipline of objectification. One can access this domain by pondering about the (subjective) conditions of possibility of objective knowledge. One can also reach it by suspending the fragmentation of the field of experience into coarse categories required for objective knowledge, and by relaxing the interest of knowledge initially directed towards restrictive parts of experience.

Yet, despite this philosophical critique, most of the overt characterizations of introspection given by the psychologists themselves remained in line with dualism. The two-realms and two-directions-of-gaze model was still pregnant at the turn of the nineteenth and twentieth century. Wilhelm Wundt (1901) thus wondered "how can our own mental life be made the subject of investigation like the objects of this external world of things about us?". Similarly, Edward Titchener (1912) approved the idea that "introspection is simply the common scientific method of observation, applied from the standpoint of a descriptive psychology". He then stated the different directions of gaze by which one should characterize the two kinds of "observation": "the method of psychology is observation. To distinguish it from the observation of physical science, which is inspection, a *looking-at*, psychological observation has been termed introspection, a *looking-within*" (Titchener, 1916, p. 20). Later textbooks of psychology usually retained the standard conception of introspection as *observation of some internal occurrence*, e.g. "introspection

is most simply defined as the direct observation of one's own mental processes" (Moore & Gurnee, 1935, p. 30). The paradigm of detachment thus pervades even introspective psychology.

It is on this unsophisticated epistemological ground that nuances and doubts grew up. Wundt resisted from the outset the rough definition of introspection as "inner *observation*", and rather referred to "inner *perception*", thus accepting a distinction first introduced by Brentano (Brentano, 1874/1944). According to Brentano, inner observation cannot be the "true source of psychology", for observing a mental event by fully focusing one's attention towards it would just lead to its disappearance. The true source of psychological inquiry is then inner *perception*, that does not require that attention be focused on some mental object, but only that, when attention is focused on some (usually external) object, it remains broad enough to notice other events such as the mental processes that underlie the act of attending. One can thus *perceive* a vibration of the telescope while *observing* a planet. This defocusing of the field of attention performed in "inner perception" has also been called "non-observational awareness" (Marcel, 2003). As for Titchener, he relied on the "introspective habit" of trained subjects, who were able "not only to take mental notes while the observation is in progress, without interfering with consciousness, but even to jot down written notes" (Titchener, 1916, p. 22). But what is this special ability trained subjects acquire when they perform introspection in the style of Tichener? A reasonable assumption, in line with Brentano's and Wundt's characterization of "inner perception", is that it is the ability to detect laterally occurrences that are not in the main focus of attention.

This, at any rate, fits remarkably well with E. Husserl's characterization of phenomenological reduction, which is the chief method to give access, not to the "inner world", but rather to the *whole field of pure experience* before exclusive intentional focusing has narrowed down the region of our full awareness. Phenomenological reduction, says Husserl (Husserl, 2002, p.11), helps revealing the "sides" (or the margins) of our experience that are overlooked as long as exclusive concern for objects prevails. Husserl insisted on the full openness of the subject to the manifold of lived experience during phenomenological reduction (Depraz, 2008, p.103). Even when Husserl used a metaphor of "splitting" of the subject in reflection, he mentioned that, by such splitting, I become "*at the same time* plainly seeing subject and subject

of pure self-knowledge"[4]. Later on, this move was confirmed by M. Merleau-Ponty, according to whom the phenomenological attitude means (in terms borrowed from Bergson) that, "instead of wanting to raise ourselves above our perception of things, we plunge into it to dig it out and *enlarge* it" (Merleau-Ponty, 1989, p. 22; Bergson, 1934, p. 148).

True, one must not overlook Husserl's own forceful denial that the phenomenological enquiry relies on some variety of introspection. He gave three major reasons for this denial: (i) Introspection, he wrote in his *Ideen I*, arises from a state of *positional* consciousness (which means that in this case consciousness *posits* an intentional object, be it in the focus or in the margin of attention); by contrast, in the genuine phenomenological stance, consciousness remains "non-positional"[5]. (ii) Being "positional", and therefore directed towards some sort of transcendent object, introspection remains fallible as any empirical investigation is; by contrast, being non-positional and therefore immersed in immanence, the phenomenological stance is supposed to reach absolute certainty. (iii) Phenomenology is not concerned by single events of mental life, unlike the primary step of introspection; it aims at elucidating the invariants (or "essences") of lived experience.

But, notwithstanding these differences, part of Husserl's characterization of the phenomenological stance supports a new understanding of introspection. Intro-spection here appears as (or is replaced by) a mental state in its own right, a state of broadened awareness, rather than being taken as a homonuclear act of observation of some other mental act or mental state. "Reflection" in a phenomenological sense no longer means a sort of specular (transcendent) observation, but rather a *modification of consciousness*, a *transmutation of lived experience as a whole*, a series of immanent *modes of capture of essences* (Husserl, 1913/2004, §78). To stress the difference without breaking lexical continuity, we can give a slightly different name to this renewed concept of "reflection". We propose "coreflection". The latter neologism may prove useful to convey two semantic shifts. According to the first shift, we are no longer concerned by a mere asymmetric revelation of the "seeing subject" by the "subject of self-knowledge", but by their symmetric *co*-definition within the experiential field of somebody who has practiced the

[4] E. Husserl, Erste Philosophie (1923/4). Zweiter Teil: Theorie der phänomenologischen Reduktion. [First philosophy (1923/24). Second part: theory of phenomenological reduction.] Ed. R. Boehm, Martinus Nijhoff, 1959. French translation: E. Husserl, Philosophie première, P.U.F., 1972, p. 156

[5] See a discussion in (Flajoliet, 2006).

phenomenological "reduction". According to the second semantic shift, the so-called "reduction" represents in fact an enlargement of the span of experience, and this can be evoked by the three first letters of the word "coreflection": "cor" for the Greek "khôra" which Plato used in the *Timaeus* to mean space, or interval.

Full realization of this alternative status of introspection is commonplace nowadays. G. Ten Elshof (2005) thus claims that introspection can still be considered as a kind of perception, provided one recognizes that the essential act of any perception is not only redirecting attention but also *changing its span*. Similarly, by making a cogent synthesis of Brentano's and Wundt's thoughts, J. Sackur (Sackur, 2009) defines introspection as a process of perception *expanded* to what is usually neglected, or to what is usually at the periphery of the attention field. Introspection, far from being like a gaze on some object (be it focused or expanded), is tantamount to (re) establish an intimate and close *contact* with what is to be explored (to wit the field of lived experience) (Petitmengin & Bitbol, 2009). The metaphor of the sense of touch (with closed eyes), or smell (Kriegel, 2013), here replaces the metaphor of the sense of vision.

Two major developments of our *Weltanschauung* and of the cognitive sciences can explain why this alternative, non-observational and non-visual, conception of introspection is now much easier to accept than it was at the beginning of the twentieth century. One of them is our growing familiarity with contemplative methods, whose aim is to stabilize attention and use this stabilization in order to get a precise knowledge by *acquaintance* of the subtlest aspects of mental processes[6]. Along with this perspective, the idea of "non-positional" consciousness, or of intimate contact with experience, as opposed to the old-fashioned observational view of introspection, is no longer problematic. Thus, according to A. Wallace, "Unlike objective knowledge, contemplation does not merely move towards its object; it already rests in it" (Wallace, 2006).

The other development that makes the non-observational conception of introspection easier to accept can be found in the cognitive sciences. It is the widespread recognition (Schooler, 2002) of a background short-term cognitive

[6] In meditation, stabilizing attention is allowed by long sessions of concentration on a single felt or imagined process (such as breath or pictures) ; and contact with the manifold processes of mental life is realized not only by broadening the field of attention, but also by dropping "all aim and objective" in full, open, non-directional, mindfulness. See e.g. (Genoud, 2009; Wallace, 1998).

unconscious (Hassin, Uleman & Bargh, 2006), in addition to the long-term affective unconscious delineated by Freud. Provided the word "unconscious" is not taken at face value, but rather identified to "unreflective", this allows one to confirm the image of focus and margin of conscious awareness that sounded so problematic during the first wave of introspective psychology (Bode, 1913).

Recent methods of verbal report and introspection fully take this conception into account. The elicitation interview method (Vermersch, 1994; Depraz, Varela & Vermersch, 2003; Petitmengin, 2006; Petitmengin & al., 2009) that we currently practice can be characterized as a strategy for progressively unfolding initially "pre-reflective" aspects of lived experience, by asking subjects to rehearse and even to *re-enact* this experience while broadening their field of attention. Here, *retrospection* (as opposed to "thinking-aloud" protocols) is systematically used. But this is not only to meet the traditional objection according to which observation disturbs the observed process if it occurs simultaneously to it (an objection automatically inactivated by the rejection of the observation conception of introspection). This is also to enable patient expansion of awareness, part after part of a selected slice of experience. The success of such procedure confirms that episodic recollection is an excellent way to reinstate immersion within a broadened field of experience (Marcel, 2003). Another, very different, method has also been developed to overcome the problem of bringing to awareness as many pre-reflective aspects of experience as possible. Its name is "descriptive experience sampling method" (Hurlburt & Heavey, 2006). It consists in interrupting subjects in the course of their tasks by means of a beep triggered by a random timer, and asking them to report on whatever was going on in their minds a few seconds before the beep. This allows something like "tomography" of moments of experience of which subjects remain usually unaware (because when no beeping occurs, they immediately switch to the most relevant aspects of their main target rather than pondering upon its experiential context).

To sum up, there are two crucial points on which the current definition of introspection differs from the classical one, thus offering it a better opportunity of development: (i) overt cultivation of contact with and growing awareness of an all-pervasive experience, rather than observation directed towards some "inner" sphere of processes; (ii) techniques for encompassing pre-reflective (or "cognitively unconscious") parts of experience in successive fields of attention. Both moves might motivate rejection of the word "intro-spection"

and use of alternative expressions instead (e.g. “expanded mindfulness”), but it is convenient to keep the old word with us in order not to minimize a certain amount of historical continuity.

2. Does introspective examination disturbs its “object”?

Let’s come now to the objection that introspection alters the mental process to be known. There are at least three varieties and many sub-varieties of this objection.

A. Observational distortion

The attitude or operation of introspection *disturbs* the mental flux to be known. This objection was already formulated by Hume: “its evident this reflection ... would so disturb the operation of my natural principles as must render it impossible to form any just conclusion from the phenomenon” (Hume, 1739 / 1978, Introduction). And it was considered as a problem to be solved by the introspectionists: “If you try to report the changes in consciousness, while these changes are in progress, you interfere with consciousness” (Titchener, 1916, p. 22).

B. Temporal distortion

This objection comes in two major guises that we will now document.

B.1 One problem is a discrepancy between the fluent nature of experience and the request of stability of knowledge contents. Kant (1786/2002, Introduction) thus claimed that there can be no knowledge of the soul, because the latter develops in time, whereas one should be able to immobilize it somehow in order to extract some knowable invariant. A different (somewhat reciprocal) difficulty was pointed out by Wittgenstein (1964/1980). According to him language, whose use is extended in time, can by no means catch experience in its present unstable actuality.

B.2 Another problem (that may be a consequence of the first one) is that what can be captured and mastered in experience is only its *past* unfolding. G.H. Mead and J.P. Sartre (2000) thus pointed out that the “I” itself can only be considered as a reconstruction, or that the “I” is always in the past. But if this is the case, isn’t there a risk of deformation or oblivion? Can’t there be *a posteriori* falsification of the history of lived experience, by the processes that

D. Dennett calls "Orwellian" and "Stalinesque"[7]? Isn't experience thus replaced with a rational reconstruction made out of prejudice?

C. Interpretative distortion

The categories that subjects apply when they describe their own experience are theory-laden (Gopnik & Meltzoff, 1994; Robbins, 2004). This is a real problem since, as shown by Nisbett and Wilson (Nisbett & Wilson, 1977; Johansson et al., 2006), subjects are very bad at theorizing about their own mental processes. Moreover, the use of words alters the experience to be described, and they are even likely to be unable to capture anything properly in experience (this is the charge of *ineffability*).

This series of objections is not as threatening as it looks. Indeed, observational, temporal, and interpretative distortions can only be called "distortions" with respect to experience as it is in itself, previous to any attempt at observing, catching, and interpreting. In other terms, the previous objections rely on some version of the myth of the "given" (Garfield, 1989). But if we distance ourselves from this myth, a very different picture arises.

An examination of the claim according to which certain processes are "disturbed" (Jack & Roepstorff, 2002) by observation and/or verbalization can be taken as a first step towards the new picture. Speaking of a process *an sich* that is unfortunately disturbed by the coarse instruments we use in order to have access to it, only makes senses if there is a way of accessing it independently of these coarse instruments. But if there is nothing even in principle to compare with the instrumental outcomes, this is wild speculation. Such a simple remark is (or should be) a keystone of the interpretation of quantum mechanics. True, the metaphor of an object disturbed by the experimental contraption has usually been accepted by physicists in the first years after quantum mechanics was formulated; and it is still used in popular science books. But it became clear in the following years that, if taken seriously, this metaphor could only lead to the accusation of "incompleteness" of quantum mechanics. This accusation in turn fed the persistent dream of a "hidden variable theory". The metaphor of disturbance was then soon

[7] Retrospective alteration of history can be obtained in two ways, according to Dennett. In the Orwellian way, somebody first makes one conclusion based on partial evidence, and then changes her memory of having made this previous conclusion in order to accommodate further evidence. In the Stalinesque way, somebody does not make any intermediate conclusion but entirely reconstruct the whole sequence ex post facto, when all the evidence is available.

discarded by Bohr, and replaced by the claim that a phenomenon is co-*defined* by the experimental conditions of its manifestation, rather than *disturbed* by them. Here, the phenomenon is taken as inseparable of its experimental context. The new physics is seen as bearing immediately on technologically holistic phenomena, rather than mediately on putative properties "revealed" yet "distorted" by the apparatus.

A similar move has been suggested for introspection. Husserl's sharp reply against the early opponents of introspection (Husserl, 1913/2004, § 79) was exactly along these lines. He noticed that when one casts doubts on the possibility of faithfully capturing lived experiences in reflection, one thereby presupposes some form of *knowledge* about what *are* these lived experiences *prior to any reflection*. But this is either self-contradictory (if knowledge of experience can only be obtained by reflection), or self-mandatory (if one is summoned to define alternative, and elusive, ways of self-knowledge). The only way out of this dilemma, as expressed by B. Shanon (1984), is then to accept that introspection bears directly on reflective experiences rather than indirectly on the experience the reflection is supposed to be about. To be sure, not caring for anything like representational faithfulness of reports is provocative, but this decision has the merit of pointing towards alternative epistemologies and alternative strategies. One such strategy is precisely to emulate the epistemological approach of standard quantum mechanics, and elaborate an overtly non-representational science of experience.

3. Is one systematically mistaken about one's own experience?

Part of this objection is grounded on the observation that it is very easy for subjects to go astray about the *stimulus* that was applied to them in order to trigger a certain experience. Titchener himself, in his defence of systematic introspection, was extremely diffident about the ability of subjects to identify a stimulus: "The subject may see what was not there at all, may fail to see much of what was there, and may misrepresent the little that he really perceives; introspection adds, subtracts, and distorts" (Titchener, 1912; Schwitzgebel, 2004). More recently, criticisms have been formulated against the propensity subjects have to say that they see more than they can evidence (Dennett, 1992, 2002), or against their unability to see major parts of what occurs in front of them if their attention is distracted (as shown by experiments of "change blindness" (Silverman & Mack, 2006)). However, this

charge might well be excessive or misplaced. In a non-representationalist epistemological framework, the issue of the truth or reliability of introspective descriptions is likely to be given a completely new meaning.

The first criterio of truth that comes to mind under the presupposition of a representationalist theory of knowledge, is that introspective descriptions should be faithful to the experimental or environmental *input* that triggered the experience reported. This (too) simple idea has long been criticized in old introspectionism, and replaced with the criterion that an introspective description should only be faithful to a slice of experience (rather than to what it is an experience *of*). Titchener thus wrote: "The question, … so far as the validity of introspection is concerned, is not whether the reports tally with the *stimuli*, but whether they give accurate descriptions of the observer's experimental consciousness; they might be fantastically wrong in the first regard, and yet absolutely accurate in regard to conscious contents" (Titchener, 1912). Here, it looks like Titchener accepts the correspondence theory of truth which goes along with a representationalist epistemology, although he applies it to "conscious contents" rather than to "*stimuli*". We will come back to this point soon, but let us first dig more carefully into what the followers of the American introspectionist school called "the *stimulus* error" (Boring, 1929, p. 33).

This presciption *not* to seek correspondence between introspective data and *stimuli* might well have been directed against the first German school of introspection, namely Wundt's. But even in this case, the criticism is excessive. Indeed, with the help of the instruments of his laboratory, Wundt focused his inquiry on very limited introspective reports having the form of judgments of time-characteristics (duration or simultaneity), number, and intensity of *stimuli*. And, under strict experimental control, his introspecting subjects turned out to be reasonably faithful to the *stimuli* that were imposed to them (Wundt, 1901, p. 31). A modified version of Wundt-like introspection has been revived recently with considerable success (under the name "quantified introspection" (Corallo et al., 2008)), and it also yields a positive outcome about the accuracy of simple reports. Here, the reports bear not on the stimuli themselves, but on the time spent by subjects to perform a certain task involving simple stimuli. The suspicion of inaccuracy about *stimuli*, being partly misplaced, is then not sufficient to motivate the rejection of introspection.

Another indication that introspective reports may be less inaccurate about their *stimuli* than is usually thought, can be found in disguised introspective work of the allegedly behaviorist era. One such research casts doubts on a widespread anti-introspectionist prejudice of cognitive scientists (after Dennett): the prejudice according to which subjects are systematically wrong about their pretending to see a whole scene extended in space, since they are in fact unable to describe most details of this scene when they are asked to do so. A classical work by G. Sperling (Sperling, 1960)[8] indeed showed that things might be much more intricate than this, and less challenging for first-person access. Sperling briefly confronted subjects with a 4x4 table of letters, and asked them to report the letters they could remember. Subjects usually claimed they had an iconic memory of the whole table, but, irrespective of the size of the table, they could hardly report more than 4 letters out of it. Was their claim of being able to *see* the whole table after its presentation completely illusory? Further inquiry ruled out this negative interpretation of the initial reports. Subjects were asked to concentrate on a single line in the table, and to list the letters of *this* line. The outcome is surprising: subjects were able to report about 3-4 letters of *any* line chosen at random by the experimenter. So, we are inclined to accept that they indeed had a short-term iconic memory of the whole table. Accordingly, it was advocated recently (Block, 2011) that the initial introspective report of the subjects was much more accurate than what is usually suspected.

The way this accuracy was brought out is also very instructive: (i) put subjects in a situation of success rather than a situation of failure (i.e. choose the task in which subjects display optimal performance); (ii) help them by asking focused questions about what they lived, rather than dispersing their attention by abstract questions. This is precisely the strategy that is followed in the method of interview we practice (Petitmengin, 2006).

Another *locus classicus* of the criticism of introspection, from which J.B. Watson inferred that a true science of mind could only be grounded on the study of behavior, is the famous unresolved quarrel of “imageless thought” (Ogden, 1911; Woodworth, 1906). This time, the threat to introspectionism looks even more serious than before, since the issue no longer bears on the ability of introspective reports to be faithful to the stimulus that triggered experience, but on their faithfulness to experience itself. In the heyday of introspectionism, the researchers of Titchener’s school at Cornell University

[8] Quoted and discussed by J. Sackur (Sackur, 2009).

claimed to have brought out the presence of sense elements, kinesthaetic feelings, and images associated to *every* thought process (Titchener, 1909), whereas the researchers of the Würzburg school, such as Külpe, Mayer, and Orth (Humphrey, 1951), declared that there exists imageless and even "nonsensory" thought. These conflicting claims were associated with mutual methodological criticism (Nahmias, 2002). As K. Danziger pointed out (Danziger, 1980, 1994), this quarrel showed how "theoretical differences could readily be made to take on the form of differences in the data themselves". But careful examination of the texts in which the debate about imageless thought developed has shown that the nuclear proto-interpreted data could after all be isolated from the school-related theoretical bias, and that in this case, no true divergence persisted (Hurlburt & Heavey, 2001; discussion in Goldman, 2001). Subjects of both schools indeed reported the existence of "vague and elusive processes, which carry as if in a nutshell the entire meaning of a situation" (Titchener, 1910/1980, p. 505-506), but they did not *interpret* these reports the same way; and both school probably missed a more faithful description of them in terms of "felt meanings" (Gendlin, 1962).

More than a failure of introspection, this indicates what kind of work should be done in order to reach a possibility of intersubjective agreement: stepping down on the scale of rational reconstructions, explanations, or generalizations, and sticking to the "how" of experience (Petitmengin, 2006). In any experimental science, identifying "facts" requires a process of descent along the hierarchy of theory-ladenness; not of course in order to reach a utopic realm of "pure non-interpreted content", but only to pick out a level of interpretation that is beyond discussion in a certain state of culture and research.

But how exactly can one ascertain the "faithfulness" of first-person reports, independently of any relation with the stimuli that triggered experience? One may distinguish two levels of faithfulness assessment: (a) signs of reliability, and (b) criteria of validity.

(a) As we have just seen about the "quarrel of images", there is one index whose presence leads to strong suspicions: this is lack of consensus about general structures of lived experience. Conversely, one may take consensus about structures as an index of faithfulness, although this consensus might well be partly induced by theoretical (or sub-theoretical) prejudice. To avoid the latter bias as much as possible, we need *individual* signs of reliability that may help us to increase the degree of credibility of each interview taken apart. Such signs are currently in use, and their significance has been carefully

discussed (Vermersch, 1994; Petitmengin, 2006; Hendricks, 2009). They are detected in the form of bodily attitudes and rythms of speech that evoke actual contact with one's experience during the process of reporting. However, one must keep in mind that such signs are taken as good ground for reliability only because they are connected with first-person access of the interviewers to the experiential correlates of similar signs within their own bodies. This suggests that faithfulness of first-person reports can be ascertained only by intersubjective criteria; there is no external "absolute" evidence.

(b) The same can be said when criteria of validity, or even *truth*, of these reports are sought. Indeed, there is at least one thing that we can say for sure: there is no way of comparing directly an experience *an sich* and its alleged report. This is obvious for experimenters, but this is also clear for subjects themselves, since their own act of "comparison" is a new experience in which the former experience to be reported is merged and recast. So, how can we sort out this difficult epistemological situation? By relying on sound epistemology, rather than on the old representationalist and dualist epistemology.

To take a significant step in this direction, we may conveniently come back to Kant. The age-old objection of skeptics according to whom we have no "absolute" access to things (no access apart from the causal relations we have with them), and that therefore we can say nothing about what they are *in themselves* apart from the effect they have on us, was addressed by Kant in a very innovative way. He first acknowledged that we indeed have no apprehension of objects apart from our very procedure of access (Kant, 1800/1988). Then, instead of trying to prove the correspondence between knowledge contents and some independent object "out there", he *defined* the object as whatever is shaped by the class of perceptual/intellectual operations used in the act of knowing. The stable component of experience is considered "objective" *by definition*, and not in virtue of its (doubtful) correspondence with some extra-experiential reality. This suggests that skepticism about any region of knowledge cannot be overcome by relying on some external warrant, but only by using *internal criteria*.

Accordingly, when we look for criteria of validity of first-person reports able to resist to skeptical doubts, we bypass the fruitless search for their *correspondence* with putative "private objects" and rather try to establish criteria of *self*-validation. We also exploit the opportunities of *mutual* validation offered by articulating the domain of first-person reports with several areas of

cognitive science.

This strategy fits with current philosophy of science, which is undergoing a major paradigm shift. The traditional debate about whether scientific theories are able (or not) to provide us with a faithful description of an independent reality is fading away. Experimental gestures, mathematical practices, and social debates are no longer seen as mere neutral windows opening on "pure", "independent" reality. Instead, they are understood as an interfacial matrix of on-going agency, out of which strategies of theoretical prediction and conceptions of reality able to guide them co-emerge (Pickering, 1995; Gooding & al., 2005; Galison, 1987). Here, as in Kant, answering skeptical doubts no longer amounts to display a one-one correspondence between theoretical symbols and real properties. It rather requires to find patterns of technological actions that have stabilized, have been adopted collectively for their success, and have then been connected to one another in coherent networks. The new kind of answer to skepticism relies on a pragmatic coherentist conception of truth, rather than on a correspondence theory of truth.

The same attitude towards skepticism can be adopted when the validity of first-person reports is at stake (Shanon, 1984; Piccinini, 2003; Piccinini, 2009; Petitmengin & Bitbol, 2009). These authors pointed out that standard critiques just show that introspective data cannot usually be evaluated on the basis of correspondence; and that this is not to be wondered about or regretted, since after all no other data, including in experimental science, are *really* evaluated this way. The alternative is then evaluation on the basis of performative coherence, where "coherence" can concern several levels of practice: internal coherence in self-assessment and report; interpersonal coherence in dialogue (see above); and triangulated coherence in a network connecting introspective reports with experimental (neurological) practice.

This retreat from the correspondence theory of truth to an extended version of the coherence theory of truth however does not mean that there is no prospect to *improve by way of coherence the probability of correspondence* between an introspective report and the experience it is meant to describe. The elicitation interview method is especially suited for that purpose, in view of its ability to focus the attention of subjects on the aspects of their experience which they better access, and avoid overinterpreting them. It has thus been shown that one can considerably improve the standardly defined faithfulness of first-person reports precisely in the experimental situation that has been taken for more than thirty years as the archetypal

rebuttal of introspection, namely in the Nisbett & Wilson (1977) setting. This improvement, that raises the correspondence between an initial experience of choice of presented faces and the later report of this experience from about 30% to about 80%, has been obtained by inserting an elicitation interview between the moment of the choice and the moment of the final report (Petitmengin & al. 2013).

4. Can knowledge about subjects be somehow objective?

The fourth and final group of objections focuses on the purely subjective status of introspective descriptions, and on the fact that the situation it concerns is irreproducible. Thus, according to Wundt's early but harsh criticism, unless it is constrained by a strong experimental environment of control, introspection is doomed to extreme idiosyncrasy: "introspective reports offer no means for independent checks by which they may be evaluated. Indeed, the reports are irreplicable not only by others but even by the particular introspector himself" (Shanon, 1984). If this is so, a verbal report of introspection only concerns the person who reports at a certain time; it teaches us nothing about other persons, and perhaps not even about oneself at any other time.

This is probably the most serious objection of all, but as we will soon see, the renewed conception of objectivity that arises from a non-representationalist view of science also suffices to meet it.

The challenge is expressed as follows: what do these strange tales told by subjects about their own experience teach us about the world? Isn't their significance restricted to each one of the subjects who provide them? Shouldn't one therefore understand the reluctance of mid-twentieth-century psychology towards the participative, empathic or idiosyncratic aspects of introspection that only worsen the wandering of the science of mind in the swamp of subjectivity? In order to persuade ourselves that this objection is not as devastating as it seems, we can use once again a certain similitude between introspective psychology and microphysics. The questions just raised indeed remind us of two related questions a Copenhagen quantum physicist might have asked. According to Bohr's analysis, each quantum phenomenon is a unique and irreversible event arising from the interaction between a micro-object and a macroscopic measuring apparatus at a certain time; moreover, there are only few and very stringent circumstances in which

the phenomenon can be reproduced when the measurement is repeated on the same object. What do such isolated micro-phenomena teach us about the object *as it is in itself,* independently of the measuring apparatus and its interaction with it? Isn't their significance restricted to single runs of the micro-experiment? This puzzlement by no means hindered the development of quantum mechanics into one of the most powerful physical theories in history. We then just have to find out what, in the methods of physics, made this overcoming of the (virtual) objection possible even before it was formulated.

To begin with, one must remember a consequence of Kant's redefinition of objectivity: objectivity is not something to be found ready-made *out there*, but a project of operational extraction of invariant structures out of a cluster of appearances. So, the issue as to whether or not single events teach us something objective is to be decided on a methodological, *not* on a metaphysical plane. Extracting invariant or covariant structures relies on a process of ascent in generalization and theoretical abstraction, symmetrical of the process of descent which is necessary to reach a nucleus of discourse that can be considered as "factual" or "data-like". In other terms, objectivity is generated ("constituted" writes Kant) by selecting an appropriate level of generality or coarseness, such that invariant structures may be extracted at that level. In the domain of validity of quantum physics, this procedure is implemented thus. One first renounces objectivation at the level of individual phenomena occurring in space-time (this is the reason why the ordinary concept of minute point-like bodies endowed with local properties is in jeopardy). Then, one ascends towards the level of statistical variables. Indeed, the strict reproducibility and indifference to measurement order, is usually missing at the level of individual values, is recovered at the level of their statistics. Finally, one ascends a step further, towards the upper level of formal tools able to generate as many statistics as measurement types, and as many probability assessments as measurement tokens. These formal tools are nothing else than the state vectors in a Hilbert space. State vectors are precisely the maximal invariant structures used by quantum physicists; they therefore play the role of objective entities without bearing the smallest resemblance with our archetypal image of the objects of physics, namely material bodies.

The procedure should be the same for introspection: descent and ascent.

(1) Descent towards minimally interpreted descriptions of the subtlest lived events, without any attempt at asking the *subject* to reconstitute her own cognitive processes (which are actually just as little accessible to subjects as

to scientists), or to explain her "reasons" *in abstracto,* or to stipulate her intended meaning. In other terms, a very careful process of phenomenological reduction must be asked to, or induced in, the introspecting *subjects*.

(2) *A posteriori* ascent of the *scientists* who are analysing the introspective reports construed as data, towards structures generic enough to be seen as stable and invariant across subjects and circumstances. As B. Shanon (1984) cogently pointed out, "While single pieces of data provide only a limited, haphazard view of the phenomenological domain of interest, the corpus in its totality can reveal regular, systematic patterns. The corpus reaches a state in which an increase in the number of tokens ceases to increase the variety of types".

This two-step procedure is exactly the one we apply when we practice the method of elicitation of experience by interviews: (i) guiding subjects towards exquisite contact with their experience and undoing any rational reconstructions that may interfere with their task of description; (ii) retrieving the data extracted from these disciplined descriptions and extracting generic structures out of them.

Conclusion

We gather from these objections and sketchy replies that the most crucial weakness of the introspectionist wave of the turn of the nineteenth and twentieth centuries is likely to have been its unconditional acceptance of the classical, dualist, representationalist theory of knowledge. But since then, many blows have been struck against this theory by contemporary epistemology and cognitive science (Varela, Thompson & Rosch, 1991; Thompson, 2007). It is now time to take this momentous turn into account when dealing with introspection, both by a proper conception of what can be expected from it, and by some concrete methods able to implement this conception. Under a non-dualist/non-representationalist assumption, what is expected from introspection is definitely *not* to monitor the "inner" realm in the same way as natural sciences monitor the "outer" realm. Instead, introspection here becomes just a historic name for a program of changing the focus of attention within the one and all-pervasive field of lived experience, from the narrowly focused state and coarse-grained categories needed by natural sciences to a broader range of interest and refined categories. Introspection should then be aimed at disclosing the initially

unreflected and unattended part of lived experience, and thereby throw light on experienced (yet usually unnoticed) counterparts of the cognitive processes. Ability to bring this information to a satisfactory level of reliability is conditional upon elaborating criteria of mutual performative coherence between the various expressive data obtained in a session of assisted introspection. It also relies on a process of extracting generic structures that have intersubjective value, beyond individual reports. All these features are *de facto* realized by a few recent methods of first-person access, especially by the elicitation interview technique we practice.

References

Bergson, H. (1934*). La pensée et le mouvant*. Paris : P.U.F.

Bitbol, M. (1996). *Schrödinger's Philosophy of Quantum Mechanics.* Dordrecht : Kluwer.
--- (1998). "Some steps towards a transcendental deduction of quantum mechanics". *Philosophia naturalis,* 35, 253-280.
--- (2000). *Physique et philosophie de l'esprit.* Paris: Flammarion.

Block, N. (2011), "Perceptual consciousness overflows cognitive access", *Trends in Cognitive Sciences,* 15, 567-575.

Bode, B.H. (1913). "The Method of Introspection". *The Journal of Philosophy, Psychology and Scientific Methods,* 10, 85-91.

Bohr, N. (1934). *Atomic Theory and the Description of Nature.* Cambridge: Cambridge University Press.

Boring, E.G. (1929). *A history of experimental psychology.* New-York: Appleton-Century-Croft. In Skinner, C. (1935). *Readings in psychology*. New-York: Farrart & Rinehart.

Brentano, F. (1874 / 1944). *Psychologie du point de vue empirique*. Paris: Aubier.

Comte, A. (1830 / 2001). *Cours de Philosophie Positive*. Booksurge Publishing.

Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). "Limits on Introspection. Distorted Subjective Time During the Dual-Task Bottleneck". *Psychological Science,* 19, 1110-1117.

Danziger, K. (1980). "The history of introspection reconsidered". *Journal of the History of the Behavioral Sciences,* 16, 241-262.
--- (1994). *Constructing the Subject, Historical Origins of Psychological Research.* Cambridge: Cambridge University Press.

Dennett, D. (1992). *Consciousness Explained*. Santa Ana, CA: Back Bay Books.

--- (2002). “How could I be wrong? How wrong could I be?” *Journal of Consciousness Studies*, 9, 13–16.

Depraz, N. (2008). *Lire Husserl en phénoménologue*. Paris: P.U.F.

Depraz, N., Varela, F., & Vermersch, P. (2003). *On becoming aware.* Amsterdam: John Benjamins.

Flajoliet, A. (2006). Husserl et Messer. *Expliciter: Journal de l’association GREX (Groupe de recherche sur l’explicitation)*, 66, 1-32.

Galison, P. (1987). *How Experiments End.* Chicago: The University of Chicago Press.

Garfield, J.L. (1989). “The Myth of Jones and the Mirror of Nature: Reflections on Introspection”. *Philosophy and Phenomenological Research.* 50, 1-26.

Gendlin, E. (1962). *Experiencing and the Creation of Meaning*. Chicago: Northwestern University Press.

Genoud, C. (2009). “On the Cultivation of Presence in Meditation”. *Journal of Consciousness Studies,* 16, 117-128.

Goldman, A..I. (2001). “Epistemology and the evidential status of introspective reports”. in Jack A. & Roepstorff A. (ed.) (2001). *Trusting the subject* 2?. Imprint Academic.

Gooding, D. C., Gorman, M., Tweney, R., & Kincannon, A. (Eds.). (2005). *Scientific and Technological Thinking.* Mahwah, NJ: Erlbaum.

Gopnik, A. & Meltzoff, A.N. (1994). “Minds, bodies and persons: Young children’s understanding of the self and others as reflected in imitation and "theory of mind" research”. In: Parker S. & Mitchell R. (eds.) (1994). *Self-awareness in animals and human.* Cambridge: Cambridge University Press.

Hassin, R.R., Uleman, J.S. & Bargh, J.A. (eds.) (2006). *The New Unconscious.* Oxford: Oxford University Press.

Hendricks, M. (2009). “Experiencing level: An instance of developing a variable from a first person process so it can be reliably measured and taught”. *Journal of Consciousness Studies,* 16, 129-155.

Høffding, H. (1905). *The Problems of Philosophy*. London: The MacMillan Company.

Hume, D. (1739 / 1978). *A treatise of human nature.* Oxford: Oxford University Press.

Humphrey, G. (1951). *Thinking.* London: Methuen.

Hurlburt, R.T. (1990). *Sampling normal and schizophrenic inner experience.* New-York: Plenum Press.

Hurlburt, R.T. & Heavey C.L. (2001). “Telling what we know: describing inner experience”. *Trends in Cognitive Sciences* 5, 400-403.

Hurlburt, R.T. & Heavey C.L. (2006). *Exploring inner experience.* Amsterdam: John Benjamins.

Husserl, E. (1913 / 2004). *Ideas: General Introduction to Pure Phenomenology.* London: Routledge.

--- (2002). *Husserliana, Edmund Husserl Gesammelte* Werke, XXXIV, *Zur phänomenologischen Reduktion.* Dordrecht: Kluwer.

Jack, A., Roepstorff, A. (2002). "The 'measurement problem' for experience: damaging flaw or intriguing puzzle?". *Trends in Cognitive Sciences* 6, 372-374.

James, W. (1890). *The Principles of Psychology*. Baltimore: Holt.

Johansson, P., Hall, L., Sikström, S., Tärning, B. & Lind, A. (2006). "How something can be said about telling more than we can know: On choice blindness and introspection". *Consciousness and Cognition* 15, 673–692.

Kant, I. (1786/2002). *Metaphysical Foundations of Natural Science*, in: *Theoretical Philosophy After 1781*. in: Allison, H., Brandt, R., Guyer, P., Meerbote, R., Parsons, C. D., Robinson, H., Schneewind, J. B. &. Wood, A. W. (eds.). (2002). *The Cambridge Edition of the Works of Immanuel Kant in Translation*. Cambridge: Cambridge University Press.
--- (1800 / 1988). *Logic.* New-York: Dover.

Kriegel, U. (2013), "A hesitant defense of introspection", *Philosophical Studies* (Forthcoming)

Lachaux, J.-P. (2011), "If no control, then what ? Making sense of 'neural noise' in human brain mapping experiments using first-person reports", *Journal of Consciousness Studies,* 18, 162-166.

Locke, J. (1690 / 1975). *An essay concerning human understanding.* Oxford: Oxford University Press.

Marcel, A. (2003), "Introspective report", *Journal of Consciousness Studies,* 10, 167-186

Merleau-Ponty, M. (1989). *Éloge de la Philosophie*. Paris: Gallimard.

Moore J.S. & Gurnee H. (1935). *The foundations of psychology*. Reprint in Skinner, C. *Readings in psychology,* New-York: Farrar & Rinehart.

Nahmias, E.A. (2002). "Verbal Reports on the Contents of Consciousness: Reconsidering Introspectionist Methodology". *Psyche*, 8 (21).

Natorp, P. (1912) *Allgemeine Psychologie nach kritischer Methode*. Tübingen: J.C.B. Mohr. French translation by Dufour, E. & Servois, J. (2007). *Psychologie générale selon la méthode critique*. Paris: Vrin.

Nisbett, R.E. & Wilson T.D. (1977). "Telling more than we can know: Verbal reports on mental processes". *Psychological Review,* 84, 231-259.

Ogden, R.M. (1911). "Imageless thought: résumé and critique". *Psychological Bulletin*, 8, 194.

Overgaard, M., Koivisto, M., Sørensen, T.A., Vangkilde, S., Revonsuo, A. (2006). "The electrophysiology of introspection". *Consciousness and Cognition,* 15, 662-672.

Petitmengin, C. (2006). "Describing one's subjective experience in the second person: An interview method for the science of consciousness". *Phenomenology and the Cognitive Science,* 5, 229–269.

Petitmengin, C., Navarro, V., Baulac, M. (2006). "Seizure anticipation: Are neuro-phenomenological approaches able to detect preictal symptoms?". *Epilepsy and Behavior,* 9, 298-306.

Petitmengin C. & Bitbol M. (2009). "The Validity of First-Person Descriptions as Authenticity and Coherence". *Journal of Consciousness Studies,* 16, 363-404.

Petitmengin, C., Bitbol, M., Nissou, J-M., Pachoud, B., Curallucci, H., Cermolacce, M., & Vion-Dury, J. (2009). "Listening from within". *Journal of Consciousness Studies*, 16, 252-284.

Petitmengin C., Remillieux A., Cahour B., Thomas S. (2013). A gap in Nisbett and Wilson findings? A first-person access to our cognitive processes, *Consciousness and Cognition* (Forthcoming)

Piccinini, G. (2003). "Data from Introspective Reports". *Journal of Consciousness Studies*, 10, 141-156.

Piccinini, G. (2009). "First-Person Data, Publicity, and Self-Measurement". *Philosophers Imprint,* 9, 2009.

http://hdl.handle.net/2027/spo.3521354.0009.009

Pickering, A. (1995). *The Mangle of Practice*. Chicago: The University of Chicago Press.

Prinz, J. (2004). "The fractionation of introspection". *Journal of Consciousness Studies*, 11, 40–57.

Robbins, P. (2004). "Knowing Me, Knowing You. Theory of Mind and the Machinery of Introspection". *Journal of Consciousness Studies,* 11, 129-143.

Roustang F. (2009). *Le secret de Socrate pour changer la vie.* Paris : Odile Jacob.

Rudrauf, D., Lutz A., Cosmelli D., Lachaux J.-P., and Le Van Quyen M. (2003). "From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being". *Biological Research*, 36, 21-59.

Sackur, J. (2009). "L'introspection en psychologie expérimentale". *Revue d'Histoire des Sciences,* 62 , 5-28.

Sartre, J.P. (2000). *La Transcendance de l'ego.* Paris: Vrin.

Schooler, J.W. (2002). "Re-representing consciousness: Dissociations between experience and meta-consciousness". *Trends in Cognitive Sciences*, 6, 339-344.

Schwitzgebel, E. (2004). "Introspective training apprehensively defended: Reflections on Titchener's lab manual". *Journal of Consciousness Studies,* 11, 58–76.
--- (2011), *Perplexities of consciousness,* MIT Press.

Shanon, B. (1984). "The Case for Introspection". *Cognition and Brain Theory,* 7, 167-180.

Silverman, M. & Mack, A. "Change blindness and priming: When it does and does not occur". *Consciousness and Cognition*, 15, 409-422.

Sperling, G. (1960). “The information available in brief visual presentations”, *Psychological Monographs* 74 (9), 1-29.

Ten Elshof, G. (2005). *Introspection Vindicated; An Essay in Defense of the Perceptual Model of Self Knowledge.* Aldershot: Ashgate.

Ten Hoor, M. (1932). “A Critical Analysis of the Concept of Introspection”. *The Journal of Philosophy*, 29, 322-331.

Thompson, E. (2007). *Mind in Life*. Cambridge: Harvard University Press.

Titchener, E.B. (1909). *Lectures on the experimental psychology of thought processes*. New-York: MacMillan.
--- (1912). “The Schema of Introspection”. *American Journal of Psychology,* 23, 485-508.
--- (1910/1890). *A Textbook of Psychology.* New York: Scholars’ Facsimiles and Reprints.
--- (1916). *A Textbook of Psychology*. New-York: McMillan.

Tye, M. (2009). *Consciousness revisited.* Cambridge: MIT Press.

Varela, F.J., Thompson, E., Rosch, E. (1993). *The Embodied Mind: Cognitive Science and Human experience*. Cambridge: MIT Press.

Vermersch, P. (1994). *L’entretien d’explicitation.* Paris: ESF.
--- (1999). “Introspection as practice”. *Journal of Consciousness Studies,* 6, 15-42.

Wallace, B.A. (1998). *The Bridge of Quiescence.* Chicago: Open Court Press.
--- (2000). *The Taboo of Subjectivity: Towards a New Science of Consciousness.* Oxford: Oxford University Press.
--- (2006). *Contemplative Science*. New-York: Columbia University Press.

Watson, J.B. (1913). “Psychology as the Behaviorist Views it”. *Psychological Review,* 20, 158-177.

Wittgenstein, L. (1964 / 1980). *Philosophical Remarks*. Chicago: The University of Chicago Press.

Woodworth, R. S. (1906). “Imageless Thought”. *The Journal of Philosophy, Psychology and Scientific Methods,* 3, 701-708.

Wundt, W. (1901). *Lectures on Human and Animal Psychology.* London: Swan Sonnenshein & Co. (translated from the second German edition of 1896).

La psycho-phénoménologie, théorie de l'explicitation

Maryse Maurel
(CESAME, IREM, IUFM, Nice - France)
m.maurel04@wanadoo.fr

Ce texte a pour but de donner un aperçu de la partie de la psycho-phénoménologie qui fonde l'entretien d'explicitation, ses techniques et son utilisation.

Dans une première partie, je montrerai qu'il y a eu pour moi au départ un besoin professionnel, celui de la *description de la subjectivité*[1] en première ou en deuxième personne et je décrirai rapidement l'entretien d'explicitation ; dans une deuxième partie j'aborderai l'objet de mon intervention en développant quelques aspects théoriques de l'explicitation qui permettent de fonder la possibilité et l'efficacité des actes de *l'introspection rétrospective*[2] pour accéder à un vécu passé - ou ressouvenir - et le décrire.

Je parle ici en tant que praticienne et chercheure qui, dans sa pratique et dans ses recherches, a utilisé un outil, l'entretien d'explicitation, et qui s'est intéressée à ses fondements théoriques [1].

[1] Les mots en italiques sont les mots du vocabulaire spécifique de l'explicitation. Je m'efforcerai, tout au long de ce texte, d'en donner une définition simplifiée au risque d'utiliser un langage un peu familier. Il se peut que la définition ne suive pas immédiatement la première occurrence du mot.

Le mot subjectivité peut être entendu au sens de monde intérieur ou de pensée privée selon le monde dans lequel vous évoluez.

[2] Il ne s'agit pas de pratiquer une introspection tout en continuant à vivre, il s'agit de se retourner vers un moment du passé dans le but de le décrire.

1. Un besoin professionnel et un outil : l'entretien d'explicitation

1.1 Au début, un besoin

Je travaille au sein de deux groupes de recherche, le GREX et l'IREM de Nice. Le GREX est le Groupe de Recherche sur l'Explicitation. Il fonctionne depuis plus de vingt ans sous la responsabilité scientifique de Pierre Vermersch [9]. L'IREM est l'Institut de Recherche sur l'Enseignement des Mathématiques de l'université de Nice où j'ai été chercheure en didactique des mathématiques et enseignante. Autour des années 90, j'ai quitté un poste de professeur de lycée pour rejoindre l'université de Nice (IREM de Nice et enseignement en première année d'université). La mission que j'avais reçue de l'institution était celle d'une meilleure intégration des étudiants arrivants. Le groupe de recherche de didactique des mathématiques dont je faisais partie ne cherchait pas à produire des cours d'enseignement initial qui sont ce qu'ils sont dans les contraintes imposées par l'enseignement au fil des changements de programme. Nous cherchions plutôt à connaître l'état des connaissances d'un élève ou d'un étudiant de mathématiques, à travers les erreurs qu'il produit, pour l'aider à reconstruire à partir de là des connaissances plus adéquates du point de vue mathématique et plus conformes à celles attendues dans le système scolaire. Dans ce groupe, l'examen des copies, brouillons et tests ne nous fournissait pas des renseignements suffisants pour nos recherches et nous pensions déjà que *seul le sujet sait ce qu'il a fait et comment il l'a fait* à la condition de pouvoir y accéder. Nous postulions que *le sujet a une cohérence interne*, qu'il est capable d'apprendre et qu'il ne fait pas n'importe quoi. D'où, dès le début, dès 1990, mon intérêt pour la technique d'entretien proposée par Pierre Vermersch, ma formation à cette technique et l'adhésion au groupe GREX, financé au début par le gouvernement, puis constitué en association après l'interruption du financement.

Le but de Pierre Vermersch, au début de ce programme de recherche, était

> De « prendre en compte le point de vue du sujet et s'intéresser à la cognition subjective. […] Plutôt que de déclarer a priori la cognition subjective non étudiable scientifiquement, la question qui sera posée c'est de savoir : à quelles conditions peut-on l'étudier objectivement ?
>
> La problématique de l'explicitation est centrée sur la cognition dans l'action.
>
> L'explicitation est la mise à jour, par la verbalisation du sujet, des connaissances implicites contenues dans l'action". ».

(Rapport GREX, Ministère de la Recherche et de la Technologie, mai 1992)

J'ai adhéré à ce programme de travail. Pour moi, l'entretien d'explicitation me permettait d'accéder aux actions mathématiques d'un étudiant apprenant des mathématiques, en évitant les interprétations, pour décrire sa logique interne et les connaissances inscrites dans ses actions.

Au début, j'ai mené des entretiens d'explicitation de recherche pour voir l'objet "apprentissage de l'algèbre élémentaire" de plus près, pour l'explorer, pour le constituer plus précisément en objet d'étude. J'ai mené aussi des entretiens auprès de certains étudiants volontaires pour en savoir plus sur leurs processus d'apprentissage des mathématiques. Je voulais aussi tester les potentialités de l'outil "explicitation". Je ne pouvais imaginer à ce moment – là que je pourrais en faire une utilisation en classe – sentiment de ridicule, risque d'intrusion dans la pensée privée des étudiants. Et un jour, j'ai osé, j'ai osé demander à un étudiant en séances de Travaux Dirigés[3] : « Et quand vous ne trouvez rien, vous trouvez quoi ». Contourner le déni pour chercher l'information sur ce qu'il a fait parce qu'il a nécessairement fait quelque chose, c'est ce que je venais d'apprendre à faire dans ma formation aux techniques de l'explicitation. Et là, stupéfaction de ma part, l'étudiant ne sourit pas, il n'est pas étonné, il ne me regarde pas comme si je disais des choses bizarres et il parle, il parle, je l'accompagne, il décrit tout ce qu'il a fait sans pouvoir arriver à des résultats identifiables par lui comme tels. Diagnostic interne de ma part et proposition de travail dans sa direction.

C'est ainsi que j'ai importé l'explicitation dans mon enseignement en l'intégrant à ma façon de travailler et en l'adaptant à un travail collectif dans la classe. Et c'est le début d'une longue histoire où j'ai suivi un chemin de travail et de passion [4].

Je précise ici que l'entretien d'explicitation a toujours été pour moi un outil de recueil d'informations et d'aide au diagnostic et que j'ai trouvé ailleurs, dans mon expertise d'enseignante et de chercheure en didactique des mathématiques, les outils d'intervention.

1.2. Le point de vue en première personne, point de vue fondateur

L'explicitation s'inscrit dans un choix délibéré : prendre en compte le point de vue du sujet et s'intéresser à toutes les couches de la subjectivité : Qui

[3] Les Travaux Dirigés sont des séances de 1h30 ou 2h pour faire des exercices et travailler à l'assimilation du cours.

parle ou fait ? De quoi parle-t-il ? Comment fait-il pour en parler ? Que fait-il ? Quelle est la chronologie de ses actions ? Quelles sont ses prises d'informations sensorielles ou autres ? Quels sont ses états internes ? Quelles sont ses croyances ? Quelle est la tonalité émotionnelle ? Où sont localisés les ressentis internes ?

Dans ma pratique professionnelle je me suis limitée à *la couche des actions du sujet pour obtenir les connaissances inscrites dans l'action* [15].

Or, prendre en compte le point de vue du sujet, c'est faire le choix épistémologique du *point de vue en première personne*.

Le *point de vue en troisième personne* est celui d'un observateur comme les chercheurs de la psychologie expérimentale. La personne est l'objet de mon attention et de mon observation ; je lui propose des tâches, des tests ou des questionnaires ; je recueille ses réponses et ses comportements, éventuellement enregistrés ou filmés. Pendant tout le 20ème siècle, la psychologie a été construite sur ce qui est observable et enregistrable, c'est ce que nous appellerons le *point de vue en troisième personne*, c'est-à-dire le point de vue classique d'un observateur extérieur qui prend le sujet comme objet d'étude sans s'intéresser à sa vision du monde. Notons que ce point de vue ne s'occupe pas de la façon dont le sujet va chercher les informations pour répondre aux tests ou aux questionnaires.

Adopter un *point de vue en première personne*, c'est développer une science du sujet, une science de la vie subjective, du point de vue de celui qui la vit. S'il s'agit d'étudier mon expérience personnelle, je suis la seule à pouvoir y accéder et donc à pouvoir la décrire en produisant un point de vue en première personne. Mais l'accès rigoureux du sujet à ses propres vécus demande le dépassement d'un certain nombre d'obstacles :

> Vivre l'expérience subjective est spontané, sans préalable ni conditions ; décrire, analyser l'expérience subjective est une expertise. (Pierre Vermersch, Pour une psycho-phénoménologie, Expliciter n° 13, Février 1996, p. 1.)

Autrement dit, en termes plus imagés, la fréquentation d'un jardin ne donne pas la compétence du jardinier pour faire pousser les fruits et les légumes. Il faut savoir accéder à son expérience subjective et il faut disposer de catégories pour la décrire.

Dans le cas où le sujet ne possède pas cette expertise, les productions d'une introspection spontanée sont très pauvres et il faudra l'accompagnement d'un questionneur expert pour obtenir des informations utiles à ses buts professionnels – pratique ou recherche – ou aux buts de

formation du sujet ; cet expert recueillera ainsi un *point de vue en deuxième personne*. L'entretien d'explicitation permet au chercheur ou au praticien de recueillir ce point de vue en deuxième personne, en aidant le sujet à décrire son monde intérieur et en recueillant les verbalisations ainsi produites pour les travailler ensuite.

Le point de vue en première personne est celui de l'expert. Comme tout autre individu il est le seul à avoir accès à sa subjectivité et il peut en faire la description selon les catégories de son objet d'étude parce qu'il sait le faire.

> Le seul vécu auquel [un sujet] ait intimement accès sur le mode direct est le sien, les autres ne seront jamais qu'une interprétation basée sur une empathie. Dans les deux cas, cependant, ce qui peut être pris en compte pour la recherche c'est ce qui peut être verbalisé, ce qui produit des données objectivables, et ce qui peut être verbalisé de son propre vécu dépend de la possibilité de le conscientiser. (Expliciter n°39, Conscience directe et conscience réfléchie, Vermersch P., mars 2001, page 10.)

Je précise bien qu'il ne s'agit pas de se passer du point de vue en troisième personne, qu'il s'agit ici de le compléter par des données issues du point de vue en première ou en deuxième personne, en vérifiant leur compatibilité, chacun des points de vue enrichissant l'autre et le validant [5].

1.3. L'entretien d'explicitation brièvement

L'entretien d'explicitation vise la description d'une situation passée, temporellement indexée[4]. La pratique de l'entretien d'explicitation s'apprend *expérientiellement*[5], c'est-à-dire en le faisant et en le pratiquant, dans des stages de durée minimale cinq jours où est défini le vocabulaire de l'explicitation et où il prend tout son sens. Pour résumer brièvement, l'entretien d'explicitation est un ensemble de techniques de questionnement et d'accompagnement à une introspection rétrospective guidée sur une situation spécifiée passée [6] pour :

- passer un contrat, vérifier le consentement du sujet,
- lancer une *intention éveillante* vers une situation spécifiée passée qui convient pour le but de l'entretien,

[4] Il ne s'agit pas de décrire ce que je fais le matin quand je prépare mon thé, mais ce que j'ai fait très exactement dimanche dernier, par exemple, quand j'ai préparé mon thé. C'est pour cela que nous qualifions cette situation de "situation spécifiée".

[5] Un apprentissage expérientiel est un apprentissage où l'on fait soi-même l'expérience de ce que l'on apprend.

- accueillir le produit du *réfléchissemen*t (ensemble des actes par lequel le sujet accède à un vécu passé et s'en fait un quasi revivre afin de le décrire) et aider à la mise en mots (verbalisation) avec des questions ouvertes (appelées *relances*[6] qui accompagnent le sujet dans sa pensée, qui guident son exploration du vécu passé),

- amener le sujet dans une position particulière de parole dite *position d'évocation*,

- focaliser les relances sur la description de l'action : il y a de la connaissance dans l'action, Piaget l'a dit [15], les ergonomes le savent bien, parce qu'ils vérifient sur le terrain qu'il y a toujours un écart entre l'action professée et l'action agie,

- accueillir l'émotion si elle vient, et relancer sur l'action car nous ne sommes pas des psychothérapeutes.

Il est impossible de prévoir à l'avance ce qui va venir, l'accompagnement doit être très souple et très ouvert et surtout non inductif.

Comme nous le disons au GREX, « Je suis toujours étrangère à la subjectivité de l'autre ».

2. L'explicitation devient un objet d'étude

L'entretien d'explicitation s'est constitué comme pratique. La psycho-phénoménologie a été élaborée pour fonder cette pratique et pour décrire le monde intérieur d'un sujet, qui devient ainsi accessible grâce à cette pratique [7] [8].

Qu'est-ce qui permet de rendre compte du fonctionnement de l'entretien d'explicitation, autrement dit de fonder cette pratique ?

2.1. Un modèle de la conscience

Un modèle de ce que l'on veut étudier permet de générer de nouvelles questions, d'orienter le regard dans la bonne direction pour apercevoir des propriétés qui ne se révèlent que si on a l'idée de les questionner et si on a des mots pour les décrire, ce qui est la fonction de tout modèle théorique.

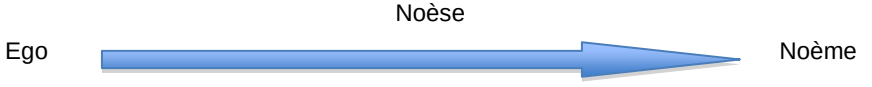
Ce modèle de la conscience a pour but d'éclairer les techniques d'accès à l'évocation d'un vécu passé spécifié, la possibilité de l'émergence du quasi-

[6] Nous appelons 'relances' et non 'questions' les interventions du questionneur pour bien marquer le caractère ouvert et non inductif de ces interventions.

revivre de ce passé évoqué, le choix des relances pour accompagner le sujet vers la position d'évocation et pour obtenir les informations recherchées et tout ce qui caractérise l'aspect très technique de cet entretien.

Dans la préface de son ouvrage "La mémoire, l'histoire et l'oubli", Paul Ricœur, philosophe phénoménologue récemment disparu, pose les trois questions suivantes : de quoi y a-t-il souvenir ? De qui est la mémoire ? Comment se souvenir ? [16] Se souvenir, c'est avoir un souvenir, être un sujet qui vise ce souvenir mais aussi faire quelque chose pour se mettre en quête du souvenir. Dans ces trois questions se trouve résumée la structure tripartite de la conscience : qui ? (Je ou ego), quoi ? (contenu ou noème), comment ? (actes ou noèse).

La conscience phénoménologique étant toujours intentionnelle, c'est-à-dire toujours conscience de quelque chose, il y a toujours un Je, un ego, qui tourne son attention vers - qui vise – qui prend comme objet attentionnel - quelque chose. C'est le *pôle égoïque*. Le quelque chose est le contenu de la visée, l'objet attentionnel ou *pôle noématique* et les actes de la visée constituent les actes attentionnels, le *pôle noétique*.



Selon le but de l'entretien, nous sommes amenés à questionner sur l'une ou l'autre des parties de cette flèche : à quoi fais-tu attention (noème) ? Que fais-tu quand tu fais attention à cela (noèse) ? Qui es-tu quand tu fais cela (ego) ?

L'apport de Husserl [13] est d'avoir proposé la notion de préréfléchi (irréfléchi pour Sartre).

Nous pouvons alors distinguer (au moins) deux modalités de la conscience : la conscience directe et la conscience réfléchie : nous parlons de *conscience réfléchie* pour désigner ce que je sais en sachant que je le sais, c'est-à-dire ce qui est déjà conscientisé ; nous parlons de *conscience directe* pour désigner ce que je sais sans savoir que je le sais, c'est ce que nous nommons *préréfléchi* ; au moment où l'information me parvient, elle m'affecte mais je ne suis pas présente à sa saisie.

L'intérêt de distinguer le *préréfléchi* du *réflexivement conscient* est de ne pas considérer ce dont je ne suis pas consciente réflexivement - ce que je ne sais pas que je sais - comme une absence, mais comme ce que je n'ai pas

encore conscientisé, comme quelque chose pour lequel j'ignore encore la possibilité de l'accès, comme quelque chose qui est disponible si je sais l'éveiller et l'accueillir.

En résumé, l'entretien d'explicitation permet l'accès au préréfléchi et permet de décrire ce que je pense ignorer.

2.2. La passivité

Le modèle de la passivité selon Husserl, justifie la possibilité d'accès à des données auxquelles nous pensons ne pas pouvoir accéder ; en effet à tout moment, sans être conscients de le faire, nous engrangeons des informations sur nos vécus, sur tout ce qui nous affecte et dont nous ne sommes pas nécessairement conscients au moment où nous sommes affectés. On peut en rendre compte de différentes façons, en invoquant le modèle de la passivité chez Husserl [14], la mémoire concrète autobiographique selon Gusdorf [12], ou encore le modèle organismique des praticiens du focusing[7] [17] [11]. Tous ces travaux arrivent à la même conclusion : je sais beaucoup plus de choses que je ne le crois, et une partie de ma mémoire se joue sur le mode préréfléchi.

Ce qui est sédimenté dans le champ de ma passivité est le fruit des rétentions et synthèses successives et permanentes. Le processus des rétentions agit à mon insu. Le processus des synthèses passives apparaît en négatif quand il ne fonctionne plus comme dans la maladie d'Alzheimer.

[7] La dimension expérientielle (dimension organismique) est repérée par Rogers comme étant une source interne d'informations. Cette affirmation, Gendlin cherche à lui offrir une théorie : la théorie de l'experiencing qu'il modélise en termes de processus.
Cette dimension « ne renvoie pas seulement à la structure physique et biologique de l'individu, mais à l'individu en tant que totalité psycho-physique interagissant comme un tout avec son environnement. » (note du traducteur du « développement de la personne » -1966)
Le terme « organismique », souvent utilisé par Rogers (Le développement de la personne, 1966), renvoie à la notion d'experiencing, puisqu'il s'agit de ce qui est corporellement vécu et ressenti, en rapport avec le contexte relationnel. Gendlin parle plus souvent d'experiencing, mais il emploie aussi ce terme, en référence à Rogers, dans certaines expressions, « organismic knowing » en équivalence à « experiential knowing » (savoir organismique, expérientiel), « organismic experiencing » (« expérience organismique »).

Un exemple de rétention : l'exemple du son de Husserl

Avoir dans l'oreille un son qui n'est plus émis, ce n'est plus de la perception, c'est déjà de la mémoire. C'est le phénomène de *rétention* qui dure plus ou moins longtemps selon les personnes. Puis, après ce temps de rétention, le vécu auditif "sombre" dans l'oubli. Dans un premier temps, je peux encore en retrouver le souvenir, mais je ne peux plus le retrouver "comme si je l'entendais encore". Puis je l'oublie. Le vécu de ce son est alors au degré d'activité zéro. Attention, il n'a pas disparu, il est seulement inactif. Il n'est pas mort, il est endormi et peut être réveillé.

En résumé, l'éveil est possible par principe. Il reste à voir comment éveiller un ressouvenir.

2.3. L'éveil d'un ressouvenir

Est-il possible d'accéder aux informations contenues dans un vécu ancien que l'on croit oublié complètement ou en partie ?

L'éveil involontaire

Cette mémoire issue des synthèses passives successives peut être réveillée à tout moment de façon inopinée. C'est le goût de la madeleine qui a déclenché involontairement pour Proust l'accès à son enfance. Le goût de la friandise lui était associé. Nous avons tous fait ce genre d'expériences et de nombreux exemples existent dans la littérature.

L'éveil provoqué

Pouvons-nous en rendre l'accès délibéré, pouvons-nous créer des madeleines proustiennes à la demande ? La réponse est oui. C'est ce que montrent les nombreux exemples recueillis en entretien d'explicitation [2] [3]. Ce n'est donc pas une question théorique. Nous constatons dans les entretiens d'explicitation que nous conduisons que le questionné retrouve des informations passées qu'il ne pensait pas pouvoir retrouver. Tout se passe comme s'il ne savait pas d'avance qu'il disposait des informations que nous cherchions. Comment expliquer cette possibilité sur le plan théorique ?

Quand le maître dit à un élève : « Je vous propose de prendre le temps de revenir sur ce qui s'est passé mardi dernier quand vous étiez en devoir surveillé et de laisser revenir ce qui vous revient comme ça vous revient »,

nous disons que la maître lance une *intention éveillante* en proposant à l'élève de viser ce moment précis de mardi où il était en devoir surveillé. L'élève l'a vécu et l'a mémorisé, au moins de façon passive. Très souvent, au début, la visée est vide, c'est-à-dire que rien ne revient. Si le maître contourne le déni[8] et cherche des éléments sensoriels ou des éléments de contexte, l'un de ces éléments va être activé, nous ne savons pas lequel, mais ce sera le premier fil à tirer pour opérer le réfléchissement de ce vécu passé, c'est-à-dire le passage de certains éléments de ce vécu de la conscience préréfléchie à la conscience réfléchie. Nous parlons de *réfléchissement* ou *d'acte réfléchissant* pour désigner l'ensemble de ces actes.

Il y a alors *remplissement intuitif* du ressouvenir, intuitif au sens de Husserl, parce que les informations - sensorielles et cognitives par exemple - arrivent sur un mode non loquace, non encore mis en mots au niveau de la conscience réfléchie du sujet.

C'est ensuite au maître de faire l'accompagnement adéquat pour aider l'élève à verbaliser et pour diriger son attention vers ce qui est pertinent dans l'entretien en cours, selon le contrat passé entre le maître et l'élève, pour compléter le remplissement.

Mon ressouvenir a existé indépendamment de sa conscientisation ; quand j'aurai conscientisé mon ressouvenir, je l'aurai recréé sous une forme sémiotique qui m'est personnelle et quand je l'aurai mis en mots, il pourra être étudié et partagé au sein d'une communauté de praticiens ou de chercheurs ; pour un élève, il pourra être travaillé avec le maître.

Notons bien la différence du rapport que j'entretiens avec mon passé entre l'acte de *penser à mon vécu* et l'acte de *rendre de nouveau présent ce vécu* dans la position d'évocation.

Notons aussi que la perception est acquisition de l'objet tandis que le ressouvenir de l'objet, son évocation, est re-présentation de celui-ci.

Ce passage creuse la différence entre perception et ressouvenir, il ne s'agit jamais dans le ressouvenir d'une "présence en chair et en os" que seule la perception peut donner, mais d'un "passé en chair et en os". Ce qui est posé, c'est donc la relation entre perception comme acte originaire dans la présence et le ressouvenir qui n'est plus un acte originaire de connaissance

[8] Si l'élève dit :
- Je ne me souviens de rien,
le maître peut continuer en disant :
- Et quand vous ne vous souvenez de rien, qu'est-ce qui vous revient en premier ? ou De quoi vous souvenez-vous ?

de l'objet, mais un acte originaire de redisposition, de réactivation de l'acte perceptif passé et de ce qui a été perçu. Le ressouvenir introduit une modification de la conscience, il s'agit toujours du même objet mais en tant que perception passée.

En résumé, c'est *l'éveil provoqué* d'un ressouvenir par *une intention éveillante* et le *remplissement intuitif* de ce ressouvenir par *l'acte réfléchissant* qui nous donnent accès à des informations préréfléchies.

Le rappel ne peut se faire par un acte volontaire, au risque de se lancer un défi de mémoire [18] [19], mais par une posture d'accueil, de lâcher prise, de laisser venir et de saisir par contiguïté les éléments du ressouvenir qui vont le compléter à partir du premier élément contacté.

2.4. Les effets perlocutoires et le travail sur les relances

Les *effets perlocutoires* sont les effets que je fais à l'autre avec mes mots [10], et en particulier avec mes relances quand je l'accompagne dans un entretien d'explicitation.

Exemple de la phrase rituelle

J'avais pris l'habitude dans mon enseignement de commencer toutes les séances de Travaux Dirigés (séances d'exercices) par une *phrase rituelle* :

« Et maintenant je vous propose de prendre un moment, chacun pour vous, pour laisser revenir ce qui vous revient, comme ça vous revient quand vous prenez le temps d'y penser, de ce qui s'est passé dans la séance de la semaine dernière et de ce que vous avez fait depuis, chez vous ou ailleurs, des questions que vous vous êtes posés, des choses que vous avez comprises, des difficultés que vous avez rencontrées ou tout autre chose qui vous intéresse ; et quand vous serez prêts, nous pourrons en parler ensemble. »

Par cette phrase, toujours la même, qui n'étonne personne, j'invite les étudiants à se tourner vers leur monde intérieur, à faire une visée de la séance précédente pour laisser revenir ce qui leur revient. Je leur propose ainsi de suspendre tout ce qu'ils sont en train de faire, de lâcher les préoccupations du moment, pour diriger toute leur attention vers leur activité mathématique de la semaine et leur vécu de cette activité. C'est une façon de lancer une intention éveillante au champ de leur passivité, d'induire un lâcher prise, une posture d'accueil, c'est une façon de désigner l'objet attentionnel (mon activité mathématique depuis une semaine) et, ce qui n'est pas

négligeable, c'est aussi une façon de leur proposer de se tourner vers eux-mêmes, de laisser de côté la posture naturelle pour adopter une posture "mathématique", d'où le silence qui s'installe de façon indirecte, sans crier, sans exiger. C'est une façon de donner une consigne de travail pour obtenir en même temps le silence sans avoir à le demander.

La phrase magique de l'entretien d'explicitation :

Nous commençons tous nos entretiens d'explicitation par la négociation du contrat qui fixe l'objectif assigné à l'entretien, par la vérification de l'accord du questionné et par la *phrase magique* : « Et maintenant, si vous en êtes d'accord, je vous propose de prendre le temps de laisser revenir ce qui vous revient de ce moment où … vous prenez tout le temps qu'il vous faut et quand vous serez prêt vous me ferez un signe. »

Nous choisissons ensuite la phrase suivante en fonction du but de l'entretien et nous pouvons maintenant mesurer les effets différents de « qu'est-ce qui vous vient en premier de ce moment ? » et de « qu'est-ce qui est le plus important pour vous dans ce moment ? » selon que le but de l'entretien est de reconstituer la chronologie ou de saisir ce qui fait sens pour le sujet.

Il y a des mots à éviter absolument comme « essayez de vous souvenir » qui oriente le sujet vers un acte volontaire alors qu'il est important d'être dans une posture de lâcher prise et d'accueil de ce qui va venir et que nous ne connaissons pas encore.

Il y a des mots dont nous connaissons bien les effets parce que nous les avons étudiés sur nous, entre nous. En séminaire expérientiel, nous avons pris l'habitude de continuer ce travail en demandant à notre sujet de faire des retours en temps réel sur l'effet de nos mots sur lui.

Des phrases classiques :

Et quand vous ne faites rien, qu'est-ce que vous faites ? (contournement du déni).

Et quand vous dites que c'est terminé, comment vous savez que c'est terminé ? (recherche du critère de fin).

Et quand vous dites que vous avez compris, qu'est-ce que vous avez compris ? (recherche du critère compréhension).

Et lorsque vous voyez ce que vous voyez, y a-t-il autre chose ? (induction à changer la direction attentionnelle vers ce qui n'est pas au centre du champ attentionnel).

Conclusion

Que retenir en priorité de cet immense champ de connaissance que dévoile la psycho-phénoménologie ?

- L'importance de la rupture épistémologique qui nous amène à nous intéresser au point de vue en première personne ;
- l'existence du préréfléchi et de ce réservoir de mémoire précieux pour chacun de nous qu'est le champ de notre passivité ;
- la possibilité de l'éveil provoqué d'un ressouvenir ;
- la preuve pragmatique par les entretiens que nous faisons, que nous analysons et que nous publions, de cette possibilité ;
- l'importance des effets perlocutoires dans l'accompagnement d'un sujet qui décrit sa subjectivité et la précision nécessaire de l'ajustement des relances pour obtenir l'effet recherché et accompagner la posture d'accueil.

Bibliographie

Sur le thème de l'explicitation

[1] MAURELM., (2008), La psycho phénoménologie, théorie de l'explicitation, *Expliciter* 77, pp 1-29. Sur le site www.grex2.fr

[2] MAUREL M., (2009), The explicitation interview: examples and applications, in Ten years of viewing from within, *Journal of Consciousness Studie*s, 16 (10-12), edited by PETITMENGIN C., pp 58-89, Exester EX5 5YX, UK.

[3] MAUREL M. (2009), L'entretien d'explicitation, exemples et applications, *Expliciter* 80, pp 1-17. Sur le site www.grex2.fr (version française du précédent)

[4] MAUREL M., SACKUR C., (2010), *Faire l'expérience des mathématiques. Entre enseignement et recherche*, ALÉAS, Lyon. (Pour une utilisation dans l'enseignement)

[5] PETITMENGIN C., (2004), Peut-on anticiper une crise d'épilepsie ? Explicitation et recherche médicale Version préliminaire d'un article soumis à la revue Intellectica, *Expliciter 57*, page 2.

[6] VERMERSCH P., (1994, 2006, 2010), *L'entretien d'explicitation*, ESF, Paris. (Pour une description de la pratique)

[7] VERMERSCH P., (2009), Describing the practice of introspection, in Ten years of viewing from within, *Journal of Consciousness Studies*, 16 (10-12), edited by PETITMENGIN C., pp 20-57, Exester EX5 5YX, UK.

[8] VERMERSCH P., (2008), Décrire la pratique de l'introspection. Esquisse d'un article, *Expliciter* 77, pp 33 – 59. Sur le site www.grex2.fr (version française du précédent)

[9] VERMERSCH P., (2012), *Explicitation et phénoménologie*, PUF, Paris. (Pour une réflexion théorique)

Et tout ce qui peut vous intéresser et qui est téléchargeable gratuitement sur le site du GREX www.grex2.fr

Sur les emprunts théoriques

[10] AUSTIN J.L., (1970), *Quand dire c'est faire*, Paris, Le Seuil.

[11] GENDLIN E., (2006), *Focusing, au centre de soi*, éd. de l'Homme, Montréal.

[12] GUSDORF G., (1951), *Mémoire et personne*, PUF, Paris. la mémoire concrète.

[13] HUSSERL E., (1950), *Idées directrices pour une phénoménologie*, Gallimard, Paris.

[14] HUSSERL E., (1998), *De la synthèse passive*, Jérôme Millon, Grenoble.

[15] PIAGET J., (1992), *Réussir et comprendre*, PUF Paris.

[16] RICŒUR P., 2000, *La mémoire, l'histoire et l'oubli*, SEUIL, Paris.

[17] ROGERS C.R., (1968) (2005), *Le développement de la personne*, InterEditions.

[18] SHACTER D., (1997), *Memorys Distorsion: how Mind, Brains, and Societies Reconstruct the past*, Cambridge, Harvard University Press.

[19] SHACTER D., (2003), *Science de la mémoire : oublier et se ressouvenir*, Paris, Odile Jacob.

A Mathematician's View on Mathematical Creation

Pedro J. Freitas
(Departamento de Matemática e Centro de Estruturas Lineares e Combinatórias, Universidade de Lisboa)
pjfreitas@gmail.com

In this paper we present a few personal views on aspects of mathematical creation, illustrated with some examples.

1. Discovery and Invention

There is a long debate on whether mathematics is discovered by humanity, as something existing beyond the concrete human mathematicians, in a Platonic view, or if it is something invented by these mathematicians, even if inspired by the physical world around them or the state of mathematical research at the time. The generally accepted view is the Platonic one, as mathematical results seem to endure throughout time, with the same statements and the same proofs, even though the mathematical work is often described as a creative one. We now take a look at both these elements in mathematics: statements of theorems and proofs.

Once a theorem is proved, it is clear that the result is established once and for all, and cannot be disproved, provided the proof is carefully read and scrutinized. The statement itself, however, can be more contingent than it looks like at first glance. Take, as an example, the famous Pythagorean theorem. One can say that the result is unavoidable, or necessary, but it does depend on the definition of a right triangle, which might leave more room for imagination that it might look at first sight. At the time of Euclid, of course, this definition was clear; however, after the establishment of the geometry on the sphere and of hyperbolic geometry, this concept became broader, as to

include right triangles for which the Pythagorean theorem is no longer true. Einstein himself refers to this fact in very clear words, in a conversation with Rabindranath Tagore [1]:

> I believe, for instance, that the Pythagorean theorem in geometry states something that is approximately true, independent of the existence of man.

A similar situation arises with another very basic mathematical result: the existence and uniqueness of prime factorization for integer numbers. Again, if we consider the usual integers, the result is true. It remains true even for Gaussian integers (the set of complex numbers with integer real and imaginary parts); however, it is not true for some other sets of ”integers” inside the complex numbers: for instance, the complex numbers of the form $a + b\sqrt{5}$, with a and b integers.

Even though these are not as familiar to most people as the usual integers, there are very good reasons for these numbers to be considered “integers”, and to expect them to have a behavior similar to that of ordinary integers. There is even a rather dramatic episode related to this issue, a mistake made by Lamé in an attempt to prove Fermat’s last theorem. Lamé presented a proof of Fermat’s famous conjecture, which depended on uniqueness of factorization in sets of numbers such as the one we mentioned above, which is not true. This fact was brought to light by Liouville just after Lamé’s talk, and the proof stood just a partial one.

One may object that the first definitions of “triangle” and “prime number” are more natural that the others presented here. This may be true at first glance, but, eventually, these non-intuitive objects end up having a more important role than the other ones. This was the case with hyperbolic geometry, which gained a more important role with the development of Riemannian geometry and relativity. Another interesting example of this phenomenon is the definition of a real valued function. At first, only continuous functions were considered as an object of study. Gradually, the concept became broader, as to include functions that were previously considered abhorrent, such as the Dirichlet function, which has value 1 on the rational numbers and value 0 on the irrationals. There is no hope of drawing a graph of this function, given the density of both these sets inside the real numbers. We’ll come back to this concept in a little while.

So, we conclude that even though the results are necessary, the mathematical objects about which we speak can vary greatly, and this will of course influence the theorems one can establish about them. This choice of

objects is thus an element of mathematical creativity, guided in part by the applications of the mathematics in question.

Another place for creativity in mathematics is, of course, proofs. It well known that one can find many proofs for the same result, and they can be quite different from one another. As an example, consider again the Pythagorean theorem. The proof given by Euclid in the *Elements* is based on Figure 1, sometimes called “the bride’s chair”, “the peacock tail” or “the windmill”.

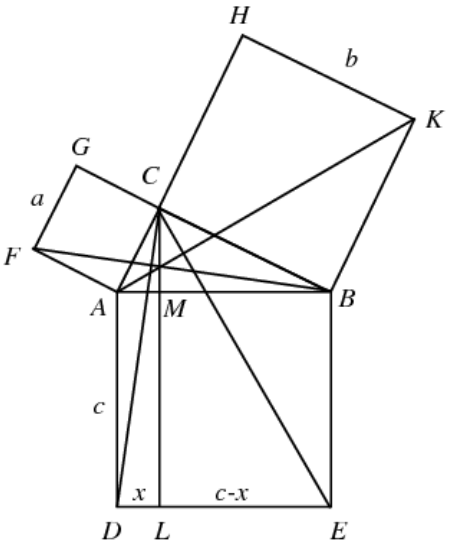


Figure 1. Illustration for Euclid’s proof of the Pythagorean theorem

It relies on comparing the areas of the three squares, by decomposing them into triangles. These triangles start as halves of the small squares, then they slide along some straight lines, preserving area, and finally end up as halves of the rectangles that comprise the large square. It is not a very simple proof — in fact, it is said that Schopenhauer called it “a brilliant piece of perversity”. Two other proofs, given by the Indian mathematician Bhaskara, are illustrated in Figure 2.

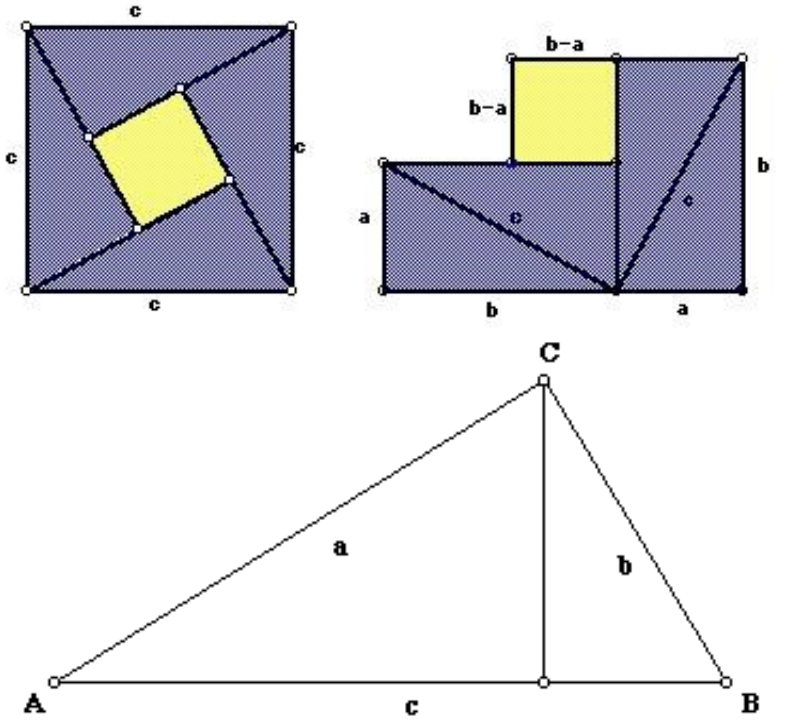


Figure 2. Illustrations for Bhaskara's proofs of the Pythagorean theorem

The first two images illustrate a proof, which is arguably the simplest one, also based on area decompositions. It is quite straightforward to deduce the argument just by looking at them. The second proof, illustrated by the large triangle decomposed into two smaller ones, depends on properties of similar triangles, and it no longer involves considerations about areas. The three triangles shown are similar and by comparing the lengths of corresponding sides, we end up proving the theorem.

These are only three out of hundreds of proofs of this result. This abundance of possibilities points to the element of creativity that exists in finding and producing a proof, probably influenced by both the individual that produces the proof and the culture this person is immersed in. The fact that this result has many possibilities of application also helps explaining this abundance.

To conclude, we could say that, even though there are rules to be followed when proving a given result, these cannot account for all the diversity we find.

2. Rigor and Intuition

Speaking of proofs, the usual understanding about mathematics is that a certain result can only be considered as established once a proof is given. This is an accurate view: no one considered Fermat's last theorem as a theorem before Wiles' work, and no one considers that the Riemann conjecture about the zeta function is true at the time this text is written, even though there is a significant list of mathematical results dependent on this conjecture being true. However, there's more to mathematical certainty than rigorous proof and there's more to proof than mathematical certainty.

To illustrate my first statement, consider Fourier analysis. The statement that a function will coincide with its Fourier series depended, at first, on the functions considered. This was proved to be true for periodic functions with a known formula (d'Alembert and Euler, 18th century), and then Fourier (beginning of the 19th century) ventured to state that the result would hold for a larger class of functions, giving a formula for calculating the coefficients of the series. The theory lacked rigor even for the standards of the time, but nevertheless Fourier's theory won the Grand Prix de l'Académie des Sciences, with a jury that included Legendre, Laplace and Lagrange. The very statement issued by the jury confirms this situation:

> [T]he manner in which the author arrives at these equations is not exempt of difficulties and... his analysis to integrate them still leaves something to be desired on the score of generality and even rigor.

Only with the definition of the Lebesgue integral, at the beginning of the 20th century, did the theory become completely clear. It was finally proved by Carlson, in 1966, that if a function is Lebesgue square-integrable then its Fourier series converges almost everywhere. Nevertheless, the absence of this final result didn't keep engineers, physicists, and even mathematicians from using the Fourier series as a tool.

Another concept that was used way before a reasonable rigorous definition was given was that of an infinitesimal. Newton and Leibniz used this concept when developing the infinitesimal calculus, again facing criticism in their own time. One of the most famous critics of this lack of rigor was George Berkley who wrote the famous sentence:

May we not call them the Ghosts of departed Quantities?

Calculus was of course used since it was established, but it was only in the 19th century, with Cauchy, Bolzano and Weierstrass, that the notion of limit

was rigorously defined. Interestingly enough, this definition did away with the notion of infinitesimal as a quantity, defining it in terms of sequences or neighborhoods. However, in the 20th century, the notion of infinitesimal as a number was given a rigorous definition, first by Robinson and then by Nelson, who actually managed to define them as real numbers.

Even nowadays, mathematicians and physicists will use concepts that are still not completely established, such as the Feynman integral. To this day it was impossible to find a measure affording this integral.

On the other hand, even when there is a rigorous definition of a concept or a proof of a result, mathematicians will still look for alternative ways to establish the result. It is very frequent that the first proof of a hard result is very long and elaborate, and new proofs are welcome. There's more than a need for certainty involved in a proof: mathematicians look also for understanding. As Gian-Carlo Rota puts it in [2]:

> This gradual bringing out of the significance of a new discovery takes the appearance of a succession of proofs, each one simpler than the preceding. New and simpler versions of a theorem will stop appearing when the facts are finally understood.

Bill Thurston also states this very clearly in [6]:

> What we are doing is finding ways for people to understand and think about mathematics. The rapid advance of computers has helped dramatize this point, because computers and people are very different. For instance, when Appel and Haken completed a proof of the 4-color map theorem using a massive automatic computation, it evoked much controversy. I interpret the controversy as having little to do with doubt people had as to the veracity of the theorem or the correctness of the proof. Rather, it reflected a continuing desire for human understanding of a proof, in addition to knowledge that the theorem is true.

Thus, for a mathematician, a proof encompasses not just the logical certainty of a result, but also, and maybe more significantly, the deeper understanding of *why* the result is true, even though the question of what this "why" means cannot be formulated in a clear mathematical way.

3. The individual and the collective

The usual view on the development of mathematical tends to underline the effort of individual people, who made significant progress in the advancement of mathematical knowledge and understanding. Mark Kac, in [4], offers an interesting quote on brilliant scientists:

> In science, as well as in other fields of human endeavor, there are two kinds of geniuses: the "ordinary" and the "magicians." An ordinary genius is a fellow that you and I would be just as good as, if we were only many times better. There is no mystery as to how his mind works. Once we understand what he has done, we feel certain that we, too, could have done it. It is different with the magicians. They are, to use mathematical jargon, in the orthogonal complement of where we are and the working of their minds is for all intents and purposes incomprehensible. Even after we understand what they have done, the process by which they have done it is completely dark. They seldom, if ever, have students because they cannot be emulated and it must be terribly frustrating for a brilliant young mind to cope with the mysterious ways in which the magician's mind works. Richard Feynman is a magician of the highest caliber. Hans Bethe, whom [Freeman] Dyson considers to be his teacher, is an "ordinary genius."

One could easily carry these definitions to the field of mathematics — Terence Tao would be a good candidate for a magician. However, in spite of the colorfulness of the description, it is undeniable that the body of existing mathematical results, and the applications of these results, influence the discovery of new ones, and even the proofs of these new results. It is the case, quite frequently, that more than one mathematician arrives at a given result independently and simultaneously.

In the Introduction of [3], John Gribbin makes this point very clearly — he does so in describing scientific discovery, but we believe it can be also applied to mathematical developments.

> It is natural to describe key events in terms of the work of individuals who made a mark in science […]. But this does not mean that science has progressed as a result of the work of a string of irreplaceable geniuses possessed of a special insight into how the world works. Geniuses maybe (though not always); but irreplaceable certainly not. Scientific progress builds step by step […], when the time is ripe, two or more individuals may make the next step independently of one another. It is the luck of the draw, or historical accident, whose name gets remembered as the discoverer of a new phenomenon.

The case of the establishment of infinitesimal calculus by both Newton and Leibniz is a very known example. The fact that both were very gifted mathematicians is certainly important, but the fact that both created the theory at the same time is a sign that the body of mathematical knowledge was ready to welcome the new theory. Newton's famous quote attests to this:

If I have seen further it is by standing on ye sholders of Giants.

Another famous example of this phenomenon is the discovery of hyperbolic geometry. Farkas Bolyai, in spite of much effort, was unable to find a model proving the existence of such geometry. However, his son Janos

Bolyai (much against his father's advice) managed to succeed, at the same time as Lobachevsky, who worked on the subject independently. It was maybe Gauss's towering influence that made it possible for both mathematicians to succeed (Gauss himself had thought about the subject, even though he hadn't published anything).

So, even though the individual effort of brilliant minds cannot be erased, it is also important to notice that the state of mathematical knowledge at a given moment in a sense engenders the new results and developments.

With the latest possibilities in communication, afforded by the internet, a new type of mathematical collaboration became possible.

One of the most famous instances of this is the Polymath project, started by Tim Gowers. This is a site [8], where problems are stated and contributions are welcome. Gowers himself describes the project as follows:

> It seems to me that, at least in theory, a different model could work: different, that is, from the usual model of people working in isolation or collaborating with one or two others. Suppose one had a forum for the online discussion of a particular problem. The idea would be that anybody who had anything whatsoever to say about the problem could chip in. And the ethos of the forum — in whatever form it took — would be that comments would mostly be kept short. In other words, what you would not tend to do, at least if you wanted to keep within the spirit of things, is spend a month thinking hard about the problem and then come back and write ten pages about it. Rather, you would contribute ideas even if they were undeveloped and/or likely to be wrong.

The project stemmed from Gowers' blog [9], where he suggested that his readers contribute ideas towards finding a new proof of the Hales-Jewett theorem; and explicitly asking the question "is massively collaborative mathematics possible?". The problem became known as Polymath 1. Terence Tao also got involved, with people contributing suggestions his own blog [10], and finally, in 2009, the new proof was found and two papers were published under the pseudonym D. H. J. Polymath, one of them in the very respected Annals of Mathematics.

In the same spirit, there is another site called MathOverflow [7]. In this site, anyone can post a question on a mathematical research topic, and answers are given by other users. The site was started by Berkeley graduate students and postdocs A. Geraschenko, D. Zureick-Brown, and S. Morrison on 28 September 2009 (Terence Tao pointed out that the newsgroup *sci.math* was similar, even though MathOverflow has newer web features). According to Wikipedia, questions are answered an average of 3.9 hours after they are posted, and "Acceptable" answers take an average of 5.01 hours.

Again, the speed and breadth of this interchanging of information only became possible in the late 20th century with the Internet, and may add a distinctive new feature to the way mathematics is created. The coming decades will tell if this way of creating new mathematics will prove relevant or not.

Conclusions

Having in mind that it is quite difficult (and probably even dangerous) to expect final conclusions in subjects such as this one, I could summarize the ideas in this essay as follows:

— Even though mathematical statements seem to have an intrinsic immutable quality, a good deal of creativity is necessary in developing new mathematics;

— Even though mathematical rigor is necessary in stating mathematical results, it is equally important to pay attention to partially established results and to the understanding of mathematics that a proof of a theorem brings

— Even though most mathematical results can be attributed to the work of brilliant individuals, it is also important to pay attention to the collective state of mathematics and to modest contributions when analyzing mathematical progress.

References

[1] David Gosling, *Science and the Indian Tradition: When Einstein Met Tagore (India in the Modern World)*

[2] Gian-Carlo Rota, *The Phenomenology of Mathematical Proof*, Synthese, May 1997, Volume 111, Issue 2, pp. 183-196.

[3] John Gribbin, *Science, a History,* Penguin Books, 2003.

[4] Mark Kac*, Enigmas of Chance: An Autobiography*, Univ of California Pr ,1987.

[5] Philip Davis and Reuben Hersh, *The Mathematical Experience*, Birkhäuser, 1981.

[6] William P. Thurston, *On Proof and Progress*, American Mathematical Monthly, 1994.

[7] http://mathoverflow.net/

[8] http://polymathprojects.org/

[9] http://gowers.wordpress.com/

[10] http://terrytao.wordpress.com/

On The Source of Mathematical Intuition

António Machiavelo
(Centro de Matemática da Universidade do Porto)
ajmachia@fc.up.pt

Introduction

The main question we here address is the following: how can we, human beings, deduce without any apparent recourse to experience, i.e. seemingly *a priori*, truths about the universe in which we live in? And what is the source of the intuitions that guide us in the quest of those truths? Our main focus here are *mathematical* truths. Therefore, we also have to deal with what exactly is a *mathematical truth*.

Let us look at an example that will be used throughout the paper. Consider a *graph*, which is just a

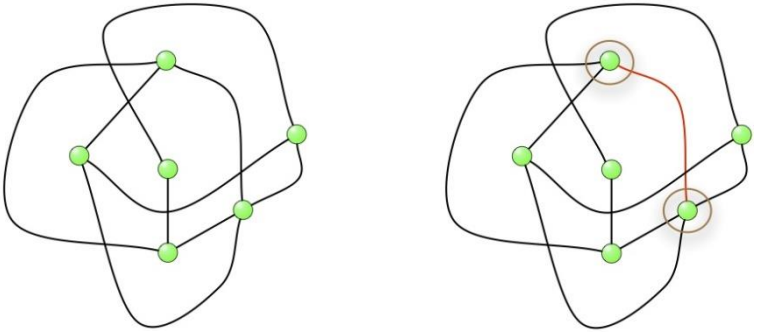


Fig. 1: A graph

set of points, some of which are connected by lines (see the left side of Fig.

1). The shape of the lines is irrelevant, all that matters is whether two points are or not connected by a line. In this context one usually calls *vertices* to the points, while the lines are called *edges*. The number of lines (edges) that come out a given point (vertex) is called the *degree* of that point (vertex). It is very well-known, and easy to see, that the following proposition, which we will henceforth call *P*, holds true:

Proposition *P*: *In any graph, the sum of the degrees of all vertices is equal to the double of the number of edges.*

The reason is simply that when one adds all degrees, one is adding all lines twice, since a line comes out of exactly two vertices (see the right side of Fig. 1).

Now, this implies, for instance, that one can never connect five things so that each one is connected to precisely other three. Be it five branch offices of a business that someone wants to connect with fiber optic cables so that each one is connected to exactly other three, or five capitals that an airline company wants to connect with flights so that from each capital one can fly to exactly other three, or simply five rocks that one wants to connect with ropes so that each stone is connected to exactly other three, it just cannot be done! Why? Because in each one of these tasks one is looking for

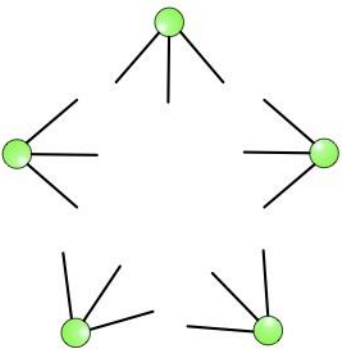


Fig. 2: It cannot be done!

something that is equivalent to the construction of a graph in which the sum of the degrees of all vertices is 15, an odd number, which cannot be, by proposition *P*.

In this way, one has concluded, without any doubt whatsoever, that a very great number of tasks cannot be done, and that was accomplished without no

need to make a single experiment. How is this possible? And where do the ideas and intuitions behind the argument come from?

1. What exactly is Mathematics?

It is not easy to define what Mathematics is about. To simply say that it is the "science of numbers" is so vague and inaccurate as saying that literature is the "art of letters". In the first place, Mathematics deals not only with numbers, but with a rather extensive panoply of objects like geometric figures, sets, functions, algebraic structures, topological spaces, graphs, and so on, some of which have no connection to numbers. Secondly, as with literature, where one merely uses letters to convey thoughts and feelings, in Mathematics, when numbers are used, is mostly to convey thoughts and relationships. Note that, although numbers intervene in the statement of Proposition *P*, this proposition is not about numbers, but about some relationship among the number of points and the number of lines.

Sometimes, it is said that Mathematics is a language, which one then claims to be universal, often comparing it with music. But the really interesting question is what does this language expresses What does Mathematics study? It is more or less clear that, roughly, Physics studies the laws of interaction of matter and energy; that Chemistry studies the interaction of molecules and the properties of the compounds that they form; that Biology studies living organisms, mainly their internal organization, and that Ethology studies the external, individual and social, behavior of complex living organisms. But, even roughly, what part of reality does Mathematics study? Or, is it the case that it does not study anything real? But then, how to explain its truly amazing descriptive and, especially, its predictive power?

A striking example of both these powers is the discovery by James Clerk Maxwell (1831--1879), on paper (!) and with the paramount help of mathematics, around 1864, of electromagnetic waves. He realized that light is such a wave, and that there are many more kinds of these waves, whose existence was only experimentally confirmed more than two decades later, by Heinrich Hertz (1857--1894),[1] who wrote:

> It is impossible to study this wonderful theory without feeling as if the mathematical equations had an independent life and an intelligence of their own, as if they were wiser than ourselves, indeed wiser than their discoverer, as

[1] See [3], Chap. XX, and [4], Chap. 6.

> if they gave forth more than he had put into them... [2]

Other examples would be: the discovery, in 1900, of the quantum nature of the atomic world, by Max Planck (1858--1947), who was literally forced by mathematics to accept physical interpretations he did not like in the least;[3] Riemannian geometry, developed by Bernhard Riemann (1826--1866), around 1854, inspired by the work of Gauss (1777--1855), and later elaborated by Beltrami (1835--1900), Christoffel (1829--1900), Lipschitz (1832--1903), Ricci (1853--1925) and Levi-Civita (1873--1941), which played, more than half a century later, a crucial role in the general theory of relativity of Albert Einstein (1879--1955);[4] the prediction in 1928 of anti-matter made by Paul Dirac (1902--1984),[5] which was experimentally confirmed four years later.

These are just some examples of what has been called by the physicist Eugene Wigner[6] the "unreasonable effectiveness of Mathematics in the Natural Sciences".[7] This unreasonable effectiveness does show that whatever the language of Mathematics expresses, it must have some real content. This has been eloquently articulated by Galileo, in a famous passage of his 1623 book *Il Saggiatore* (Chap. 6):

> Philosophy is written in this grand book - I mean the universe - which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language and interpret the characters in which it is written. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures, without which it is humanly impossible to understand a single word of it.

It is also often said of Mathematics that its deals only with approximations to reality. That since perfect triangles or perfect circles do not exist, results about triangles or circles can only be used within prescribed degrees of error. I will return to this later, but now I just wish to note that Proposition *P* presented above is not an approximation to reality, but an exact description of a feature of reality. It is indeed utterly impossible to connect five rocks, or any other five things, so that anyone of them is connected to exactly three other, or in any other way which violates what that proposition states. The relation

---

[2] Quoted in [5], p. 101.
[3] 6], p. 4.
[4] See [7], Chap. 37, and §4 of Chap. 48.
[5] See [8], p. 392.
[6] Nobel laureate in Physics, 1963.
[7] In [9]. See also [11,12].

stated by the proposition is absolutely necessary, as all of its instances. But the interesting point is that, while one can easily imagine, and even conceptually play with worlds that have different physical laws, like one in which the gravitation law would be inversely proportional to the cube of the distance, instead of the square, one cannot envision an universe where Proposition *P* is false. As Raymond Smullyan writes in [13, p. 47]:

> The physical sciences are interested in the state of affairs that holds for the actual world, whereas pure mathematics and logic study all possible state of affairs.

This, to me, hints at the fact that, in Mathematics, one is studying some sort of deep structural laws on which the universe is built upon, something that underlies the physical laws of Nature, maybe even some structural laws that must be satisfied by any possible Universe. Pushing the point a bit further, it does seem that when the Universe come into existence at the so called "Big Bang" (a not very good name, by the way), it already come with some structural fabric and that, somehow, Mathematics is the area of human knowledge that studies precisely that fabric, a sort of logic inner fabric underlying everything.

2. Mathematical Objects

The ontological status of mathematical objects has been a source of philosophical debate since, at least the time of Plato. Whether mathematical objects are real or ideal, are discovered or created, has been discussed for millennia. And the controversy is pretty much alive in the 21th century, as shown by a series of short papers published by the *Newsletter of the European Mathematical Society*, [14, 15, 16, 17, 18, 19, 20]. Lots of philosophical "theories" have been conceived to try to answer those puzzling questions. Although all of them do make pertinent and interesting remarks on these matters, none of them seems to quite yield completely satisfactory answers. Either they do not adequately explain how can mathematics have so many and quite impressive applications to the "real" world, or they do not clearly unravel in exactly what form are mathematical objects real.

But before we tackle this question of the status of mathematical objects, one should first try to make clear what exactly does one mean by "object". We have to be very careful here, since we humans have a more than natural tendency to attribute reality, or existence, overwhelmingly to material objects

that our senses can directly detect. Now, even if one tries to restrict the notion of "real objects" to things that are "physical" in some sense, one immediately runs into some difficulties, as for example: are electromagnetic waves "real" objects? What about gravity? These do seem to have a form of existence that is quite different from, say, a rock.

But even a "physical" object like a person, for example[8],has layers of complexity that, although well known, are seldom thought of. To see this, let us consider some of the levels at which one can describe a human being (see Fig. 3) To a doctor, he is but a set of organs and its interrelations - let me draw here the reader's attention to the importance of these interrelations: rearranging the organs has absolutely dramatic consequences! Now, to a biologist, he is a set of cells and their (vital!) interrelations. To a physicist, he is but a set of atoms and their (crucial!) interrelations. But, and I find this extremely curious, according to the 1932 Nobel laureate in physics, Werner Heisenberg (1901--1976), elementary particles are "mathematical forms" ([6], p. 36) and, in general, (p. 51):

> The 'thing-in-itself' is for the atomic physicist, if he uses this concept at all, finally a mathematical structure.

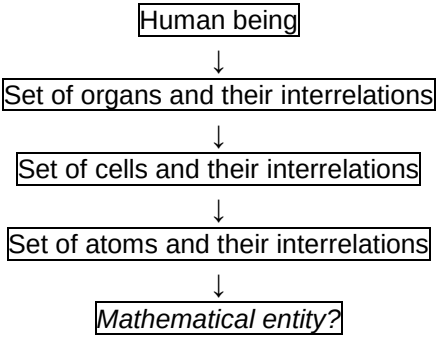


Fig. 3: What is an object, really?

Therefore, the question of knowing what a "real" object is, in order to eventually help clarify what a mathematical object might be, leads us right into mathematical structures!

To complicate things even further, let us observe that a human being is a

[8] I do not believe in things for which there is no evidence for their actual existence, so I am here disregarding beliefs in the existence of supernatural (whatever that means!) components of humans or other animals.

set of atoms, together with their very special interrelationships, that varies with time! In an address to the National Academy of Sciences of the USA, in 1955, titled *The Value of Science* (included in [21], pp. 240--248), Richard Feynman[9] (1918--1988) noted:

> [the] phosphorus that is in the brain of a rat --- and also in mine, and yours --- is not the same phosphorus as it was two weeks ago. [...] the atoms that are in the brain are being replaced: the ones that were there before have gone away.
> So what is this mind of ours: what are these atoms with consciousness? Last week's potatoes! They now can remember what was going on in my mind a year ago --- a mind which has long ago been replaced. [...] the thing which I call my individuality is only a pattern or dance [...]. The atoms come into my brain, dance a dance, then go out --- there are always new atoms, but always doing the same dance, remembering what the dance was yesterday.

That is, humans and animals in general are much more like rivers than like rocks: they are patterns, rather than "fixed" physical objects.

From all of this, what I want here to emphasize is that relations between "physical" objects are as real and important as the objects themselves. And that there are laws and patterns ruling these interrelations which are as real as anything else. Mathematics seems to capture some of these inner relations that are just not visible to the naked eye. As Rudy Rucker, in p. 4 of [22], writes:

> Mathematics is the study of pure pattern, and everything in the cosmos is a kind of pattern.

Numbers themselves are but representations of some special kinds of relationships. To make this clear, let us first point out the distinction between a number, e.g. 6, and its representation[10]. In fact, "6" is not the number six (see Fig. 4).

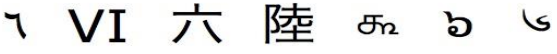


Fig. 6: Several representations of the number 6

[9] Nobel laureate in physics, 1965.

[10] The distinction between a representation and the thing being represented is eloquently illustrated by some paintings of René Magritte (1898--1967), namely the one titled "La trahison des images" (see http://en.wikipedia.org/wiki/The_Treachery_of_Images), which consists of a drawing of a pipe together with the sentence "this is not a pipe". This is entirely true: there is no pipe in the painting, only a representation of it!

So, what is the object represented by "6"? What does it refer to? More precisely, what exactly is the number six? Well, it is a certain "quantity", which is a certain property of a collection of objects. It is actually the common property of all collections that have that particular number of elements, and it captures a certain relation that those collections all have among themselves.

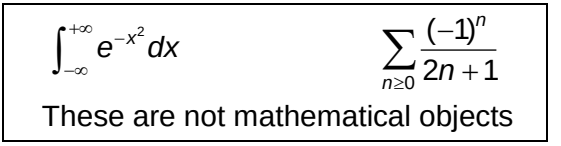


Fig. 5: Paraphrasing Magritte

In conclusion, mathematical objects encode some subtle relationships, and are not to be expected to literally exist out there in the same manner as a physical object exists, although one should be careful, since a detailed analysis shows that even "physical" objects may be quite more intricate than realized at first glance. So, numbers, triangles, circles, and other mathematical entities are but constructs that represent deep, hidden relations, and are not to be taken verbatim. But these relations are as real as any other "objects".

3. An Evolutionary Perspective

More than two millenia ago, Plato could not explain how humans seem to be born with some form of knowledge --- some sort of software, as we could now call it ---, and could not explain how can one reach previously unknown truths from deductions alone, except by arguing for the existence of another sort of parallel world, a world of "forms", and for the pre-existence of a "soul" that would have inhabited that world before being "attached" to a body. Now, most people do not seem to realize that those mysteries were solved, in a much more satisfactory way, about 150 years ago.

Before explaining how, let me rephrase what I tried to convey in the last section, that the "forms" of Plato do exist in this world, not in a mysterious and intangible ideal world. They are the laws governing the interconnections of matter and energy, and of more subtle properties, like "quantity" and various kinds of relations among things, and also the laws governing the

interconnections between those "first-order" laws, maybe even some "higher-order" laws. They are all part of a sort of inner structure of the Cosmos. Mathematical objects (not their representations!) are the elements of that structure. But then how do we have access to that "mathematical structure"?

The answer was given by one of the most brilliant and diligent humans of all times: Charles Darwin (1809--1882), who perfectly summarized it in the so called "notebook M",[11] in which one can find (p. 128, in an entry dated 4 September 1838):

> Plato says in Phaedo that our "necessary ideas" arise from the preexistence of the soul, are not derivable from experience --- read monkeys for preexistence.

This is just a note that Darwin wrote to himself, but after one understands the history of life on this planet, which was made possible by the seminal discovery of "natural selection and descent with slow modification", its meaning becomes clear. We humans are the result of thousands of millions of years of selection, of real experiences made by countless generations of all our ancestors, from all the species of which we are descendants. There is therefore a vast array of experience contained in our genetic code, experiences that we draw upon to explore the Universe that surrounds us. So, when a human being is born is not some sort of blank slate, but comes equipped with powerful tools to understand Nature.

Now, the discovery of the mechanism of "natural selection and descent with slow modification"[12] or the "theory of evolution", as it is commonly known,[13] explained so many things that were previously completely baffling, and made intelligible an huge array of data and observations about living organisms previously scattered and mystifying. It stimulated, and continues to stimulate, fruitful research in several areas of biology.[14] However, after more than 150 years it is still not properly understood by many people, and there are too many misconceptions[15] and completely wrong ideas about it.

Among the main erroneous ideas that interfere with an understanding of the theory of evolution, let us mention the following: (a) life evolves purely randomly, which arises from not realizing that there is a sharp distinction between the randomness of mutations and the mechanism of natural

[11] Available online, at http://darwin-online.org.uk.
[12] [23]. See also [24, Chaps. 3 and 4.
[13] The term "evolution" is not the best, since it gives the wrong idea of a "progress", but unfortunately "natural selection and descent with slow modification" was just too big.
[14] See [25], Chaps. 7-10.
[15] See [26] and http://evolution.berkeley.edu/evolibrary/misconceptions_faq.php.

selection, which is anything but random; (b) to be "fit", meaning well adapted to a particular environment, implies to be ruthless and strong; (c) evolution implies a continuous progress from "inferior" animals to "superior" ones; (d) it justifies mean, cruel and immoral behaviour; (e) it justifies the "law of the jungle". Partly, these confusions came from the fact that there has been an exaggerated emphasis on competition over cooperation in descriptions and popular introductions of the theory of evolution. Here we limit ourselves to note that a human being is in fact the result of tremendously complicated symbiotic relationships. In our digestive tract alone there are hundreds of species of bacteria, essential to our survival, and their total number is ten times greater than the total number of human cells in the body[16]!

The wrong, but very pervasive, ideas about the theory of evolution, together with the fact that this theory removes humans from a central pedestal above all other living creatures (which hurts our natural anthropocentric feelings), lead to an emotional denial, be it conscious or unconscious, of the "transcendently democratic"[17]and profound consequences of the insights of Darwin. A perfect example of this, and quite relevant for the subject of this essay, is the following passage from [27][18] (p.19), where Roger Penrose clearly states why he prefers Plato's intangible, ideal world:

> How do I really feel about the possibility that all my actions, and those of my friends, are ultimately governed by mathematical principles of this kind? I can live with that. I would, indeed, prefer to have these actions controlled by something residing in some such aspect of Plato's fabulous mathematical world than to have them be subject to the kind of simplistic base motives, such as pleasure-seeking, personal greed, or aggressive violence, that many would argue to be the implications of a strictly scientific standpoint.

This shows that the author felt into the trap of some of the above mentioned misconceptions. It comes then as no surprise that he writes a little later:19

> it remains a deep puzzle why mathematical laws should apply to the world with such phenomenal precision.

In a Darwinian perspective this mystery starts to fade away, since, as Carl Sagan explains in [28], pp. 232--233:

[16] See http://en.wikipedia.org/wiki/Gut_flora.
[17] See [24], p. 67.
[18] Which is, nevertheless, an amazing book, a true tour de force!
[19] [27], pp. 20--21.

> we can imagine a universe in which the laws of nature are immensely more complex. But we do not live in such a universe. Why not? I think it may be because all those organisms who perceived their universe as very complex are dead. Those of our arboreal ancestors who had difficulty computing their trajectories as they brachiated from tree to tree did not leave many offspring[20].Natural selection has served as a kind of intellectual sieve, producing brains and intelligences increasingly competent to deal with the laws of nature. This resonance, extracted by natural selection, between our brains and the universe may help explain a quandary set by Einstein: The most incomprehensible property of the universe, he said, is that it is so comprehensible.

I have always found it rather curious that everyone is so amazed with the extraordinary fine-tuning between some characteristics of some animals and their environment, and do not notice that the same applies to the human animal. They seem to assume, explicitly or, most of the time, implicitly, that there is a fundamental separation between our mental capabilities and Nature. The mind is the product of a natural selection that operated over a vast period of time, and is just as part of Nature as anything else. It contains remarkable adaptations of humans to their environment, including the capabilities of pattern detection, abstraction, and the organization of information. As Rudy Rucker so well puts it, in [22], p. 16:

> That our mathematics is effective for manipulating concepts is perhaps no more surprising than that our legs are good for walking.

4. The Source of Intuition

As the success of Physics in describing a huge amount of diverse phenomena shows, the Universe clearly seems to have a sort of inner mathematical "texture". Now, we humans are the product of a natural selection process that produced our brains, which can, through pattern recognition and abstraction, access at least part of that texture. This has allowed our species to uncover some parts of our Universe that totally escape detection by our senses, like radio waves, for instance. As sketched in Fig. 6, through an amazingly rich evolutive heritage, our brains are able to capture the mathematical structure of the Cosmos, and this has allowed us to enlarge our horizons, by uncovering parts of the Universe that were previously unknown to us.

[20] Obviously, this is just a caricatural example.

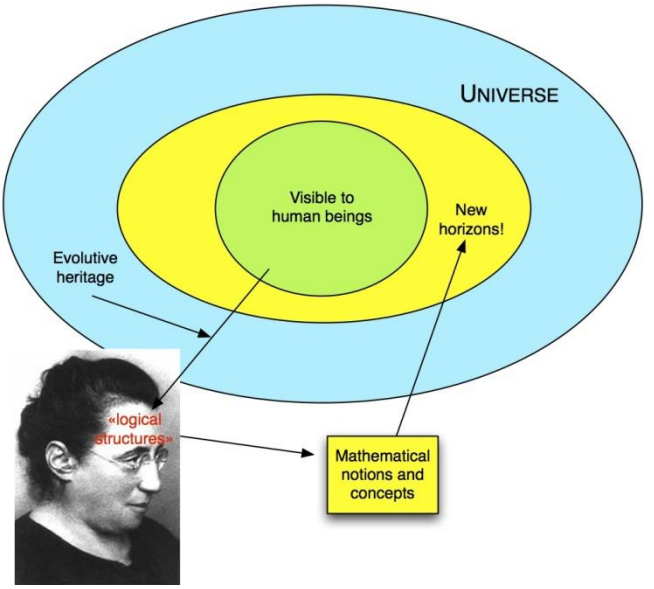


Fig. 6: The universe, human beings and mathematics

One can now see that working on an open problem in Mathematics, for example, helps to sharpen some of our main evolutionary tools, testing our intellectual limits, and this effort may lead to build the tools to overcome some of those limits. The evolutionary advantages of this should be obvious. Of course, any progress in any given problem will be but a tiny and very humble piece of knowledge about the intimate structure of our universe, but it is still worthwhile. Leonhard Euler (1707-1783) said it best [29]:

> knowledge of every truth is a worthy matter in itself, even of those which seem unrelated to popular use; we have seen that all truths, at least those which we are able to understand, are so greatly connected with one another, that we cannot consider any one of them altogether useless without some rashness.
>
> And so, even if a certain proposition seems to be this way, so that regardless of whether it turns out to be true or false, it would be of no benefit to us anyway, still the method itself, by which we would established its truth or falsity, nevertheless may be useful in opening up the way for us to discover other, more useful truths.

What I have tried to argue above is that, in order to understand what Mathematics is about, one must first realize that, besides physical objects (whatever they really are), and as importantly, the world contains some sort of

intrinsic logical inner structure. And our brains have been selected to apprehend it, to some extent. Working on a mathematical problem, as abstract as it may be, is to uncover a tiny piece of that inner structure. The source of the intuitions that guide us in these investigations resides on our immensely rich evolutionary heritage.

Of course, how exactly does that genetic heritage comes alive in each one of us is still largely unknown, and to unravel its secrets will represent a tremendous challenge for generations to come. In the same vain, the problem of knowing precisely what is the mechanism behind mathematical intuitions, whatever exactly that means, the problem of knowing precisely what intuitions are behind a result like Proposition *P* discussed above, remains to be understood. But, and this has been the central point of this paper, one simply cannot do that without a proper evolutionary perspective.

References

[1] A. Machiavelo, *A Natureza dos Objectos Matemáticos*, Gazeta de Matemática 161 (2010) 7-16.

[2] A. Machiavelo, *On the Importance of Useless Mathematics*, in Ehrhard Behrends, Nuno Crato, José Francisco Rodrigues (eds.), *Raising Public Awareness of Mathematics*, Springer-Verlag, 2012, pp. 397-408.

[3] Morris Kline, *Mathematics in Western Culture*, Oxford University Press, 1964 (original from 1953).

[4] Robert Osserman, *Poetry of the Universe: A Mathematical Exploration of the Cosmos*, Anchor, 1996.

[5] Giona Hon, Bernard Goldstein, *Hertz's Methodology and its Influence on Einstein*, pp. 95--105, in: Gudrun Wolfschmidt (ed.), *Heinrich Hertz (1857-1894) and the Development of Communication: Proceedings of the Symposium for History of Science, Hamburg, October 8-12, 2007*, Books on Demand, 2008.

[6] Werner Heisenberg, *Physics and Philosophy: the Revolution in Modern Science*, Penguin Books, 1989 (original from 1958).

[7] Morris Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, 1990.

[8] Werner Heisenberg, *Development of Concepts in the History of Quantum Theory*, American Journal of Physics 43 (1975) 389-394.

[9] Eugene P. Wigner, *The Unreasonable Effectiveness of Mathematics in the Natural Sciences*, Communications in Pure and Applied Mathematics 13 (1960) 1--14.

Reimpresso em [10], vol. III, pp. 116-125.

[10] Douglas M. Campbell, John C. Higgins (eds.), *Mathematics: People, Problems, Results*, Wadsworth International, 1984.

[11] R. W. Hamming, *The Unreasonable Effectiveness of Mathematics*, The American Mathematical Monthly 87 (1980) 81-90.

[12] Mark Colyvan, *The Miracle of Applied Mathematics*, Synthese 127 (2001) 265-277.

[13] Raymond Smullyan, *Forever Undecided: a puzzle guide to Gödel*, Oxford University Press, 1987.

[14] E Brian Davies, *Let Platonism Die*, Newsletter of the European Mathematical Society, June 2007, pp. 24-25.

[15] Ruben Hersh, *On Platonism*, Newsletter of the European Mathematical Society, June 2008, pp. 17-18.

[16] Barry Mazur, *Mathematical Platonism and its Opposites*, Newsletter of the European Mathematical Society, June 2008, pp. 19-21.

[17] David Mumford, *Why I am a Platonist*, Newsletter of the European Mathematical Society, December 2008, pp. 27-30.

[18] Philip J Davis, *Why I Am A (Moderate) Social Constructivist*, Newsletter of the European Mathematical Society, December 2008, pp. 30-31.

[19] Martin Gardner, *Is Reuben Hersh 'Out there'?*, Newsletter of the European Mathematical Society, June 2009, pp. 23-24.

[20] E Brian Davies, *Some Recent Articles about Platonism*, Newsletter of the European Mathematical Society, June 2009, pp. 24-27.

[21] Richard Feynman, *What Do You Care What Other People Think?*, Bantam Books, 1989.

[22] Rudy Rucker, *Mind Tools: the Five Levels of Mathematical Reality*, Houghton Mifflin, 1987.

[23] Charles Darwin, *On the Origin of Species by means of Natural Selection, or the preservation of favoured races in the struggle for life*, John Murray, 1859. Available in *The Complete Work of Charles Darwin Online*, at http://darwin-online.org.uk.

[24] Carl Sagan, Ann Druyan, *Shadows of Forgotten Ancestors*, Ballantine Books, 1993.

[25] John A. Moore, *Science as a Way of Knowing: the Foundations of Modern Biology*, Harvard University Press, 1993.

[26] T. Ryan Gregory, *Understanding Natural Selection: Essential Concepts and Common Misconceptions*, Evolution: Education and Outreach 2 (2009) 156-175.

[27] Roger Penrose, *The Road to Reality: a Complete Guide to the Laws of the Universe*, Vintage Books, 2005.

[28] Carl Sagan, *The Dragons of Eden: Speculations on the Evolution of Human*

Intelligence, Hodder & Stoughton, 1977.

[29] L. Euler, Theoremata circa divisores numerorum (E134). *Novi Commentarii academiae scientiarum Petropolitanae* 1, 1750, pp. 20-48. Reprinted in *Opera Omnia*: Series 1, Volume 2, pp. 62-85. Original article available online, along with an English translation by David Zhao, at www.eulerarchive.org.

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

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA