Many ontological, epistemological and ethical issues arise when one deals with the topics of minds and persons and the connections between them. In particular, different approaches to the nature of the thinking subject – as cognitive agent, perceiver, consciousness, reason, inquirer, subjectivity and power to act – may converge and, eventually, conflict. In order to address the issues posed by (some) minds being persons, the present volume brings together authors working within different philosophical traditions, whose work was initially presented in Porto, in 2011, at MLAG's 1st Graduate Conference.

Edited By

Rui Vieira da Cunha University of Porto
Institute of Philosophy
Clara Morando University of Porto
Institute of Philosophy
Sofia Miguens University of Porto
Institute of Philosophy
Department of Philosophy

From Minds to Persons
Rui Vieira da Cunha | Clara Morando | Sofia Miguens

From Minds to Persons

Rui Vieira da Cunha
Clara Morando
Sofia Miguens

COMPETE PROGRAMA OPERACIONAL FACTORES DE COMPETITIVIDADE
QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL PORTUGAL 2007.2013
UNIÃO EUROPEIA Fundo Europeu de Desenvolvimento Regional
U.PORTO FACULDADE DE LETRAS UNIVERSIDADE DO PORTO
FCT Fundação para a Ciência e a Tecnologia MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA
mlag MIND LANGUAGE ACTION GROUP
mlag MIND LANGUAGE ACTION GROUP DISCUSSION PAPERS

Edited by
RUI VIEIRA DA CUNHA, CLARA MORANDO, SOFIA MIGUENS

FROM MINDS TO PERSONS

PROCEEDINGS FROM THE FIRST MLAG GRADUATE CONFERENCE

Edição

U.PORTO
FACULDADE DE LETRAS
UNIVERSIDADE DO PORTO

Apoio

Instituto de Filosofia

Ficha Técnica

Título
From Minds to Persons - Proceedings from the First MLAG Graduate Conference

Coleccção
MLAG Discussion Papers, vol. 6

Editores
Rui Vieira da Cunha, Clara Morando, Sofia Miguens

Impressão
Invulgar Graphic - Penafiel

Depósito Legal
381060/14

ISBN
978-989-8648-05-1

ISSN
1646-6527

Este trabalho é financiado por Fundos FEDER através do Programa Operacional Factores de Competitividade – COMPETE e por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito do projecto «PEst-C/FIL/UI0502/2011» (FCOMP-01-0124-FEDER-022671).

TABLE OF CONTENTS

INTRODUCTION

From MLAG's first ever externally funded projects (*Rationality, Belief, Desire I* and *II)* to its latest one *(The Bounds of Judgment – Frege, Cognitive Agents and Human Thinkers)*, it has always been our goal to bring together graduate students doing their dissertations (Masters or PhDs) at The University of Porto to work on issues connected with the groups' Projects. This intention has been complemented with the natural ambition of presenting the work being done by those graduate students in many of our conferences and colloquia. Yet never to this date had we arranged for a specific forum that could fulfill that purpose.

The 1st MLAG Graduate Conference provided such forum, and its main purpose was the presentation and discussion of ongoing doctoral research in the four structuring domains of MLAG. The conference, which also included presentations by invited speakers from Universities which collaborate or have collaborated on MLAG's projects (Universities of Lisbon, Paris, Jyväskylä, and Santiago de Compostela), took place in the University of Porto, on November 10th and 11th 2011. This book is one material result of that productive encounter and every essay in it was presented and discussed at the Graduate Conference.

The book starts with **Manuela Teles'** article ***Brewer against Tradition***, on the question of whether perceptual experiences have representational content. Teles begings by analysing Brewer's early position on the nature of perceptual experiences, focusing on the article "Do Sense Experiences Have Conceptual Content?". She then proceeds to "Perception and Reason", where Brewer claims that versions of foundationalism, coherentism and reliabilism within analytical epistemology are not by themselves sufficient to properly account for the grounding of empirical knowledge, since subjectivity (the subject's grasp) is not taken in consideration and the relevant contents of experience are absent. Brewer has been insistent on

the fcat that conceptual demonstrative contents of perceptual experience cannot be ignored; the nature of demonstrative thought was in fact crucial for his early take on perception. It once led to the endorsement of a representationalist view. His later refusal of such a position, from "Perception and Content" to the present, lies on the assumption that representationalism about perception constrains us to strictly distinguish between what, in perceptual experience, as considered from its "interior" or "moment of occurrence", is supposed to be seen as a propositional attitude and what belongs to the "subjective nature" of the same experience – and that is an almost impossible task. If we are to make a methodological distinction between the perceptual experience and what constitutes the range (or the single) empirical beliefs within the former, the very notion of content disappears. That is why Teles calls Brewer's "Content View" "the Content Dilemma" approach. The article ends with Brewer's alternative explanation of illusions and delusions.

Roberta Locatelli's paper ***Disjunctivism and the puzzle of phenomenal character*** goes around the problem of phenomenal disjunctivism, listing, firstly, its main characteristics as pointed out by the philosopher Michael Martin. The question arises whether if it is really possible to make a full distinction between a veridical experience and a hallucinatory one, and whether if the characterization of perceptive experience as having a phenomenal character (or an introspective one) really makes sense or if it just leads to a misguiding error whose unique purpose is simply to entail a distinction for methodological reasons, that for the author appear as belonging to a *naïve realistic* view, without truly any *raison d'être*. Issues as the indiscriminability between different kinds of perceptions, also that which describes the hallucination phenomenon as being under a negative conception of what should be the main traces just by focusing on their absence, and the problem of the phenomenal naïve realism, all these raise the question about epistemic sovereignty in a way that does not cope with the prior importance of perception considered only in itself. Perception is a process too complex just to be described as a phenomenal experience, without considering at the same time its sheer richness in operative sensory factual and cognitive assumptions. A good disjunctivism, in the author's view, means to envisage perception as independent of its 'subjective-phenomenal' experience, in order to assure a solid explanation to subsisting pathologies in cases of hallucination. So, this means that the same hallucinatory phenomenon should be absolutely distinguished from the general experience of perception, since what lies underneath as its explicating source is not the phenomenal experience of

hallucination in itself, but a problem in the perceptual basic system.

Paulo Jesus' paper ***Selfhood as grammatical responsibility: Between the will-to-understand and the will-to-explain*** raises the problem of an *ontological difference* between the realm of reasons and the realm of causes in several authors (Kant, Wittgenstein, Anscombe, etc.), who have discussed this very issue among their various works. Reasons and causes do not come into a communal sharing of *a world* since they are among them incommensurable, despite the confusion which often arises when considering both. The way reasons resemble causes is so different from the one we find in the stance of causes, considered in themselves, that we are necessarily led to assuming a monist ontology for the first ones (for reasons that in the end subsume to one single Reason, to the One), and for the second ones (for causes which permit what we could call the chaos of diversity, a multiplicity that seems not to impose a leading point of departure all into the rest). What happens is that for the realm of causes, the chain of successive happenings they draw up in nature, is at times confused with the very different nature of *a* reason. This does not mean that, in the last instance, we do not subscribe a chain of causes under *a* reason. So the monist ontology we find in the order of reasons also applies, in a subtle fashion, to the order of causes, even though that very same monist ontology has to cope with a dynamics whose main function is to done *matter* to *a* reason. "A deep ontological homogeneity" needs, in order to be what is, to rest under a "surface heterogeneity", and in the same stroke it's explicating and organizing virtues. The Kantian "I think" is here converted into a "self-alert phenomenology" where the action of the "I feel" (the translation now of the "I think") serves precisely as *the* reason for a manifold of "actings" and several "making to happen", giving rise, then, to the emergence of the self: a competent player within various semiotic-phenomenological games where the meaning is a construction that follows the silent-(un)silent dynamics of the body(ies), the language(s) and the world(s).

Clara Morando's essay puts up the problem of reflexivity and pre-reflexivity as it appears on Sartre's account of consciousness. Turning around on how this issue is developed in the first works of Sartrean psycho-phenomenology – mainly in *The Transcendence of the Ego* –, we are led into the idea that rests within (or throughout) the Cartesian *cogito* a non-capturable trace of a un-reflexivity, showing us the urgent need for a clarification into the true backgrounds of philosophical activity as a reflexive one, as not being its true core. Consciousness is likewise a Janus-faced operation where it lies as prior to any objectifying (reflexive)

commitment a pre-reflexive soil of conscious activity, calling our attention to something as another kind of "reasoning". The reference to authors like Kant, Husserl and Merleau-Ponty, concerning questions as what the nature of consciousness means for each of them, respectively the transcendental unity of apperception, the self-unifying stream of consciousness/the transcendental Ego, and the pre-predicative subjectivity, helps us to make evidence on the way of functioning of a consciousness whose first trait is precisely its intentionality.

What is music, metaphysically speaking? This is the question tackled by **Vitor Guerreiro**'s ***Thinking Clearly about Music***, an essay that leads us through the theories on the nature of music and at the same time tries to get us closer (or as close as possible) to a more accurate concept of music. Beforehand, one fact about this essay should be noted: although the author begins by asking what is music and finishes with a definition of music, most of the thinking done is not directly concerned with music or with the definition of music but with theories of music. The worries that guide this paper are focused on what kind of philosophical theory of music we should strive for and the concern with a definition of music is, at most, secondary, even if some definitions are presented and one in particular defended, in the closing remarks. Guerreiro's way of forcing us to reflect on the kind of philosophical theory of music we should strive for begins with the distinction between natural-kind theories of art (NKTA) and cultural-kind theories of art (CKTA). The distinction, presented by Dickie in his 1997 article "Art: Function of Procedure, Nature or Culture?", is based on a complex separation between natural-kind activities (NKA) and cultural-kind activities (CKA), the difference here being that the former are done spontaneously by living organisms (or genetically fixed – although further in the essay a way of distinguishing NKAs and CKAs is presented that involves the appeal to biological rigidity).

The distinctions, however, do not end here and Guerreiro uses the notion of conceptual dependence to introduce a separation within the class of CKAs, between conceptually dependent CKAs and conceptually independent CKAs. "*Conceptually dependant CKAs are those that essentially involve the act of counting some X as some Y in a context*", claims Guerreiro, who uses Nelson Goodman's example of the configuration of stars as constellations. Now, Guerreiro does not embrace Goodman's case for constructivism: he merely seems to be claiming that some CKAs are conceptually dependent, while others are conceptually independent, and that the use of language is at the heart of such distinction. Such conceptual

relativity, Guerreiro claims, is still consistent with realism, arguing that conceptual independent facts are prior to conceptually dependent facts. Guerreiro then uses Searle's *causal agentive functions* and *status functions* to distinguish between traditional functionalist theories of art and institutional/proceduralist theories of art, in particular, and between NKTA and CKTA, in general. The main difference is that NKTA appeal to causal agentive functions (like the functions of artifacts in general, these depend on physical structure and intention but are not language dependent) while CKTA appeal to a status-function (which are language dependent, collective intentionality dependent and the backbone of institutional reality).

Through consideration of status-functions, and, in particular, of one of their features, their self-referentiality, Guerreiro raises a novel argument against the institutional theory, a new objection not to be confused with the traditional objections of circularity. Guerreiro's objections is that an institutional concept of art presupposes a more basic functional concept, because every institutional kind is grounded on a network of practices, relations and causal roles, much in the same way conceptual dependent facts are grounded on conceptually independent facts. The upshot is that a NKTA, provided it is supplemented with an explanation of how the status-function actually works, can give a satisfactory explanation of art (and music). This also means that such an enhanced NKTA will deal with concepts of art (and music) which have two-layers: one element of causal agentive functions and one element of status-function. Naturally, this is the point where the author returns to the concept of music, where the connection between thinking clearly about music and the definition of music is rendered explicit. Working on a definition of music presented by Andrew Kania as (1) any event intentionally produced or organized (2) to be heard, and (3) either (a) to have some basic musical feature, such as pitch or rhythm, or (b) to be listened to for such features, Guerreiro tries to show that it accords with his previous reflections on NKTA. Specifically, the author strives to show how the disjuncts of 3 in the definition (a and b) are well adjusted to reflect the two-layered structure of causal agentive functions and status-functions that NKTA deal with. This adjustment is important for the author also because it is a reason to accept the disjunctive definition as really reflecting something about the objects that fall under the extension of the concept being defined and not just being *ad hoc*. It is precisely because there are both cases of objects that fall under the concept of music due to their physical structure (causal agentive functions) and cases of objects that fall under music due to collective representation

(status-functions) that one needs a disjunctive definition.

Leaning on the Kantian motto "you will not learn from me philosophy, but how to philosophize, not thoughts to repeat, but how to think", **Tomás Magalhães Carneiro's** essay shows how the exercise of "daring to think" must be one whose premises lie not on the process of sterile repetition of ideas, which leads to an "intellectual anesthesia", but on the assumption that learning "how to think" (and not "what to think") is the key-passage to a spiritual emancipation. Through the Philosophical Dialogue, behavioral traits as laziness and cowardice tend to be progressively substituted by a "Socratic skepticism", improving student's critical thinking into a more open-minded way of interpreting the world as a non-dogmatic compound of knowledge and events. Anxiety plays a key-role in this entire educative scenario as the proof it is coming from the assuming of the essential heterogeneity about the right manner to think what is there to think. Only from this point of view can we understand the true role the Human Being has as a creative viewer of *a world*. Thus, the "ideal student" does not exist, since it is not supposed to just exist only one correct view but several possible views about the same "problem", the same "question", the same "world". The unique concern taken as indispensable is the one fulfilling the main lines of a coherent argumentation process, which should lead to (a) plausible truth(s). Listing the rules, obstacles and tools to achieve a successful Philosophical Dialogue constitutes one of the main steps the teacher has to take in order to make possible "the game of philosophy", working as a referee of this very process. Stating also the values as the ones indispensable to build sustained and harmonious co-operative dialogues, among students, is, thus, the goal-work to be done in the classroom "environment".

João Machado Vaz' paper ***Power and Beauty In Psychopathology. Eugen Bleuler's concept of schizophrenia*** deals with the historical evolution of the concept of Schizophrenia along the twentieth century, recalling however important philosophical concepts, mainly some of those from the previous century, the epoch of German Romanticism, as *Will* (Schopenhauer) and *Power* (Nietzsche). We find here the author's intent to pursue a deeper clarification of what we may understand as the leading *spiritual forces* under several psychopathologies, namely of schizophrenia. The perspective the author endorses reveals the belief in a sort of internal convergence between semi- or even unconscious *powers* that are not able to overcome the irruption of the power of conflict and the will (as being also a power) to surpass it. The core-thesis is that between the individual

and the collective there is often a gap or even an abyss of forces in the relation man has to and with the world. The feeling of the world's crushing into our deepest desires, and amounting, then, to a complete annihilation of all of them (or of some of them) leads, in the end of the spectrum, to an inversion of the balance of the forces, to a situation where the "pathological man" assumes an omnipotent posture. The specific case of schizophrenia has to do, not rarely, with "an abnormal sense of power", accompanied often with a singular aesthetical experience where the value of the power turns to the value of the *will to power*.

João Santos' essay called ***Experiencing the World: John McDowell and the Role of Sensibility*** deals with the issue of knowing whether we can find a sheer *co-operation* between the Kantian figures, *sensibility* and *understanding* (or even a radical metaphysical separation between them), through a re-fashioning of the terms of the discussion as presented in some contemporary analytic philosophers, such as John McDowell and Wilfrid Sellars. McDowell's proposal on the "absence of a metaphysical gap" between *mind and world* lies on the assumption that the role of sensibility as a conceptualizing activity already bears the power of a keynote in his whole work. Thus, he assures that the content of experience and the content of judgments do not represent distinct realms within the subject's transcendental structures of knowledge. This particular view stands against the one defended by the Myth of the Given, which considers the facts from the 'world as first perceived by the *a priori* forms of sensibility, and then, only after, to be put under the scope of the understanding's structures. This very "framework of the Givenness", as Sellars calls it, is also the one characteristic of traditional empiricism, leading McDowell to think there is a breach between the conceptual and the non-conceptual, which does not allow for the possibility of overcoming the Myth of the Given. If this were so it would not be possible to envisage even a minimal bridging from the world to the mind. So: McDowell argues for the unboundedness of the conceptual, which means that there is no metaphysical gap between things and ourselves, that the traditional Kantian proposal, or the way it has traditionally been regarded, of the distinction phenomenon/things in themselves, does not make sense at all. Since the subject's openness to the world is now considered as absolute in the way that every object of the world is *possibly thinkable.* This same view entails, thus, another sort of transcendental idealism, which is, at the same stroke, a transcendental realism. McDowell thinks that all of these insights are, although in a non-explicit fashion, present in the first *Critique.* The main purpose of the essay's author, João Santos, is to regard McDowell's *unboundedness of the*

conceptual claim as a non-totally clarified and sustaining position, which has led him to believe that even if the Kantian "transcendental machinery" covers the potential to explore the form of a content, that does not mean that that which is non-discursive or non-conscious to be totally unveiled by what McDowell calls *sensibility in operation*.

It is a quote from Descartes *Discourse on Method* that opens **Tero Vaaja**'s ***Meeting Other Minds*** essay. Such a quote allows our author to start by briefly confronting the literature on whether Descartes endorsed some version of an analogical argument for the existence of other minds or whether he thought that a single judgment is enough to assure us of the same conclusion – because he worked on the assumption that all and only beings with human shape and form have a mind. Either way, one can at least safely say that Descartes' opened up the logical possibility of a human-shaped automaton lacking a real human mind. The case being so, Vaaja's starting point is thus the problem that, unless one shares Descartes' mentioned assumption, some form of an argument from analogy is needed to assert the existence of other minds. Vaaja points out the support the analogical argument has received from many philosophers but at the same time does not fail to mention the many criticisms it has endured, from the weak basis of generalization it starts with to the unverifiable conclusion it leads to, not forgetting Wittgenstein's conceptual (formulation of the) problem. More importantly, Vaaja's own critique of the argument is stated in a very clear way: even if one were to accept the argument from analogy, all one would gain would be a probability, albeit a very high probability, of the existence of other minds. These minds' existence, in themselves, would still be beyond our reach and requiring an inferential leap to be gained knowledge of. This way of proceeding does not seem to be, Vaaja argues, the way we actually behave when we attribute a mind to other human shaped beings – we respond to some behaviors assuming a mind is being made manifest there, we do not observe such behaviors and then infer a mind from them. Vaaja's first attempt to deal with the problem is by resorting to a criterial approach, based on Wittgenstein's later works. This endeavor is grounded on the idea that some behavioral characteristics are a special kind of evidence of the existence of mental states, because there is a logical link between them such that no inference is required. The crucial point here is the way one interprets Wittgenstein's notion of "criterion": it is not, Vaaja argues, to be read as if the criterion entails or is logically equivalent to the existence of the phenomenon it is a criterion of. Rather, the criterion seems to be better understood as something more than a mere sign of the phenomenon but still short of a defining criterion. Drawing upon other

readings of Wittgenstein's, Vaaja concludes that criteria are expected to bear two mutually exclusive features: defeasibility and anti-skeptical power, which, as McDowell as put it, seems straightforwardly incoherent.

Vaaja does not try to refute the charge of incoherence but he strives for a reading of Wittgenstein's words on which behavioral criteria are a special kind of evidence. The answer, motivated by a passage in *Philosophical Investigations* that uses pain as an example, is based on the idea that some behavioral patterns are the paradigm cases for the correct application of the concept of a determined mental state. In the case of pain, when a person who has mastered the use of pain-language observes pain-behavior there is no inference being made, but only a single judgment of attribution. Vaaja notes, however, that this may still not be enough to satisfy the skeptic and that perhaps we still have to either recognize the need for the analogical argument or to try to defuse skepticism at a different level, showing their incoherence of illegitimacy, as Wittgenstein in *On Certainty* or as McDowell in a 1982 paper, "Criteria, Defeasability and Knowledge". McDowell's refutation of skepticism about the external world is then paralleled with a possible refutation of skepticism about other minds, prompting Vaaja to an excursion on this author's views on both problems and the way they can relate to one another. In the end, however, Vaaja seems to concede that some difference in character between those two kinds of skepticism will remain and resorts to Stanley Cavell's insight that the extent and nature of the inner lives of others is something we will never be fully certain of. A conclusion is then put forth that remembers Wittgenstein's support for the idea that it is our cognitive effort that makes human behavior meaningful, which means that even if one can understand the skeptic's supposition about other minds, one can't do anything based on it nor, as P.F. Strawson put, will oneself to believe it or not.

Daniel Ramalho's paper ***Paul Churchland's Call for a Paradigm Shift in Cognitive Science*** is both an outline and an endorsement of Churchland's philosophical work and of his arguments for the connectionist paradigm in cognitive science. To fully appreciate Churchland's work one needs to understand the paradigm he is fighting and so Ramalho begins his paper with an historical tour through the origins and evolution of the symbolic paradigm, an exercise one could also describe as trip down the memory lane of cognitive science. In fact, from Turing to Putnam, from von Neumann to Chomsky, our author swiftly but efficiently portraits the rise of computationalism and functionalism in cognitive science and the way the symbolic paradigm of cognition brings about a focus on the

algorithmic leves of cognitive processes rather than on the neurological or behavioral, thus leading to a privilege of the software over the hardware and to a search for the "Platonic Function" that can be implemented on any brain. Ramalho then turns to the historical development of connectionism, beginning with McCulloch and Pitts description in 1943 of a computational architecture inspired in the physical structure of biological brains that demonstrated connectionist systems' computational power of a Universal Turing Machine. Our author proceeds with Frank Rosenblatt's 1958 development of the perceptron neural network design, carefully explaining the computational architecture and the workings of this early model, as well as pointing out the need for a human "tutor" as one of the features that stalled research in artificial neural networks until the mid-1980's, when the backpropagation algorithm emerged and provided those networks with a learning method free of human intervention. The backpropagation algorithm allows Ramalho to introduce Churchland's theory of cognition, the theory of parallel distributed processing (PDP), and two essential features of Curchland's connectionist model of cognition: the Domain-Portrayal Semantics model and the Dynamical-Profile Approach. While the latter refers to how the brain generates consciousness from the processing of information, the former points to the way the brain acquires and stores information. Ramalho highlights the coherence between these two theoretical proposals, given that the Dynamical-Profile Approach construes consciousness as being wholly independent from its subject matter, thus making it not about what is being processed in the brain but about how.

One of the most important aspects of this essay is the overview it provides of the argumentative arsenal Churchland raises against the symbolic paradigm and in support of his plea for a paradigm shift in cognitive science. Ramalho uses three categories to accommodate all those arguments: biological plausibility, eliminative materialism and reductionism. Regarding the first category of arguments, it is safe to say they identify discrepancies between digital computers and brains in terms of their processing and structure that allow one to conclude with the symbolic paradigm's inability to account for human cognition. The second category, that of eliminative materialism, pertains to Churchland's rejection of folk-psychology as an accurate or even useful description of cognitive activity, with four reasons being identified (poor explanatory depth, narrow explanatory breadth, too much reliance on propositional attitudes and inability to deal with nonhuman animal cognition). Lastly, the third category of arguments, that of reductionism, involves the accusation that

the symbolic approach offers no promise of an intertheoretic reduction of its account of cognition, unlike, of course, the connectionist approach. A note is in order, here, to stress that Ramalho's closing comments address not only the compatibility of Churchland's reductionist proposal with the multiple realizability argument of functionalism (as long as the latter, claiming the instantiation of mental states to be possible in a multitude of different physical substrates, construes of those mental states in the light of neural networks and not digital computers) but also what lies ahead for cognitive science - in particular for Churchland's call for cognitive science to return to infancy (both in the sense of revisiting Turing's ideas and in the sense of paying more attention to the child's brain).

The idea of moral personhood under fire is **Oscar Horta**'s contribution to this volume. In it, the author aims his guns at the concept of personhood in general but, specifically, to the application of such concept in ethical discussions. Horta's two main claims in this essay are 1) that personhood, as currently understood in ethics as in other fields, cannot be considered an attribute coextensive with humanity, 2) that we should abandon the language of moral personhood altogether. In order to defend his claims, Horta starts his paper with a digression on the meaning the term person has in ethics and in other fields (metaphysics, law, common language). "Moral person" can be understood in a number of different ways but the common trait is that all of them treat those individuals that fall under such classification as privileged (morally considerable, endowed with morally relevant interests or moral agents) by opposition to those that don't. Metaphysical persons can be viewed either as primary, basic constituents of the world, or as constructs reducible to some more basic elements, legal persons (such as corporations) can be considered either as real entities or as mere fictions, common language persons are equivalent to human beings and persons *simpliciter* are persons in every sense hitherto analysed. According to Horta, there is a common assumption that humans are persons in every sense of the term (and in some accounts humans are persons *simpliciter*). In this common assumption, whatever the meaning of the term "person" is and whatever the field it is being used in, humans are referred by it. This common assumption, although universally approved, is pure and simply wrong, Horta claims. Horta's way of showing this is through the acknowledgment that such meaning actually differs significantly whether the term is used in common language or in the moral, legal or metaphysical realms. There are many humans who would not qualify as metaphysical persons under some criteria for personhood and there are nonhuman animals which would satisfy criteria for metaphysical

personhood. In much the same way, there are moral persons who are not humans and not all humans are moral persons.

Only in common language, then, is "personhood" coextensive with "humanity". But common language is no ground to accept the common assumption and so, Horta argues, one must look closer at it. The author's wager here is that if one does take such closer look, anthropocentrism will stare right back at us. Indeed, regardless of the specific criterion appealed to in order to defend moral personhood, the fact is that all of those criteria end up with the same kind of beings as moral persons: humans. The same is pretty much true for metaphysical personhood theorists, Horta claims, and also for legal personhood: in the end, it is membership of the human species that really matters. Nonetheless, Horta's attack on personhood is not yet laid to rest. There are further reasons not to use personhood, particularly in the moral field, and those have to do with it being superfluous. Consider the disjunction: either personhood is based on some morally relevant criterion or it is not based on some morally relevant criterion. If personhood is based on some morally relevant criterion, then why waste time with personhood? It seems being called a person makes absolutely no moral difference, because the moral difference operates at the level of the morally relevant criterion on which personhood was based. The other option, that personhood is not based on some morally relevant criterion, is not much better: who wants to rest their moral theory on a notion that has no morally relevant justification?

It can be argued that the real target of Horta's attack on personhood is the picture in which humans (and only humans) are entities of a certain special kind. In this anthropocentric picture, humans are the kind of beings that are morally considerable or deserve special moral consideration. And since, according to Horta and many other authors, this picture is a result (at least in part) of resorting to the notion of "person", it follows that the widespread reliance on "person" from our ethicists is to blame for such anthropocentrism. If this were all there is to it, one could then say that perhaps Horta should ponder whether or not the concept of "person" still has some task to fulfill. Horta's reflections on the conceptual elaboration of "personhood" and its other flaws, however, seem to show very clearly that he sees no use left for "person" in moral philosophy.

Rui Vieira da Cunha's paper, ***Theories of Personhood: Guilty as Charged?*** aims to discuss some of the most common objections to personhood theories. There is a growing number of objections to the possibility and the practical use of personhood and theories based on it, particularly in bioethics, and chief on this attack has been the suspicion

that many in the philosophical literature have expressed about the concept of person, both in its metaphysical and moral aspects. The author tries to summarize the objections leveled against this concept by different authors from different theoretical standpoints and argues that they fall into four groups of charges, discriminated as the over-simplification charge, the charge of vagueness/ambiguity, the cover-up/begging the question charge, and the irrelevance/superfluousness charge. In dealing with these objections, Rui Vieira da Cunha uses Bert Gordijn's (1999) basic formulations of the charges, subsuming under the former many other similar objections from other authors. One of the main claims of the essay is that we can answer all the charges in a non-problematic way, because most of these charges are unjustified or, at least, they could equally well be applied to many other concepts who play a fundamental role in philosophy, science, and in our own practical life. The author's ultimate interest, however, is not so much on the objections themselves but on showing how they all share a deeper metaphysical *cum* moral question. It is argued that in each charge we will inevitably reach the point where the metaphysical and the moral aspects of personhood meet and that is the crucial point of the paper: what kind of connection is there or should there be between metaphysical personhood and moral personhood or, to frame it in another possible manner, what is the link between the descriptive and the normative aspects of the concept of person? The author does not attempt to answer the question but only to highlight it and warn about some misconceptions of the topic.

Luís Veríssimo's essay ***Julia Driver's 'Virtues of Ignorance'*** is a criticism of some features of Driver's consequentialist account of virtue, namely of its criticism of the Aristotelian notion of virtue and of its demarcation of a special class of virtues – the 'virtues of ignorance'. The author begins his paper with an analysis of Driver's interpretation of Aristotle's understanding of virtue. According to Driver, a central feature of Aristotle's account of virtue is that it requires some sort of cognitive skill: to be virtuous one has to have a special sensitivity that makes one aware "of the morally relevant features in each case where a moral decision is called for", says Veríssimo, relating the "knowledge condition" of virtue ethics with its particularism. Although contemporary virtue ethicists, such as John McDowell and Martha Nussbaum, endorse this "knowledge condition", Julia Driver does not. In fact, Driver believes knowledge is neither a necessary nor a sufficient condition for virtue and her strategy, as Veríssimo presents it, is to proceed by providing counterexamples to the Aristotelian notion of virtue. These are situations where some sort of character trait that one wishes to consider as virtue is involved, although

knowledge is not. Such knowledge-free character traits constitute the so-called 'virtues of ignorance', with modesty being the chief example. Arguing against these 'virtues' is the bulk of Veríssimo's paper.

Driver's defense of these traits as virtuous is made firstly on the basis of them being recognized by others as improving the holder's character, a defence Veríssimo rightly criticizes. Indeed, as our author puts it, such recognition by others may not occur in every society, for instance, not to mention that it seems arbitrary to ground our assessment of virtuous traits on the recognition of others. Besides, Veríssimo points out that there is no way of knowing how many of those others are needed for the recognition to go through and, on top of all of that, even if all of the others agree on the recognition, they might very well be wrong. More promise is found in Driver's defense of these traits as virtuous by resorting to a controversial consequentialist definition of virtue: these traits are virtuous because, like other moral virtues, they tend to produce beneficial effects. Modesty is Driver's chief example and it is her requirement that any account of modesty must explain why is the assertion "I am modest" self-defeating. Driver runs through four approaches to the concept of modesty, namely modesty as the careful avoidance of boastfulness, modesty as understatement (false modesty), modesty as the recognition of luck or efforts of others and modesty as underestimation, and after pondering their problems, advocates for modesty as underestimation, a view of modesty Veríssimo thus proceeds to criticize.

In fact, for Veríssimo modesty does not involve underestimation or any ignorance trait, but rather the "disposition to frame one's self-worth in a wider picture, and thus relativize his achievements". This wider picture is one that requires knowledge, not ignorance, of one's full potential, of the talents of others in the same area, of luck and train factors, and of the variety of talents. Such a conception of modesty as our author sponsors is ultimately an ability to relativize self-worth. This allows, note, that one can them say that the assertion "I am modest" is not self-defeating while at the same time accommodate Driver's definition of virtue, by claiming that such an ability or disposition would bring benefits to others and myself. With this result in hand, Veríssimo then proceeds to apply this approach to other 'virtues of ignorance', such as 'blind charity', 'blind trust', 'the disposition to forgive and forget' and 'impulsive courage', only to conclude that in neither of these cases is ignorance required, as Driver claims, but, on the contrary, it is knowledge that can qualify such traits as virtuous on Driver's account of virtue (the likelihood of beneficial effects). The upshot

is obvious: if there is no ignorance involved in those so-called 'virtues of ignorance', then we can safely say that this is an empty class and no real counterexamples to the Aristotelian knowledge condition have actually been presented, claims Veríssimo, before closing the essay with a small digression of the advantages and the prospects of virtue ethics adopting an universalistic perspective.

Having thus introduced the essays that constitue this book, we believe it is now clear that ***From Minds to Persons*** pursues the intent of combining different approaches to the topic of the thinking subject – among others, as perceiver, as consciousness as reason, as person, as a power to act, as inquiring subject, as cognitive agent. An extended range of ontological, epistemological and ethical commitments prove to be inevitable when dealing with the question about the nature of the connection between minds and persons. In order to make that clear, a great number of philosophical traditions were gathered in this book, which also exhibits the sheer richness such a wide theme provides. There are many possible approaches to the topic of subjectivity, and the idea often arises that some of them imply antithetical positions (which in fact is true). However, despite this difficult situation, quite confusing in some ways, we think that having some of these views in perspective and even dialogue, as we have tried to do in this book, may lead us to a more comprehensive view on what are the main premises and conclusions concerning such topics (even among the ones we find to be traditionally separated), even if that is done at the cost of a contrasting exercise among them. In fact, if nothing else, we would consider that as the main achievement of the present volume. What started as the proceedings of a Graduate Conference turned out to be a book with its own unifying theme, unintentionally reflecting the common ground our different researchers tread.

Porto, December 2012

Rui Vieira da Cunha | Clara Morando | Sofia Miguens[1]

[1] The editors are very grateful, for their help in revising some of the essays, to MLAG members José Oliveira and João Santos.

BREWER AGAINST TRADITION

Manuela Teles

1

The central inquiry in contemporary philosophy of perception is whether perceptual experiences have representational content. Bill Brewer assumes this dispute to be about the subjective character of perceptual experiences. Following his perspective, representationalism and anti-representationalism can be taken to be the contemporary responses to a claim introduced by the early empiricists John Locke and George Berkeley. Both reject a traditional outcome from this claim, according to which the direct objects of perceptual experiences are mind-dependent entities of some sort: ideas, sensa, sense-impressions, etc. In different parts of his work, Brewer's proposal is that Locke and Berkeley presented different strategies to overcome this outcome, while trying to preserve their empiricist insight. My aim in this paper is to show that Brewer's presupposition is that representationalism responds to the sceptical threat inherited from Locke's indirect realism, while anti-representationalism avoids the idealist/phenomenological outcome of Berkeley's alternative to Locke. To accomplish my aim, I will focus on Brewer passage from a representationalist account of perceptual experiences, where he defended that *experiential presentation* is at the ground of perceptual experiences and is constituted by conceptual (demonstrative) contents, to an anti-representationalist account, where he rejects this on the basis of the notion of *perceptual presence*. I will question his reasons to this dropout of representationalism and consider the consequences of adopting, now, an anti-representationalist perspective.

2

Let's begin with Brewer's initial position in the nature of perceptual experiences. In "Do Sense Experiences Have Conceptual Contents?"[1], Brewer tells us that yes, they do have conceptual contents. In this paper, he intends to give us an argument to support the thesis that perceptual experiences have conceptual contents. His starting point is epistemological. Brewer's departure is that (i) sense experiential states provide reasons for empirical beliefs. The argument follows considering that (ii) a person has reasons for believing that her surroundings are the way they are experientially presented to her only if she is in a conceptual mental state. Since both premisses, (i) and (ii), obtain, Brewer concludes that (iii) perceptual experiences have conceptual content. Now, for a belief to be empirical is for it to be about the mind-independent physical world. For Brewer, a belief about the mind-independent physical world is grounded only if perceptual experiences play a certain role in their determination. To this role to be accomplished, two conditions need to be satisfied. First, the perceptual experience has to have a content of a kind that it has a certain form: it must be able to be a premise or the conclusion of an inference, i.e., a proposition. Second, the content of the perceptual experience, this proposition, has to be such constituted that the concepts that are its components are on the possession of the subject of the experience. This lead us to a (Brewerian) definition of conceptual content: it is the content of a possible judgement. I may need to use this later. If perceptual experiences provide reasons for an empirical belief, according to Brewer, then their conceptual content must be of a kind that it can (and should) be part of the reasoning to support it. But in Brewer's account this is not enough. The conceptual content of perceptual experiences has to be demonstrative in order for those reasons, it provides for an empirical belief, to be reasons *of the subject* of the relevant perceptual experience. Brewer's premises to hold that perceptual experiences provide reasons for empirical beliefs are the conclusions of two other arguments. The first is a result of Brewer's reading of P. F. Strawson. Brewer highlights from Strawson's writings that basic empirical beliefs have the contents they have only in virtue of its *relation* to perceptual experience. At stake is that *description* is not sufficient for referring spatial (physical) particulars. Successful such reference must be anchored on the experiential presentation to the subject of the mind-independent things in question. Brewer's outcome from Strawson is that the reasons-giving relation between perceptual experiences and empirical beliefs *depends* on a strawsonian demonstrative reference. Brewer says:

> "The key to understanding perceptual knowledge lies in exploring the interconnections between the philosophical logic and epistemology of perceptual demonstrative thoughts"[2].

[1] In M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology*. Oxford: Blackwell, 2005

[2] Précis of Perception and Reason, and response to commentators (Naomi Eilan, Richard Fumerton, Susan Hurley, and Michael Martin) for a book symposium in Philosophy and Phenomenological Research, 2001.

As we shall see, the demonstrative nature of the conceptual content of perceptual experiences is double. It applies both to the part of the content that refers to the object of perception and the part that refers to its predicate. The other argument on which Brewer supports his thesis that perceptual experiences provide reasons for empirical beliefs is that which concludes that between perceptual experiences and empirical beliefs have a content-fixing relation. This relation is that the perceptual experience determines the content of an empirical belief informing its subject which mind-independent it is to be about.

Brewer's motivation for perceptual conceptualism is epistemological. In his *Perception and Reason*[3], he endorses a criticism to analytical epistemology, considering that its versions of foundationalism, coherentism and reliabilism fail to ground empirical knowledge. For such epistemological theories are heedless on the subject's grasp of the relevant contents. I do not intend to say much more on this except that Brewer considers that, in ignoring the conceptual demonstrative contents of a perceptual experience, these theories cannot account for the basic ground of perception which is precisely the presentation of the mind-independent entities of the physical world as it is experienced by the subject of the relevant empirical belief. For Brewer, the analytical epistemology approach to perception is a second-order one since that which is taken into account is the subject's reflection on the perceptual experience, hers. Therefore, the perceptual experience itself, the very presentation of mind-independent entities from the physical world, is kept away from the account. Analytical epistemology, Brewer claims, has no place for an investigation of the subject's possession of empirical beliefs and understanding of the contents of those beliefs. It takes both for granted. A first-order account is needed if these are to be involved in the reasons-giving relation between perceptual experiences and empirical beliefs. This is Brewer's motivation for the conceptualist account of perception. According to him:

> "The crucial epistemological role of experiences lies in their essential contribution to the subject's understanding of certain perceptual demonstrative content"[4].

Thus, Brewer earlier perspective on perception is focused on the nature of demonstrative thought. In his conceptualist account, the subject of a perceptual experience endorses reasons to hold the relevant empirical belief while grasping those demonstrative contents. To grasp the demonstrative content of perceptual experiences is for the subject to have a "genuine" reason to endorse it in belief. Only demonstrative

[3] *Perception and Reason.*Oxford: OUP, 1999.

[4] *Philosophical Books*, Book Symposium. Summary of *Perception and Reason*, p.1.

conceptual contents are of the kind and form needed to be both identified by a subject as possible judgement and furnish her with her reasons to endorse an empirical belief. Thus, the earlier Brewer is committed to a representationalist view of perception: a particular perceptual experience has a reasons-giving relation to a particular empirical belief only if the subject understands the content of the first as the content of the second, which is demonstrative and thus, conceptual. Now, what are the later Brewer's motivations to reject this view?

3

In "Perception and Content"[5] Brewer points to a fatal dilemma for the view he just defended. He calls it now "the Content View". The Content View is representationalist in that it claims that perceptual experiences are to be characterizable by *the way* they represent things being in the mind-independent physical world. A more appropriate manner to describe this characterization is to say that, for the Content View, perceptual experiences represent the subject's surroundings. The dilemma can arrive from the representationalist view of the earlier Brewer. In following his arguments, one is driven to ask whether his perspective on perception and beliefs is such that in a reasons-giving relation between a perceptual experience and an empirical belief, the content of both is (or is not) the same. The later Brewer claims that if the representationalist answer is positive, which is her natural answer, then she is pushed to the obligation of being clear on *what* must be added to the content of perceptual experience in order to distinguish it from a propositional attitude and to account for its subjective nature. In such a position, the representationalist may wish to answer negatively. But in denying that the content of perceptual experiences and empirical beliefs is the same, she looses the very notion of content. I will call this the Content Dilemma. Brewer uses the term "genuine" to qualify *that* which is supposed to represent things as being such and such. I add that a genuine content is, roughly, a thought: that which is taken to be true (or false) in a judgement and propositional attitudes in general. Brewer develops three models of content upon which the Content View is built. I take these models to be three versions of the Content View. According to each of these versions, the content of perceptual experience is either (i) a *proposition*, (ii) a demonstrative sense which is dependent on the existence of an object, or (iii) a demonstrative sense which is dependent on the existence of an object and the instantiation of properties. Brewer's conceptualism is a representative of version (iii), which the earlier Brewer considers to be an improvement of version (ii). He explicitly says that

[5] "Perception and Content". *European Journal of Philosophy*, 2006.

his motivations for the version (iii) of the Content View was McDowell's version (ii). For Brewer, in elaborating the basic version (i) in order to allow for constituents of the world to be constituents of the content of perceptual experiences, McDowell offered the Content View the chance to defend that "perceptual experience presents us directly with the objects in the world around us themselves"[6]. I take this to be Brewer's motivation both for his earlier conceptualism and his latter anti-representationalism. I will avoid to go into McDowell's proposals themselves and try to elucidate how Brewer interprets them. Suppose that "that is green" expresses the content of a visual experience of Alfred. In the expression, "that" refers to a strawberry Alfred is seeing. "That" is also a demonstrative concept of which Alfred has possession and that is used, or may be used, in particular situations, namely, when he is having a perception. Since the concept "that" is put to work only when a particular object is there to be perceived, there is no question about the existence of the strawberry (or whatever object one can refer by "that"). On version (ii) of the Content View, the particular object, the strawberry, referred by "that" exists and the rest of the content depends on its existence. Alfred cannot fail on the existence of the strawberry and this is a way the world is. A way in which there is a strawberry that Alfred is seeing. However, it still allows for the possibility of falsehood. Alfred can fail about the way experience represents "that" to him. In fact, Alfred is colour-blind so he cannot see the strawberry as it is, red, since it is among the green leaves of the strawberry plant and he cannot distinguish red from green. The version (iii) of the Content View is Brewer's own contribution. His intention is to extend the demonstrative nature of the object reference to that of the properties reference. In his version (iii), all parts of the content have a demonstrative relation with the physical world mind-independent entities, including its predicate. Thus, seeing that which is referred by "that" as green is, for Alfred, to see it "thus" and "thus" has a demonstrative sense, which is dependent on how the world where Alfred is seeing the strawberry actually is.

For the later Brewer, the Content View is dependent on two features of the representational content that prevent it to overcome the Content Dilemma. These features are present in all the three versions. For a content to be representational, it has to be able to be false and be general. The possibility of falsehood and the involvement of generality are the main faults Brewer attributes to the Content View. They are fault in the sense that not only they preclude the Content View to be free of the Content Dilemma but mainly because they do not allow it to be a genuine account of the subjective character of a perceptual experience. The Content View stops at the experiential presentation when it should go until perceptual

[6] Idem, p.6.

presence. Here, I will consider only the possibility of falsehood and leave the involvement of generality out of my concern. According to Brewer, perceptual experiences admit the possibility of falsehood since their content can be determined as true or false based on how things are "out here". This is what representationalism is supposed to be and also the core thesis of the Content View. On Brewer's words, to a representationalist:

> "Genuine perception involves a successful match between mind and world, between content and fact, which might instead have been otherwise, in correspondingly unsuccessful cases[7].

It is worth to notice that the analogy between mind and content, in one side, and world and fact, on the other is not straightforward. If content is to be something like a thought, or a thought itself, then it is contentious that is analogous to something of a mental nature. Or at least this is so if one respect Gottlob Frege's claim, according to which the question of truth (and falsity) arrives only where there is a thought. For Frege, thoughts are objective, public, shareable. They cannot be compared to what is subjective, private and belong only to one owner. Anyway, this is not a point I wish to follow here. At least not now. For the moment, I just follow Brewer's point against the three versions of the Content View. According to him, the possibility of falsehood is present even in version (iii), where contents are demonstrative relatively both to the object and to the properties. In version (ii), the possibility of falsehood is straightforward: although that which is referred by the demonstrative "that" cannot be false, the way this constituent is represented can be false. This is one of the possibilities for the Content View to explain *illusions*. Indeed, representationalists take this explanation of illusions as perceptual experiences with false contents as a strong motivation for the Content View. Yet, the later Brewer claims, appealing to the possibility of falsehood can be used to reject the Content View. If representational contents are considered to be *dependant* on particular objects and/or instantiated properties, how can they be false? If they cannot be false, they cannot be true either and if they cannot be any of them, they simply are not representational, or even contents. Brewer's earlier version (iii) of the Content View is an example of that self-destructive feature of the Content View. In Brewer's version (iii) of the Content View, the possibility of falsehood may be not as straightforward as it is in the other two versions. Would this would protect it from the later Brewer accusation that it is self-destructive?

[7] Idem, p.4.

4

A possible way out of the difficulty for the Content View, introduced by the possibility of falsehood upon which it is dependent, would be to consider an alternative definition of illusion. I believe that it is here that Brewer's passage from representationalism to anti-representationalism begins. As we saw, his version (iii) of the Content View faces a difficulty. It is dependent on the possibility of falsehood to be representationalist. However, in trying to be faithful to the experiential *demonstrative* presentation of the mind-independent physical entities of the world, it narrows that same possibility. If contents are dependent on the existence of particular objects and of the instantiation of particular properties, they simply cannot be false. For contents, not being able to be false, as we just saw, is a step not to be contents at all. But even ignoring this fatal consequence, version (iii) of the Content View would be in trouble. If contents cannot be false but are that which characterizes a perceptual experience, and if illusions are perceptual experiences, how can the Content View explain illusions? Facing this difficulty, the proponent of version (iii), the earlier Brewer himself, would have to question the thesis that illusions are perceptual experiences with false representational content. Yet, if he wants to answer that they are, he might have to reject that all perceptual experiences have representational content and, thus, that they are to be characterized by their representational content. It he wants to answer that they are not, he might end to be committed with the very counterintuitive thesis that illusions are not perceptual experiences. I believe that the passage Brewer takes from representationalism to anti-representationalism begins when he tries to improve the situation for the version (iii) of the Content View. Faced with the difficulty, its proponent would first try to keep the thesis that illusions are perceptual experiences with representational content. But in doing so, he would have to reformulate his starting points, abandoning the demonstrative nature of the thoughts that are to be the contents of perceptual experience. He would probably want to "descend" one level, and allow that only the part that corresponds to the object referred by an expression of the content is demonstrative. This would be a version (ii) of the Content View. But then, the representationalist claim would be weaker, since one part of the content was not subject to the possibility of falsehood. But, again, being, at least in one part, dependent of the existence of particular objects, how can the representational content of a perceptual experience be false? And not being able to be false, how can it even be a content? Here, the proponent of the Content View would probably try to descend one more level and get rid of the demonstratives in order to explain illusions as perceptual experiences with representational content. But, now, without demonstrative reference, he would loose the experiential presentation of the mind-independent objects (and properties) of the physical world. The proponent of the Content View would, thus, be in such a position that he would no longer be able to sustain that there is representational content *in*

perceptual experiences. The result would be that perceptual experiences are not characterizable by their representational content. At this point, the proponent of the Content View would try to answer that, no, illusions are not perceptual experiences with false representational content. This might help him to save version (iii). But then he would have to explain how can illusion be perceptual experiences and have no representational content. The Content View stands on the possibility of falsehood to explain illusions, so, if illusions are not perceptual experiences with false representational content, how is the Content View to explain them? But even worst for the proponent of version (iii) of the Content View is that if illusions were not to be considered to be perceptual experiences with false representational content, and the representational contents of perceptual experiences is to be demonstrative in its whole, what would be the role of falsehood? And, again, if there is no possibility of falsehood for perceptual experiences, there simply is no room to truth, and, thus, no room to content. The Content Viewer would, thus, be forced to reject that illusions are perceptual experiences with false representational content and, with it, that perceptual experiences have representational content. Brewer's conclusion is that the Content View is self-destructive. In trying to give a characterization of the most basic level of perceptual experiences, that in which a subject is directly aware of the mind-independent things in the physical world by describing its content, the Content View simply looses it. The Content View has no means to explain illusions. Therefore, an explanation of illusions cannot support or motivate it anymore.

5

So, those who are still interested in characterizing the subjective nature of perceptual experiences, that same level in which mind "touches" the world, simply have to choose an alternative account of perception.

> "The only alternative to characterizing experience by its representational content is to characterize it as a direct presentation to the subject of certain objects, which themselves constitute the way things are for him in enjoying that experience. Call these the direct objects of experience: the objects which constitute the subjective character of perceptual experiences" ("Perception and Content", p.6).

Brewer claims that perceptual experiences present us with the objects in the world around us themselves. He calls this "the Object View". According to Brewer, the quest for the subjective character of a perceptual experience is an inheritance of early empiricists accounts, such as those from John Locke and George Berkeley. I will define is as:

Early Empiricism (EE):
The subjective character of perceptual experiences is to be given by citing its direct objects.

For Brewer, the direct objects of perceptual experiences are *that* of which the perceptual experience's subject is aware while having the experience. It is a salient feature in the stream of consciousness; *entities* with which the subject is acquainted. This is why to cite the direct objects of a perceptual experience is to give what *in it* is subjective. So, answering positively to the dilemma, the Content View theorist could add EE to the thesis that perceptual experiences have representational contents. A perceptual experience would, thus, be characterized by its content *and its direct objects*. Why is this not acceptable for the later Brewer? In my opinion, the answer is already, although implicitly, present in his earlier account of perception. He builds the Object View on the basis of Berkeley's critiques to Locke. What is at stake in these critiques is the metaphysical nature of the world and our minds. The earlier Brewer objections to the analytical epistemology are motivated by what he sees as a lack of place for what is genuinely subjective in a perceptual experience. The idea is that in the pure subjective level of a perceptual experience what is to find is an experiential presentation of the mind-independent entities of the physical world. And this is why epistemology has to move deeper in perception in order to explain how it provides reasons to empirical beliefs. Thus, it is precisely in Brewer's motivations against those analytical epistemological theories that one can already find a concern with what is properly subjective in a perceptual experience. For the earlier Brewer, reasons for empirical beliefs are genuine only if they are reasons for the subject to hold a certain empirical belief, and reasons are for the subject only if they are provided by her perceptual experience. What is unacceptable to the later Brewer is that his earlier view took that subjective character of a perceptual experience to be exhausted by the demonstrative representational content. Apparently, this could withdraw the representationalist from the Content Dilemma, since she would not find it necessary to add anything else to a perceptual experience characterization. But it wouldn't work. As Brewer seems to stress, the representationalist would be pushed to the Content Dilemma simply by claiming that the representational content that characterizes a perceptual experience is the representational content of a possible propositional attitude. Therefore, the point is still to give an account of what, in a perceptual experience, is subjective. And the subjective character of perceptual experiences enters in the puzzle even against the representationalist will.

6

What is Alfred aware of when he sees the strawberry? From a commonsense perspective, the answer is plain and simple: in seeing *the strawberry*, Alfred is aware of *the strawberry*. Alfred himself, while seeing the strawberry believes to be aware of *that same* strawberry. This is a perspective usually called Naïve Realism.

Naïve Realism (NR):
The objects of perception are physical things.

Some philosophical perspectives, though, reject this plain and simple answer. Their rejection is based on a set of claims that have been used as a classical inference called the Argument from Illusion. The Argument from Illusion is a contentious theme in the philosophy of perception. Several versions of it have been put forward both by its proponents and by its critics. Following John Austin[8], but not only, I will consider a version that divides it in two. It is a very short version and it goes like this. The first sub-argument of the Argument from Illusion concludes from EE and one other thesis, to be shown right away that the direct objects of illusions cannot be physical things. The other two theses are the following:

> Illusion (ILL):
> An illusion is a perceptual experience in which an object *o* looks F to a subject when *o* is not F.

The second sub-argument of the Argument from Illusion adds one third thesis on perception to the first conclusion:

> Subjective Indistinguishability (SI):
> For the subject of a perceptual experience it is impossible to discern whether it is a delusional[9] or a non-delusional one.

Its conclusion is a negation of NR: the direct objects of perception are not physical things. I am stating it as:

> AI:
> The direct objects of perception are mind-dependent.

From here, an opposing view to the Object View can be mounted. Having rejected the Content View of perception, Brewer is willing to carefully expose and reject this view. He doesn't give it a name. I will call it the Idea View[10]. According to the Idea View, the direct objects of perceptual experiences are of a mental nature. But this is an unacceptable result for the contemporary views on perception. In one hand, it is considered to be counterintuitive. In the other, it is taken to be at odds with the scientific achievements of the modern ages. In brief, the thesis that the direct objects of perceptual experiences are mental is incompatible with the contemporary realistic perspective. Both the Content View and the Object View move against the outcome of the Idea View and the difficulties

[8] Austin, J. L., *Sense and Sensibilia,* Oxford University Press, 1962.

[9] I am purposefully being heedless on the question whether only illusions should be included in what I am considering a delusional perceptual experience or if hallucinations should be included.

[10] I am deeply unsatisfied with this name and having such a hard time to find it a proper designation I wonder if that was Brewer's reason to let it unnamed... For the moment, let's think of it as a provisory name.

it brings to a realist broader perspective. The difference between them is that they use different strategies to deny the conclusion of the Argument from Illusion. While the Content View rejects EE (even if only implicitly), the Object View wants to preserve it; thus, it rejects ILL. In Brewer writings one can find Locke and Berkeley being each a representative of each of these views. His aim is to show that Berkeley's objections against Locke are to be of great interested to the contemporary *realist* views on perception. In Brewer's analogy, Berkeley was for Locke as the Object View is to be for the Content View: a block against representationalism. According to Brewer, Locke tried to overcome the anti-realist outcome of the Argument from Illusion. His strategy was to add to the objects of perception a second layer. Besides the direct objects, perceptual experiences can have *indirect* objects. For Brewer, the lockean proposal entails an inconsistency. This inconsistency has its root at the notion of *resemblance*, from which Locke's perspective is dependent. I assume that, from Brewer's point of view, this is the kind of argument one finds in Locke:

- The direct objects of perceptual experiences are mind-dependent (AI).
- Physical objects are things of which we are aware in perception (NR).

Thus,

- physical objects are the ***indirect*** objects of perception.

So, for a lockean development of the rejection of NR follows what is known as "Indirect Realism". For Brewer the problem for this argument is that it depends on an explanation of the *resemblance* between the mind-dependent direct objects of perception and its mind-independent indirect objects. In Brewer's opinion, this resemblance is something that cannot be argued for. The Indirect Realism proponent best explanation is that physical indirect objects are the causes of mental direct objects, but this says nothing about the core of a theory of perception. For Brewer, it says nothing about the subjective character of perceptual experiences and in being heedless on that it is selfdestructive. If one considers that the subjective character of a perceptual experience is to be given citing its direct objects, one ought to be able to describe it (its subjective character), describing those direct objects. Nevertheless, if those direct objects are to be some kind of correspond of those other indirect objects and those indirect objects can be described only as causes, it is hard to see how one can get to any description of the direct objects we are firstly interested in. The outcome of such an argument is that all we can know about the indirect objects of perception, i.e., those physical object that we naturally believe to be that which we perceive, is that they are causes. Brewer? point here is that in rejecting DR, this lockean perspective cannot afford the account of perception itself needs. Indirect Realism must explain the direct objects of perception with its indirect objects, however if all that can be known about material objects is that they are causes for that of which we are aware in perceptual experiences, there

is not much to explain about the resemblance they should have in order for the proponent of Indirect Realism to be able to say that some perceptual experience is a perception. Being unable to tell about those causes, we are forever ignorant of the nature of the direct objects of our perceptual experiences and, thus, banned from the possibility of explaining their subjective character. The same subjective character one aims at explaining accepting EE. The result for Indirect Realism is that, ultimately it has to abandon EE, if it is to deny NR.

> "On Locke's materialist view, the direct object of an illusion is a mind-dependent entity, which is F, which nevertheless sufficiently resembles a non-F, mind-independent object, *o*, which is also appropriately causally responsible for its production, for the later to be the physical object, which is G"[11].

For Berkeley, the same premises should result in a different conclusion:

1. The direct objects of perceptual experiences are mind-dependent (AI).
2. Physical objects are things of which we are aware in perception (NR).

Thus,

3. physical objects are mind-dependent.

Notice that this conclusion is about the nature of the physical objects, while in Locke's argument, the conclusion was about the nature of perception. For Brewer, Berkeley's objection to Locke rests on the idea that given the constraints that physical objects need to impose on the direct objects of perception in order for there to be a resemblance (and thus a perception proper), those physical objects cannot be of a material nature. Instead, those same physical objects, that according to our commonsense conception are *the objects* of our perceptions, are to be broadly conceived either as "*mereological sums of* mind-dependent direct objects of perception" or as "*actual and possible* mind-dependent direct objects of perception". For brevity, I will avoid the discussion Brewer presents on Berkeley's metaphysics and resume it to the final idea that, according to Brewer, Berkeley's metaphysics is sustained by a single ontological thesis: that all there is is "minds and their ideas". This is a twist of Locke's indirect realism. Ultimately, Berkeley turns it into a form of empirical idealism.

In rejecting Locke's addition to the Argument from Illusion, the idea that physical objects are the indirect objects of (successful) perceptions, Berkeley rejects ILL. This is his insight Brewer wishes to recover. His

[11] *Perception and Its Objects.* Oxford: OUP, 2011, p.9.

project is to rehabilitate this insight in order to arrive to a realist ontology to which Brewer calls "empirical realism". According to Brewer, the insight comes from Berkeley's response to Locke's proposals on the nature of the direct objects of perceptual experiences. Berkeley rejects Locke's argument according to which the world is material and our minds have access to it indirectly when directly perceive ideas. Very roughly, Berkeley's point is that if Locke is correct and we directly perceive only ideas, and we must, simultaneously, admit that perception is about a mind-independent world, then the conclusion must be that the mind-independent world, with its objects and properties, is immaterial, just like our ideas. Berkeley's position can be read either as an *(empirical) idealism* or as an *(epistemic)phenomenalism*. In both cases, it is at odds with Locke's position. For Locke, the two thesis about perception, that we directly perceive ideas and that perception is an access to a mind-independent world, should lead us to some form of *(indirect) realism*. Notice that from a lockean perspective, with its origin in EE, if the direct object looks F then it is F. So, an illusion is a perceptual experience in which there is an object *o* which is F and is its direct object, and an object, say *o** which is not-F and its is indirect object. This is what drives Locke to consider that the direct object is not the physical object. Now, we can understand Brewer's project as a tentative to keep the Berkeleyan insight rejecting his idealism and phenomenalism. Thus, Locke's indirect realism, or any other's, is to be rejected. This is why he considers earlier, when proposing the Object View, that it is the only alternative to the Content View in explaining illusion. The indirect realist explains illusions too. But, like Locke, it explains them appealing, again, to some sort of representation, in this case, one of a pictorial nature[12], and this pulls away that direct contact with those mind-independent objects for them to be constituents of the perceptual experience itself. But Locke's realism is to be maintained. Berkeley's idealism is supported on the existence of God and this is something at odds with a contemporary view of perception, whether it is philosophical or scientific. The opposition Berkeley presented to Locke is for Brewer an analogy of the contemporary counterparts of the question on the subjective character of perceptual experiences. EE tells us that this is to be given by citing their direct objects. But how is this to be done? How are the direct objects of perceptual experiences to be identified? Tradition offered two alternative answers to this. The first is the Content View. For it, if there are direct objects of perceptual experiences, they are to be identified by the way they are represented in it. As we just saw, ways for things to be represented are thoughts. So, the direct objects

[12] As we saw earlier, for Brewer the only difference between the sort of representation involved in indirect realism and the sort of representation involved in the Content View is the form: for one is pictorial and the other linguistic. But the difference could be more radical if one rejects his implicit consideration of content as something mental. Taking a more fregean view, one could consider that the difference at stake is in nature, for the sort of representation in an indirect realist view is mental, while the one in the Content View *is not.*

of perceptual experiences are given as thoughts. In Brewer's version (iii), these are doubly demonstrative thoughts that are possible premises or conclusions of inferences and have their components in possession of the subject of the perceptual experience. The other alternative is that the direct objects of perceptual experiences are to be identified by that of which the subject is aware. Here, we arrive at the heart of the problem that an account of the subjective character of perceptual experiences has to face.

7

It is now time to introduce Brewer's alternative explanation of illusions. According to him, it was Berkeley that offered a first version of this explanation, when rejecting Locke's realism. For Berkeley, Brewer says, a perceptual experience is delusional because the its objects themselves have a delusional nature. So, the possible errors are not in the experience, or the subject, but in the objects themselves. To avoid the difficulty introduced by illusions, one could start to reject that illusions are perceptual experiences with false representational content. This is to be done taking EE and Berkeleyian insight into consideration. According to Brewer, both allow us to explain illusions without any appeal to false representational contents. How are illusions to be explained using EE and Berkeley? Appealing to features of the direct objects of perceptual experiences themselves. Illusions are perceptual experiences which the direct objects are delusional about themselves. One traditional example is the Müller-Lyer Illusion. Brewer also takes it to present his alternative definition of illusions. The most classic Müller-Lyer Illusion is an image composed of several lines in which two are taken to be horizontal and parallel and the rest are taken form four wedge shapes. The lines are composed in such a manner that each line has two wedge shapes placed in each of its points. The wedge shapes of one line are turned inwards while the other are turned outwards. Looking at the Müller-Lyer Illusion, a perceiver will see it as containing two parallel horizontal lines with diferent lenghts. The one connected with the turned inwards wedge shapes will look smaller than the one connected with the turned outwards wedge shapes. But the lines are exactly the same length. The different positions of the wedge shapes make them look different while they are not. Suppose Alfred is now looking at a drawing of the Müller-Lyer Illusion. For the Content View, his visual experience would be such that it would contain a false representation of the length of the parallel horizontal lines. For Brewer, this explanation would bring a proponent of the Content View to those difficulties we outlined earlier. What is his alternative explanation? I sketched the answer above. The alternative explanation of Alfred's perceptual illusion is to be found not in his experience but in its direct objects. Considering that the direct objects of Alfred's perceptual illusion just are the drawing of the Müller-

Lyer Illusion and all its components, Brewer considers that the origin of the delusion is its "visually relevant similarities with paradigms of various kinds" ("Perception and Its Objects", p.8). In "Perception and Its Objects", Brewer considers a paradigm to "a kind" of which we consider the direct objects of our perceptual experiences to be in. In the case of Müller-Lyer Illusion, when seeing the drawing, Alfred takes the parallel horizontal lines to be an instance of a paradigm of two lines of different length.

The notion of "paradigm" is better exposed in "How to Account for Illusion", where Brewer presents the Object View in contrast to the Content View's explanation of illusions. According to Brewer, as we have been considering, both are contemporary responses to EE. The Content View refuses to take direct objects as that which constitutes the nature of the subjective character of perceptual experiences. This refusal is a result of the outcome that comes from EE and the Argument from Illusion. We just noticed that, if one accepts both EE and the Argument from Illusion, one ends up with a form of indirect realism. This brings with it the thesis that the direct objects of perception are mind-dependent and, thus, either a lockeanist scepticism or a berkeleyian idealism/phenomenalism. So, the Content View explanation of illusions refuses that objects are that which gives us the subjective character of perceptual experiences. Rather, it is to be given by the way those objects are presented in experience, i.e., the representational content. Now, the Object View rejects that contents are the kind of things that can give us the subjective character of perceptual experiences. Its motivation is to preserve EE but do away with the anti-realist outcome. In my opinion, although not explicitly, in Brewer writings one find as a strategy for the Object View the negation of ILL. In fact, this is precisely the task of "How to Account For Illusion". Let's go back to ILL. Above, I defined it as the thesis according to which an illusion is a perceptual experience in which an object *o* looks F to a subject when *o* is not F. This is, of course, to be contrasted by what a veridical perceptual experience would be: a case of perceptual experience in which an object *o* looks F to a subject when *o* is F. Again, the Content View explains this, say, that the strawberry looks green to Alfred, rejecting EE. Then Alfred sees the strawberry looking green because his visual experience has a false representational content. For the Object View, this cannot be. Preserving EE allows us to consider that the strawberry looks green to Alfred, not because Alfred has a direct object of a mental nature that resembles the strawberry but is green and not red, but because it is phenomenologically classified as belonging to the paradigm of things that are green. In Locke's perspective, the direct object of an illusion is a mind-dependent object which is F and resembles its indirect, mind-independent, object which is not-F. Brewer's suggestion is that the Object View should take Berkeley's insight against the lockean view. The insight is that in an illusion one perceives *o* as being F when *o* is not F because the *o* we perceive as F is

a part of a composite physical object which is not-F (as a whole). The physical object, in Berkeley's response to lockean indirect realism, is not-F but looks F because the part perceived is F. So, there is only one object in Berkeley's mentalism. And this is the insight the Object View should respect. Alfred's case, thus, should be thought as a perceptual experience in which he is seeing the same strawberry that *is* red and *looks* green to him. Being green is to be taken as a part of the all properties a strawberry has. Among those, say, is the property of reflecting light rays we take as looking green as well as light rays we take as looking red. Alfred's "illusion" consists in simply not seeing part of those light rays. Nevertheless, he sees the strawberry. In fact, green is a possible *visible* property of surfaces that, paradigmatically and in normal conditions, we take as being red. Alfred would not see a lemon looking green since being yellow does not involve the possible visible green property. In seeing the strawberry that looks green to him he is seeing the strawberry. What he does not see is one of its properties, namely, the property of emitting the light rays that most other perceivers interpret as being red.

None of this has taken us, yet, to the notion of paradigm. I believe that Brewer, or at least his Object View, could do along without it. Still, it is an important counter-part for the alternative explanation of illusions by the Content View. Brewer explains paradigms as the instantiation of kinds associated with linguistic terms that we use to make our perceptions "intelligible" to us. Intelligibly, we take similarities of what is seen with a certain paradigm to be a qualitative identity. This would be for Alfred to intelligibly take the perceptual experience of seeing the strawberry looking green to be a case in which something is similar to all the other things we subsume to the concept GREEN. The result, it seems to me, might be that illusions are not a *perceptual* phenomenon. They are not at the level of perceptual presence. Instead, they come with the "conceptual phenomenology", which is based on paradigms and similarities and, thus, connected to the intelligibility of that which we see. But if it is so, then what is exactly the difference between the Content View and the Object View? The later Brewer would answer quickly that the difference is probably small, since both appeal to concepts. Yet, contrary to the Content View, the Object View does not need to appeal to content. To grasp concepts is not the same as endorsing a judgement. Brewer appeals to conceptual phenomenology as the classificatory engagement with what is presented to us. This is not to judge something but to subsume the particular object present in perceptual experiences under a concept. This strategy allows him to avoid an explanation in which representationalist content is needed and, thus, to avoid the Content View explanation of illusions. Phenomenology is an engagement with the direct objects of perceptual experiences, with what is present.

> "The relevant phenomenological 'looks' phenomena flow directly from the core early modern empiricist insight at the heart of (OV) [the Object View]" (HAI, p.19).

Whereas, for Brewer, the Content View is not even appropriate to account for the phenomenology of perception.

8

In getting rid of the representationalist response to the Argument from Illusion, the Object View not only recovers the Berkeleyian's insight, that the direct objects of perception just *are* the physical objects of the world, as it adds a realist account of their nature, through its account for illusions. According to Brewer, there is no need to appeal either to mental entities or representational content in order to explain those phenomena that we consider to be delusive in perceptual experiences. What is needed is to add a third element to that already offered in EE. Perceptual experiences, including illusions, are to be explained considering not only the relation between their subject and their objects, but also the spatio-temporal situation in which it happens. Thus, an illusion is a manifestation of a three part relation, common to all perceptual experiences. Illusions are perceptual experiences in which a subject perceives an object as instantiating a property that belongs to a certain paradigm. In my view, a complete rehabilitation of the Berkeleyian insight, together with the preservation of EE, is achieved only if Brewer, and the proponents of the Object View, explicitly admit some sort of *illusions eliminativism*. To eliminate illusions from a philosophical account of perception is the step to eschew from it any account that depends on the intervention of any kind of representation, whether some kind of mental entity or some kind of content. This may be a radical outcome but so it was ILL. The question, then, is to consider the implications of such a radical account for the epistemological role of perception.

REFERENCES

Brewer, Bill.

Perception and Reason. Oxford: OUP, 1999.

Perception and its Objects. Oxford: OUP, 2011.

Précis of *Perception and Reason*, and response to commentators (Naomi Eilan, Richard Fumerton, Susan Hurley, and Michael Martin) for a book symposium in Philosophy and Phenomenological Research, 2001.

Précis of *Perception and Reason*, and response to commentator (Michael Ayers) for a book symposium in Philosophical Books, 2001.

"Do Sense Experiential States Have Conceptual Content?". In M. Steup and E. Sosa (eds.), Contemporary Debates in Epistemology. Oxford: Blackwell, 2005.

"Perception and Content". *European Journal of Philosophy*, 14 (2006), pp. 165-81.

"Perception and its Objects". *Philosophical Studies*, 132 (1), 2007, pp. 87-97.

"How To Account for Illusion". In F. Macpherson and A Haddock (eds.), *Disjunctivism*. Oxford: Oxford University Press, 2008.

Byrne, Alexander.

"Experience and Content". In K. Hawley and F. Macpherson (eds.), *The Admissable Contents of Perceptual Experience*. Wiley-Blackwell, 2011.

Frege, Gottlob.

"Thought". In M. Beaney (ed.), *The Frege Reader*. Blackwell Publishers Ltd., 1997.

Travis, Charles.

"The Silence of the Senses". *Mind*, 113, 2004, pp. 57-95.

DISJUNCTIVISM AND THE PUZZLE OF PHENOMENAL CHARACTER

Roberta Locatelli

The present paper stems from some trouble I have been having in understanding the commitments of what is often called phenomenal disjunctivism. The worries are connected with what seems to me to be a tension between the central role that the notion of phenomenal character plays in this version of disjunctivism and what I take to be the primary aim of any disjunctive account of perceptual experience, namely the rebuttal of the internalist view according to which only the inner, subjective features of experience play a role in the identification of perceptual states, while the features of the surrounding world are just accidental, causal, conditions, which don't really affect the nature of perceptual experiences.[1]

[1] See Snowdon (1980, 1990). It is quite common to see this view labelled 'Cartesian view of experience' (See Child 1991, 1994) or, less often, 'experiential monism' (see Kalderon, Travis, manuscript).
It is noteworthy that, as Martin (2006) points out, the label 'disjunctivism' was first introduced by an opponent of the view, Howard Robison (1985: 174; 1994 : 152) who calls 'disjunctive theory' a style of reaction to the argument from hallucination. Hinton (1967a, 1967b, 1973), who is shown to be the initiator of disjunctivism, does not use this term, but merely speaks of 'perception/illusion disjunction', exemplified by sentences such as 'I see a flash of light of a certain sort or I am having the perfect illusion of seeing one of that sort'.
Broadly speaking, being acquainted with mind-independent objects means being in an irreducible and direct contact with them. Fish, for instance, says that the term 'acquaintance' signifies an irreducible mental relation in which the subject can only stand in to objects that exist and features that are instantiated in the section of the environment at which the subject is looking. (Fish 2009, p. 14). I am not inclined to assign any substantial significance to this terminological choice, because in this context it seems to be just one attempt (among others) to convey the immediacy of the perceptual relation under the naïve realist conception.
Cf. Martin (2004, p. 273). This view is also called the 'highest common factor view' by McDowell (1996, 113): there must be something in common between cases of veridical and non-veridical experience, given that some veridical experiences are phenomenologically indistinguishable from some non-veridical experiences. Hinton calls this idea 'the doctrine of experience' where 'experience' is a 'special, philosophical notion' which is thought of as being 'a sort of bonus in addition to everything that happens physically' (Hinton, 1973, p. 11) which is countered by an '"internal" description of a sensory [...] experience' (p. 12). In a more linguistic characterisation, he also calls the view 'phenomenological specimenism', which is described as the view that some neutral experience-reports which are precise or exact enough to satisfy the following condition: a) the report states the occurrence of a specimen event which is one that could have occurred to the subject 'if

The aim of this paper is twofold. Firstly, I will try to elucidate both the significance and the motivations of phenomenal disjunctivism, as it is presented by Martin. In particular, I will try to make sense of the idea, which has puzzled many commentators as possibly incoherent, that two experiences, one veridical, the other hallucinatory, might be indistinguishable, although having different phenomenal characters.

Secondly, I will suggest that some doubts still remain beyond such an elucidation, and I will conclude that if disjuctivism aims to challenge internalism, it should avoid using the notion of phenomenal character or introspection.

The last part of this paper is aimed at understanding why, despite the problems that such an approach generates, one might be willing to state the difference between perception and hallucination in terms of a difference in their phenomenal character. I will conclude this examination by suggesting

the case had been whichever it in fact was not: illusion if in fact it was perception, perception if in fact it was illusion'. b): if one makes the report at the occurrence of a perception, the very same report could be made for an illusion the subject might have later and being 'qualitatively' indistinguishable from the first (Hinton 1980, 37).

A terminological precision is useful here. In accordance with the literature, I will use 'experience' with reference to any sensory experience, irrespective of whether the experience is illusory or successful, while I restrict 'perception' to cases in which the perception verbs are used as 'success' verbs (See Searle 1983, p. 194).

It might be seen itself as a variation on the most fundamental argument from illusion: as the latter, it relies on two crucial premises; the phenomenal principle and the generalising principle, which respectively allow for the first and second steps of the argument (See Robinson 1994, p. 87).

This example comes from Austin (1962, p. 50) and has been widely echoed by many disjunctivists. However there is a slight difference between Austin's original use and the use made of it by recent disjunctivists. Austin's target was the idea that the same kind of objects (namely sense data) must be perceived in both cases of perception and illusion (or hallucination). More recent disjunctivists's targets are less the commonality across objects of perception than the commonality across experiences themselves (Cf. Martin 2006, Travis 2004).

Martin never used this label, but it has become quite common to call his proposal by this name, or a similar one, like 'disjunjunctivism about phenomenal character'. See, for instance, Soteriu (2009), Conduct (2010), Dorsch (2011), who all follow Macpherson and Haddock's (2008) distinction between epistemological, experiential and phenomenal disjunctivism.

It is worth noting that this ontological understanding of disjuctivism has arisen only in the discussion of the last decade, while the claim that that early disjuctivists, such as Snowdon and McDowell (not to mention Hinton) are committed to any ontological claim about the nature of perception and hallucination respectively is more dubious. In McDowell's view, the difference between perception and illusion depends solely on the respective epistemic reliability. What is more controversial is whether McDowell's disjunctivism is exclusively epistemological (See Macpherson and Haddock (2008), Byrne and Logue (2009), Pritchard (2008), Thau (2004), Snowdon (2004) and many others) or if his epistemological view commits him to a disjunctivism about the nature of experience (See Snowdon 2009). The divergence may rely on the fact that Snowdon (2009) considers only McDowell (2008), which present several differences from McDowell's early position.

As for Snowdon, it is not clear how far his view is ontologically committed. However, Snowdon stresses 'how limited [...] is the commitment to the disjunctive theory' (1990) and that his main aim (1980, 1990) is to argue against the idea that 'the concept [...] of seeing is a causal concept with a separable experience required as the effect end' (Snowdon 1990, p. 61). I don't see in Snowdon (1980, 1990) any reason to think that this conceptual claim requires ontological commitments.

This is often true even of those who do not like to talk about 'qualia' and don't subscribe to the idea that qualia have an autonomous existence.

that the reasons for adopting this approach derive from an understanding of the scope and the aims of naïve realism which is far from being obvious.

1. CHALLENGING THE COMMON KIND VIEW

It is somehow misleading to consider disjunctivism a theory of experience[2]: rather, disjunctivism is a variety of views sharing a common polemical target and a very broad aim.

The common aim is the defence of naïve realism, the view according to which perceiving is being acquainted with mind-independent objects.[3]

The shared negative thesis is the rejection of the common kind assumption (from here on CKA)[4], namely, the idea that experiences form a common kind of mental state across veridical perceptions, illusions and hallucinations.[5] The two points are connected, for the common kind view seems incompatible with naïve realism: if perception and hallucination are fundamentally the same kind of mental event, it becomes problematic to claim that perception is an acquaintance with mind-independent objects for in hallucination there is no proper mind-independent object one can be related to. In fact, the famous argument from hallucination against naïve realism hinges on the CKA.

The argument from hallucination comes in a number of different formulations[6], but we can roughly identify two steps in which it is

[2] See, for instance, Block (1995, 2002), Byrne (2003, 2004), Carruthers (2000), Crane (2001) and Farkas (2008). In this regard, the first line of Fish (2008) is strikingly eloquent: 'Our *datum* then is that some mental states that are not veridical perceptions – hallucinations – can nonetheless be indistinguishable from veridical perceptions'. (p. 144; my italics).
Martin says 'matching hallucinations'. Hereafter I will more often omit this precision. I don't think this abbreviation will affect an understanding of the main ideas discussed in this section. In fact, the restriction to hallucinations which match veridical perceptions is not relevant at this stage. We will see later on that it might pose certain problems for the disjunctivist. Moreover, the precision regarding 'matching hallucination', instead of dispelling possible misunderstandings, introduces an element of ambiguity, as it seems to conflate two different conditions: that of being realised by the same proximal causes as a perception and that of being subjectively indistinguishable from a perception.

[3] See Martin 2004, p. 69.
See for instance Farkas (2006), Smith (2008), Siegel (2008) and Conduct (2010).
See also Martin (2002) where he describes the debate about intentionalism and the sense data accounts as a 'debate about appearances, about how things seem to one', (p. 376) and proposes that naïve realism is a better account of a phenomenal feature of experience: due to its transparency.

[4] McDowell doesn't generally use, to my knowledge, the locution 'naïve realism' in his writings. However, I think one can fairly apply this label to his view. Moreover, in a recent public lecture (the Inaugural Dorothy Edgington Lectures at Birbeck College, University of London), McDowell, in response to a criticism from Charles Travis, claimed that he sees his own view as a naïve realist position.

[5] I take it to be a combination of epistemological and ontological claims because in this account the ability to entertain demonstrative thought is made possible by perception in virtue of its putting the subject in a particular relation to the object.

[6] See Heather Logue (2010, 22): 'One might find this idea difficult to get one's head around. What exactly does it mean to say that the phenomenal character of my current veridical experience , that is "what it is like" for me to have it, is *constituted* by a *banana*?'. Here the concern is to do with the ontological discrepancy between

articulated. The first step stems from the statement that experiences are sometimes hallucinatory, and concludes that, at least in these deceptive cases, naïve realism cannot be true. The second step generalises this conclusion to cover all cases of perceptual experience on the basis of the CKA, using the idea that hallucinations and veridical perceptions might be indistinguishable, and so they are fundamentally identical and therefore both require the same account. Hence naïve realism cannot be true in the case of veridical perception either.

Disjunctivism contends that the argument from hallucination is fallacious because the common kind assumption is unmotivated, as indistinguishability cannot justify identity. Even if a perception and a hallucination might be introspectively indistinguishable, they do not need to share any essential core, or be identical in any fundamental way.

The fact that the incapacity of a subject to tell one thing apart from another does not assure their identity is generally not challenged: as far as I can see, no philosopher would be willing to infer, from the inability of someone to tell apart a lemon shaped soap from a real lemon, that the lemon and the soap must be ontologically the same.[7] When it comes to perceptual experiences, the reason why,most of them accept the inference from indistinguishability to identity is because of their acceptance of a fundamental assumption about experience and, often, mental states in general: the idea that experiences are identified in an internalist way, that is to say, only by reference to what is introspectively accessible, whilst the relation to an external object is just an incidental additional condition. Therefore, the ultimate target of disjunctivism seems to be this internalist approach to experience. In contrast, according to disjunctivism, mind-independent objects are not fortuitous additions to experiences, but are fully constitutive of them.

2. INDISCRIMINABILITY WITHOUT SAMENESS OF PHENOMENAL CHARACTER

If it is relatively easy to spell out what the target of disjunctivism is, suggestions about it's positive commitments have been more controversial.

phenomenal character, a subjective property of a mental state, and the physical objects in the external world. I cannot linger on Fish's proposal here, however I would like to draw attention to the fact that, unlike Martin, he provides an explicit definition of 'phenomenal character', which, for the above-mentioned reasons, is hard to conciliate with phenomenal disjunctivism.

[7] On this distinction between 'internal' and 'external' typing, cf. Snowdon 2008, p. 39.
In addition, the phenomenological tradition, which has widely informed the current use of the notion of 'phenomenal', identifies the phenomenal consciousness with a methodological suspension of any ontological commitment about what appears to one; the well known phenomenological epoché.

In fact, it is because of its fundamentally negative nature that disjunctivism comes in many varieties. It is now time for us to appreciate the peculiarities of phenomenal disjunctivism[8], the focus of this paper.

In more recent discussions, disjunctivism has often been understood as a claim about the ontology of the mind: it provides a new taxonomy according to which we shall count two (hallucination and perception as distinctive mental kinds) where we used to count one (experience as a common kind)[9].

There are different ways in which this distinction can be drawn. A very influential way of stating the distinction between perception and hallucination has been in terms of a difference in their phenomenal characters: veridical perception has a distinctive phenomenal character which cannot be shared by any hallucinations. This idea can be found stated in the work of Langsam (1997), Martin (1997, 1980, 2004, 2006) and Fish (2008, 2009). Martin, for instance, claims, that naive realists must deny 'that two experiences, one of which is indiscriminable from another, must share the same phenomenal character' (Martin 2006, 367) and assures us that 'the phenomenal characters of two experiences can be different even while one of them is indiscriminable from the other' (Martin 2006, 14).

If we are to understand this claim, we ought first to see what 'phenomenal character' means. The problem with this notion is that it is very widespread but never quite explained fully. It is a typical case of fashionable philosophical jargon: as widespread as it is ambiguous, its meaning slightly varies throughout the contexts in which it appears. It is often used as a synonym for *qualia* (generally the more innocuous variant), and should grasp the 'what-it-is-like' aspect of experience, its qualitative tone. Fish, for instance, defines phenomenal character as 'the property of the experience that types the experience by what it is like to have it' (Fish 2009, p. 16).

Consequently, it suffers from all the ambiguities of the 'qualia' and 'what-it-is-like' jargon.[10] However, at least three ideas about the notion of

[8] I use the notion of sensory exploration with reference to Mohan Matthen (2012), which discuss sensory exploration as a procedure apt to eliminate grounds for doubts about the correctness of one's experience and a distinguish between empirical doubts, which are dispelled through such a procedure, and sceptical doubts, which are impermeable to any kind of inquiry and exploration.

[9] I cannot expand on this here. I refer the reader to Snowdon (2010). Snowdon's targets are 'what-it-is-like-ness' expressions, but as phenomenal character is generally explained in terms of what-it-is-likeness, most of Snowdon's analysis can be extended to the notion of phenomenal character itself.
Here is an important methodological point which should be treated separately. Are hypothetical cases of experience pertinent (more pertinent than actual cases) to a theory of perception? Disjunctivists struggle to conciliate these sceptical scenarios with naïve realism. I do not think this is possible, but this is not even required. We can dismiss them as irrelevant.

[10] Snowdon pointed out to me that in certain cases the notion of 'experience' might be primitive and prior

phenomenal character are fairly consensual: 1) it identifies some properties of the experience itself, not of the objects experienced; 2) it refers to something subjective, accessible only from the subject's point of view; 3) it is used to refer to the qualitative, sensory aspects of the experience.

Given this definition of phenomenal character, the advocates of phenomenal disjunctivism have a problem reconciling their main claim with the fact that a hallucination might be subjectively indistinguishable from a perception (which is, as a general rule, accepted to be uncontroversial).

If two experiences are subjectively indistinguishable, then they look subjectively the same. This means that they must share the same phenomenal character, for the phenomenal character tells us how an experience appears or looks, subjectively, to an individual. So, in denying the sameness of phenomenal character while admitting that a hallucination might be indistinguishable from a veridical perception, the phenomenal disjunctivist seems committed to a contradiction.

3. THE NEGATIVE CONCEPTION OF HALLUCINATION AND THE AUTONOMY OF THE PHENOMENAL LEVEL OF EXPERIENCE UNDER QUESTION

In Martin this contradiction is avoided by combining phenomenal disjunctivism with a negative, epistemic view of hallucination, according to which the phenomenal character of hallucinations is constituted purely by its being indiscriminable from a veridical perception. This is stated by Martin as follows:

> 'For certain visual experiences as of a white picket fence, namely causally matching hallucinations, there is no more to the phenomenal character of such experiences than that of being indiscriminable from corresponding visual perceptions of a white picket fence as what it is' (Martin, 2006, p. 369)
> 'When we turn to a case of perfect hallucination, we know that the Naïve phenomenal properties which seem to be present in the case of veridical perception certainly cannot be present in the case of hallucination. Of course they may still seem to be

to the notion of perception. This is, for instance, the case of pain, which one experiences without properly perceiving it. I grant this, in fact the experience of pain doesn't allow for a distinction between correctness and incorrectness: experiencing a pain is sufficient for the pain being there. However, when talking about 'experience' I refer here to 'perceptual experience', that is to say purported experience of objects and facts about the external world, the notion that one might be tempted to consider as a general and comprehensive case covering veridical perception as well as illusions and hallucinations.

> present, and in as much as the hallucination is indistinguishable from the perception they will seem to be so.' (Martin 2004, p. 49)

Perception has a peculiar phenomenal character (composed by what Martin calls 'naïve phenomenal properties') which is partly constituted and determined by some mind-independent objects (See Martin 1997, p. 93). Hallucinations of course cannot have this phenomenal character, because there is no mind-independent object that might constitute the phenomenal character of one's hallucinatory experience, since *ex hypothesis* in hallucinations 'no appropriate candidate for awareness existed' (Martin 2004, p. 39). But this does not mean that the phenomenal character of hallucination must be constituted or determined by something different, or that we might identify other phenomenal properties, alternative to the naïve ones. The entire phenomenal consciousness of hallucinations is provided by its being 'essentially failure - they purport to relate us to the world while failing to do so' (Martin 2006, p. 372). And this is the most specific thing one can say about hallucinations:

> 'There are certain mental events [causally matching hallucinations] whose only positive mental characteristics are negative epistemological ones - that they cannot be told apart by the subject from veridical perceptions' (Martin 2004, p. 73-4; my stress).

Prima facie, this might seem odd. How can a mental state be defined only through a negative criterion? And, more importantly, how can the phenomenal character of a hallucination be determined by an epistemic condition, such as being indiscriminable from another experience? This seems to conflate two distinctive aspects of mental life, the epistemic and the phenomenal levels, which one wants to keep separate, for the latter was introduced precisely to isolate the qualitative, felt components of experience from the reflective, epistemic components (beliefs that one forms on the basis of experience or that one previously possessed, and that might influence the nature of experience and any other epistemic attitude connected with perception).

Many commentators (most famously Smith 2002, 2008) have objected that the epistemic conception is enough to highlight only the cognitive content of hallucination, but it is completely useless for explaining its sensory, felt character, that is to say its phenomenal character. And as the phenomenal character is meant to be the sensory aspect of an experience,

this seems to imply that there is nothing that having a hallucination is like. Without phenomenal character, a state can hardly be a genuine experience. The negative view of hallucination would then end up claiming that in cases of hallucination we are no better off than philosophical zombies who 'just satisfy a functionalist definition for being in a mental state but lack any phenomenal consciousness' (Martin 2006, p. 378).

The apparent oddity of this view decreases if one considers an aspect which is often neglected: that the negative view of hallucination is first of all a negative view of experience. According to Martin, the very notion of experience itself has only a negative characterisation and is defined only by reference to perception: 'Being indiscriminable from veridical perception', Martin writes, 'is the most inclusive conception we have of what experience is' (Martin 2004, 52) – 'indiscriminability provides sufficient conditions for an event's being a sensory experience' (Martin 2004, 74).

Having an experience, according to Martin, is being in a state which is subjectively indiscriminable from a state in which mind-independent objects are made manifest. This feature of indiscriminability can apply to perception (this is tautological, as a perception is always indiscriminable from itself) and to hallucination (which is defined only by reference to perception). However, even if indiscriminability from perception is a common feature across perception and hallucination, it is not the most fundamental feature of both cases: it is the fundamental characteristic of hallucination, but it is not so of perception, whose must fundamental characteristic is having certain naïve phenomenal properties.

The negative view of experience does not deny that there is anything like a conscious experience. It simply means that experience (as a common kind) 'lacks explanatory autonomy from that of veridical perception' (Martin 2004, p. 73). Sensory consciousness is not a matter of having some inner properties of the experience, and some appearances: it is, in Martin's words, a matter of having a point of view on the world. And this is so even in the case of hallucinations. He says.

> 'The negative epistemological condition when correctly interpreted will specify not a subject's cognitive response to their circumstances – and hence their knowledge or ignorance of how things are with them – but rather their perspective on the world. This is sufficient for it to be true of a subject that there is something it is like for them to be so. ' (Martin 2006, p. 376).

If one considers the negative view of experience (and, hence, of hallucinations) this way, it turns out to not be a tortuous exit strategy aimed at avoiding the apparent inconsistency between the indistinguishability of perception and hallucination, and the difference in their phenomenal character, as it might have seemed at first glance. The negative view of experience appears then to be one and the same with the general thesis of disjunctivism, that is to say, the idea that we do not need to look for a mental event which specifies what one is undergoing, whatever the world is really like. In other words, denying any explanatory role to the notion of experience is no more than a way to specify the anti-internalism propounded by disjunctivists that I gestured to before. The negative view of experience amounts to the rebuttal of the substantial understanding of phenomenal character implied by the internalist view: something inner (a property of the experience itself), something purely experiential, something that you grasp by 'looking' inwards.

So, in a sense, it is true that the epistemic conception of experience isn't able to specify the phenomenal character of an experience, if by phenomenal character one means a level of mere appearances, a flow of inner occurrences, which are prior to and independent from any uptake of it. But this conception of phenomenal character is precisely a myth, which presupposes an observational model of self-awareness. The observational model conceives self-awareness as a twofold process, upheld in a phenomenal level of consciousness, and a higher-order monitoring process which picks up the phenomenal level of consciousness (which, in turn, is prior to any cognitive access to it). Only if one thinks that the phenomenal level and the cognitive consciousness of it are two sharply distinct moments of consciousness can one complain that the negative view of hallucination provides only its cognitive aspect and not its sensory one.

For Martin this picture of introspection is misleading: in experience, there is no phenomenal level prior to and independent of its cognitive uptake, 'rather they must coincide' (Martin 2006, p. 389).

Thinking that an experience is something that can be defined in a more direct way (through the appeal to phenomenal character) is a *petition principii* which presupposes CKA (and the internalist perspective it presupposes) which should instead be established.

The epistemic conception of hallucination is in reality a criticism of a certain way of understanding phenomenal character as a substantive mental occurrence or feature, as an inner appearance.

4. PHENOMENAL NAÏVE REALISM

So far, so good. If my reading is correct, the negative view of hallucination (and experience), on which most of the criticism against phenomenal disjunctivism has focused, doesn't really pose any major problem to the possibility of taking hallucinations as cases of genuine sensory consciousness. On the contrary, it further clarifies the significance of naïve realism and gives some important and promising suggestions for the understanding of self-awareness.

What still requires elucidation is the claim that perceptions and matching hallucinations have different phenomenal characters. If the phenomenal character is how experience seems, subjectively, why should we claim that a perception and a hallucination which look the same have different phenomenal characters?

As neither the relation of constitution nor the relation of determination are relations of identity, for two things being constituted or determined by different things does not imply that they are fundamentally different, so two phenomenal characters that are constituted or realised by different conditions might still be the same.

Conceding that a perception and a matching hallucination might have the same phenomenal character does not imply that they must be fundamentally the same: they might still have different natures, namely, and respectively, the positive nature of being an occurrence of physical objects made perceptually manifest to one, and the negative nature of being indiscriminable from experiences that put us in a relation with physical objects.

However, for Martin, the acceptance of any identity between the phenomenal characters of a perception and a hallucination'would be inconsistent with the aim of the naïve realist' (1997, p. 97). The reason why naïve realism requires that the difference between perception and matching hallucination be drawn on the phenomenal level is that Martin understands naïve realism as an account of phenomenal consciousness. He says:

> '[Naïve realism] seeks to give an account of phenomenal consciousness, and hence the disjunctive account is intended to have a direct bearing on one's account of what it is like for the subject to be perceiving.' (Martin 1997, p. 97)

This is a peculiarity of Martin's understanding of naïve realism, which is most often not defined in phenomenological terms. Most often naïve realism is viewed as an ontological or an epistemic claim, or a combination of the two.

An instance of naïve realism formulated as an ontological claim can be found in Logue (2011): 'Naïve Realism [...] holds that veridical perceptual experiences fundamentally consist in the subject perceiving physical entities in her environment.'

McDowell states naïve realism in epistemological terms, as 'the idea of environmental facts making themselves available to us in perception'. (McDowell 2008, p. 380).

An example of naïve realism as a combination of epistemic and ontological claims can be found in Snowdon (2005):

> 'If an experience E is a genuine perception by subject S of object O then the occurrence of E places S in such a relation to O that were S able to entertain demonstrative thoughts (and was equipped with the necessary concepts) then S could entertain the *true* demonstrative thought 'that is O". (Snowdon 2005, p. 138).

Martin provides a significantly different definition of naïve realism. He claims:

> 'According to naïve realism, the actual objects of perception [...] partly constitute one's conscious experience, and hence determine the phenomenal character of one's experience' (Martin 2004, p. 93).

And he recommends that 'This talk of constitution and determination should be taken literally' (Ibid.). Obviously, as he notes himself later on in the same sentence, 'a consequence of it is that one could not be having the very experience one has, were the objects perceived not to exist' (ibid). When naïve realism is stated in phenomenological terms, as a claim about the nature of phenomenal character of veridical perception, one is committed to phenomenal disjunctivism. But is this formulation of naïve realism either required or justified?

Many commentators have wondered how this naïve realist claim should be interpreted, and what it might mean. How can something mental (the phenomenal character of a perception, that is: a property of a mental state) literally be constituted and determined by external objects?

Sure, we can give this formulation a charitable interpretation, and understand it as claiming that perceiving is not being aware of inner appearances or features of experience itself, rather it is being aware of some aspects of objects in the world. This is certainly part of what Martin has

in mind and I could not agree more with this. However, the awkwardness of the expression (with all the difficulties in grasping its proper sense) remains, and one might ask why naïve realism should be stated in terms of what determines and constitutes the phenomenal character of an experience, what it is like to have it. Why should the notion of phenomenal character be used at all here?

5. WHAT PHENOMENAL CHARACTER?

I will suggest that one is led to think that naïve realism has to be stated as a view about phenomenal character (and hence implies phenomenal disjunctivism) if one understands naïve realism to be connected in a certain way with the sceptical concern. However, before any attempt to diagnose the reasons that might motivate this particular way of understanding naïve realism (which, we have seen, is neither universally shared nor compulsory), it is important that we understand what 'phenomenal character' means in this context.

It is clear, in fact, that, unless we admit that phenomenal disjunctivism overtly contradicts itself, we must grant that phenomenal character is not used by Martin along the same lines as the mainstream use of the term. This is fair enough, as 'phenomenal character' is a term of art and hence its significance can be fixed with relative freedom, provided one also explains how the term is being used. This is all the more important in this case, where, as we have seen, 'phenomenal character' appears to be used in a very flexible and ambiguous way.

The problem is that the significance of 'phenomenal character' is even less clear in the writings of phenomenal disjunctivists than it is in the mainstream use of the term. In particular, Martin never defines what he means by 'phenomenal character'. This lack of explicit definition seems to be due to a purported self-evidence of the locution: he relies on the existing literature to seize the scope of the notion he uses. So, on the one hand, the use of the notion of phenomenal character by phenomenal disjunctivists alludes (at least partially) to its standard use, as 'how experiences strike to us as being to introspective reflection in them' (Martin, 2004, p. 42), or, as Fish says, 'the property of the experience that types the experience by what it is like to have it' (Fish, 2009, p. 16).

On the other hand, this definition appears to be insufficient, for 'how experiences strike to us as being to introspective reflection in them' (Martin, 2004, p. 42) is *ex hypothesis* the same for perception and hallucination. Yet phenomenal disjunctivism claims that hallucination and perception cannot share any phenomenal character, even when they are indistinguishable. So

the only way to spell out what phenomenal character means here seems to be the following: phenomenal character is what an experience seems to be to a subject *through introspection*, with the proviso that a certain kind of phenomenal character exists only if it is partly constituted and determined by surrounding objects, meaning that only veridical perceptions can have a phenomenal character (of that kind). Only with this proviso can one make sense of the idea that perception and hallucination cannot share a phenomenal character.

It is not clear whether, in this account, the fact of being (or not being) constituted and determined by some physical object makes any difference in the *phenomenology* of the experience, in the way it phenomenally strikes the subject. Once again, a positive answer would be at odds with the starting hypothesis about indiscriminability. If the answer is negative, we are left with an incertitude regarding the significance of phenomenal character. If phenomenal character is the way experience seems to one through introspection, it is not clear why the difference in the way it is realised should count, especially if the difference is not manifest to the subject. It would be like saying that the visual features of a lemon and of a perfectly crafted lemon-shaped soap are different because they are determined by two things which are ontologically different. Of course, seeing a lemon and seeing a soap are two different things. But in order to maintain that we do not need to deny that their visible appearances are the same.

It seems to me that saying that the phenomenal character of an experience is what it is like to have that experience *and* at the same time it is something that requires *constitutively* the presence of some objects results in a sort of conceptual monster, which tries to bring together two incompatible ideas. On one hand, the notion of phenomenal character is suited to type experiences on the basis of what is accessible through introspection alone, through an inner observation. On the other hand, the reference to objects seems to convey the idea of an external principle for classifying experiences; a principle that considers the external objects that one sees. The formulation of naïve realism provided by Martin seems to aim to combine these two principles of typing by suggesting that the phenomenal character, which is merely inner and subjective, can have, in itself, an ontological commitment.

I am not sure that the notion of phenomenal character is suited to support any ontological commitment at all. The notion of phenomenal character, or of phenomenal consciousness is the result of the gesture of 'withdraw[ing] my thoughts from every thing external', as Thomas Reid said. And one can hardly reintroduce an ontological commitment to the world in a notion that, by definition, withdraws it. To use a Hintonian

expression, phenomenal character seems to have been introduced to 'answer to the question as to what is happening to the subject' (Hinton, 1973) in the 'most precise way', and the question is posed in a way such that the fitting answer should exclude any consideration of what is or is happening 'outside' the subject.

The aim of phenomenal disjunctivists is, in Martin's words, to 'preserve the little knowledge that we could have through reflection on our experience' (Martin 2006, p. 57): starting from what we know through introspection alone, we must be able to acknowledge that we can still reach some knowledge about the external world (See Martin 2006, 57).

But by aiming to show that what we can know though introspection alone already contains all we need to be assured of the existence of an external world, one has already conceded the first step of a Cartesian-like sceptical reasoning. That is to say, the idea that the capacity which experience has, of putting us in contact with the world, should be evaluated solely on the basis of what one can tell through introspection alone.

Once the problem of perceptual access is arranged in this way, hallucinations start to become a problem for the idea of a perceptual contact with the world. Hallucinations risk undermining the idea of an appropriate relation to the world that perception should provide. At this stage, the only solution for avoiding the conclusion of the veil of perception would be to put hallucinations under *quarantine*, in some compartment of mental taxonomy other than perception: they should be different from veridical perception. But, at this stage, what makes them different must be something intrinsic to one's mental, inner life; it cannot be just the fact that in some cases experience is acquainted with features of the world, and in other, much rarer cases, experience is, *de facto,* not experiencing anything. The difference cannot ultimately lie anywhere other than in the phenomenal character itself.

I do not think this strategy is successful, and not because the sceptic is right, but because this line of thought, from the beginning, concedes too much to the sceptic, and once the first step has been taken, once one has decided to confine oneself to the phenomenal in order to evaluate the ontological scope of experience, one can hardly reconstitute the mind-independent world, which is, *ex hypothesis*, cast out.

Experience 'contains' all that we need to be assured of being in cognitive contact with the world. But this is the case not because all we need can be found within the phenomenal character itself, in what is accessible through introspection alone. Rather, it is because perception is not the something we meet when we turn our attention inward. It is a complex process that allows us to perform procedures of verification, such

as sensory exploration through different sensory modalities, manipulation of objects or reflection on the coherency of series of experiences, each utilising testimony and previous knowledge.

If I am right in suggesting that the notion of phenomenal character is essentially internalist, perhaps those who aim to challenge internalism in the philosophy of perception should avoid using the term. I am not suggesting banning the use of locutions such as 'phenomenal character' or 'phenomenal experience' at all (however, it would be safe to use these expressions carefully, since they are liable to mislead). What I am recommending is avoiding attributing to phenomenal character any explanatory role in the identification of perceptual states. After all, Martin recommends that we don't attribute any explanatory role to experience. But to what does phenomenal character refer, if not precisely to what disjunctivists find wrong with the philosophical notion of experience, that is to say, in the possibility of grasping the properties of an experience by positing between brackets the existence of the world which the experience purports to present?

The problem is, hence, that phenomenal disjunctivists attribute to phenomenal character an explanatory, central role, as phenomenal character is at the core of both the definition of naïve realism (the thesis which has to be defended) and the strategy adopted to do that: the difference between perception and hallucination with respect to their phenomenal character.

6. INDISTINGUISHABILITY: A LESS THAN EVIDENT ASSUMPTION

Phenomenal disjunctivism, therefore, seems to latently accept the fundamental inner principle according to which only what is introspectible counts in typing, and has primacy over experiences. However, this view also wants to include the external way of classifying into the internal one, because the inner principle contains, in itself, the ontological commitment to the external world.

One might also see this principle operating in the strategy adopted by disjunctivists against the argument from hallucination. They address all their complaints to the identity principle, while accepting all the other assumptions, and foremost the indistinguishability of perception and hallucination. Instead, I think that we should be more prudent in accepting the indistinguishability claim.

First of all, it is not at all certain that hallucinations are like veridical perceptions: many cases of hallucinations are far from being indistinguishable from a perception in any possible sense. I have in mind,

for instance, hallucinations of impossible, Escher-like figures or the reports of psychiatric patients or subjects under the effect of psychotropic drugs, who often describe their hallucinations as being qualitatively different from ordinary perceptions (confused, blurred or incoherent, or – on the contrary – extremely vivid and bright).

Phenomenal disjunctivists seem less preoccupied with ordinary cases of hallucination than with the mere hypothetical possibility of a perfect hallucination induced by a *malin génie* or a mad scientist who makes us live in a world of mere appearances. But this is the typical counterfactual reasoning that structures the sceptical threat.

This view has a side effect. Although Martin and Fish mention that not all hallucinations must be distinguishable from a perception, they say surprisingly little about non-indistinguishable hallucinations and seem unable to account for these cases.

But there is a more profound concern with the idea of indistinguishability. What does indistinguishability mean when we use it in this sense?

Martin (and with him other disjunctivists) generally claim that indistinguishability of perception and hallucination is undeniable, as it is somehow contained in the definition of hallucination: a hallucination is indistinguishable from a veridical perception in the sense that a subject undergoing a hallucination might be unable to tell that he is not veridically perceiving. But saying that one might take a hallucination at face value – say, of an oasis – is one thing, while saying that hallucinating an oasis and perceiving one are two indistinguishable phenomena is quite different.

The notion of distinguishability (which is an epistemic notion) is ordinarily used for observable external objects, and it implies the possibility of comparing two objects viewed at the same time – or in succession. In any case, we have two terms to compare. I look at two objects before me (say, a lemon and a lemon-shaped soap) and I cannot tell them apart (for instance, I believe both are lemons), or, looking at the same basket on two different occasions, I might be unable to distinguish that the lemon seen the first time has been replaced by a lemon-shaped soap later. But, as Martin himself stresses, introspection does not work like the observation of external objects, and discrimination through introspection even less so.

When talking about perceptual experiences, the problem is that we have no idea about what must be brought into comparison. Imagine that I am now hallucinating a chimpanzee and I want to decide if my current hallucinatory experience is indistinguishable from a veridical one. With what am I going to compare it? Certainly not with another current experience, because I cannot hallucinate a chimpanzee and, at the same

time, veridically see one. But it is not even clear that I can compare my current experience with one I had before. I have never seen a chimpanzee before in a university classroom. When one talks about a hallucination that is indistinguishable from a perception, the situation the hallucination is compared to is a counterfactual one: an imaginary perception matching my current hallucination. This is problematic, because a comparison with a counterfactual object or state is necessarily unfair. The second object is purposefully shaped to match the first one. The very sense of the comparison is lost, because we know the result of the comparison before performing it.

I think that the transition from the simple fact that sometimes we take hallucinations for perceptions to the fact idea that the two are indistinguishable from perception comes from a mythological view of experience: an atomistic view of experience, which purports to analyse perception as a series of snapshots which exist in pure isolation from each another. I think that this view is an insidious myth of empiricism, and yet it is still very widespread. The myth has two faces. The first is the idea that any single judgement, considered in isolation, should be justified by an experience. Holism has generally god rid of this idea. However, the other side of the myth is still quite effective. It is the idea that an isolated particular, instantaneous experience (a snapshot, or, as it is often called in the literature, 'an atomic experience') should be enough to justify some judgements about the world. Of course, experience cannot do so: if one tries to artificially fracture the flux of experience and then wonders whether we have sufficient elements to tell if any snapshot provides the appropriate relation with the world, one will end up feeling that it does not, and will consider the possibility of hallucinations as a problem for naïve realism. But this is the case only because experience has been mutilated.

Perception does not work in snapshots. It is a complex process, a flux of information organised in a system of retention and anticipation, and in which we can operate sensory explorations. If we remember these important aspects of perception, we can see that experience contains in itself all the tools we need to know whether something is a hallucination or not, and that hallucination is no threat to the validity of perception.

PROSPECTS

If the sceptical problem of hallucination can be dispelled, we still ought to explain, of course, what happens when one hallucinates; how we can be victims of pathologies of experience. How to do this will be the subject of further inquiries. However, a promising track would be to take

the idea that hallucinations are pathologies of perception more seriously. This means, in one sense, that I recommend endorsing at least two central claims of disjunctivism: 1) the idea that perception is conceptually prior to the idea of experience, to which, in turn, one should not attribute any explanatory role. 2) the refusal to extend any conclusions about hallucination to perception, for what is true of the pathology should not necessary apply to the normal case.

On the other hand, precisely because hallucination is a pathology of perception, it doesn't seem very useful to consider perception and hallucination as two 'different mental kinds'. We can account for hallucinations as the result of an interaction, or a short-circuit of experience combined with imagination, beliefs, emotions and so on, which penetrate and shape the experience. We should stop thinking of mental life as being composed of individual watertight compartments. Consciousness is animated by a mutual penetration between what it has become fashionable to call 'natural mental kinds'. Investigating his mutual penetration might be of benefit to the study of the nature of hallucination and its relation with perception.

BIBLIOGRAPHICAL REFERENCES

Block, Ned. (1995). "On a Confusion About the Function of Consciousness." Behavioral and Brain Sciences 18: 227–47.

— (2002). "The Harder Problem of Consciousness." Journal of Philosophy 99 (8): 391–425.

Byrne, Alex. (2003). "Consciousness and Nonconceptual Content." Philosophical Studies 113 (3): 261–274.

— (2004). "What Phenomenal Consciousness Is Like." In Higher-Order Theories of Consciousness: An Anthology, ed. Rocco J Gennaro. John Benjamins.

Carruthers, Peter. 2000. Phenomenal Consciousness: A Naturalistic Theory. Cambridge University Press.

Crane, Tim. (2001). Elements of Mind: An Introduction to the Philosophy of Mind. Oxford University Press.

Farkas, Katalin. 2006. "Indiscriminability and the Sameness of Appearance." Proceedings of the Aristotelian Society 106 (2): 39–59.

— (2008). The Subject's Point of View. Oxford University Press.

Hinton, J. M. (1967). Experiences. Philosophical Quarterly 17 (66):1-13.

— (1967). Visual Experiences. Mind 76 (April):217-227.

— (1973). Experiences: An Inquiry Into Some Ambiguities. Oxford: Clarendon Press.
— (1980). "Phenomenological Specimenism." Analysis 40 (January): 37–41.

Austin, J. L. (1962). Sense and Sensibilia. Oxford University Press.

Byrne, Alex, and Heather Logue. (2009). "Introduction." In Disjunctivism: Contemporary Readings, ed. Alex Byrne and Heather Logue. Vol. 1. MIT Press.

Conduct, M. D. (2011). "Naïve Realism and Extreme Disjunctivism." Philosophical Explorations 13 (3): 201–221.

Dorsch, Fabian (forthcoming). "Experience and Introspection." In Hallucination, Cambridge (Mass.), ed. Platchias Dimitris and MacPherson, Fiona. MIT Press. www.exre.org/assets/files/dorsch/ei.pdf.
— (2011). "The Diversity of Disjunctivism." European Journal of Philosophy 19 (2): 304–314.

Haddock, Adrian. (2010). "What Is Disjunctivism?" Philosophy Now 81: 21–22.

Haddock, Adrian, and Fiona Macpherson. (2008). "Introduction: Varieties of Disjunctivism." In Disjunctivism: Perception, Action, Knowledge, ed. Adrian Haddock and Fiona Macpherson. Oxford University Press.

Hawthorne, John and Kovakovich, Karson. (2006). "Disjunctivism." Aristotelian Society Supplementary Volume 80 (1): 145–83.

Kalderon, Mark and Travis, Charles. (manuscript) «Oxford Realism», http://ucl.academia.edu/MarkEliKalderon/Papers/734291/Oxford_realism.

Logue, Heather. (2010). "Getting Acquainted with Naïve Realism: Critical Notice of Perception, Hallucination, and Illusion." Philosophical Books 51 (1): 22–38.
— 2011. "The Skeptic and the Naïve Realist." Philosophical Issues 21 (1): 268–288.

Michael G. F. Martin (1997). The Reality of Appearances. In M. Sainsbury (ed.), Thought and Ontology. Franco Angeli.
— (2002). The Transparency of Experience. Mind and Language 4 (4):376-425.
— (2003). Sensible Appearances. In T. Balwin (ed.), The Cambridge History of Philosophy. Cambridge University Press.
— (2004). The Limits of Self-Awareness. Philosophical Studies 120 (1-3):37-89.
— (2006). On Being Alienated. In Tamar S. Gendler & John Hawthorne (eds.), Perceptual Experience. Oxford University Press.

McDowell, John. (2008). "The Disjunctive Conception of Experience as Material for a Transcendental Argument." In Disjunctivism: Perception, Action, Knowledge, ed. Fiona Macpherson and Adrian Haddock, 25:19–33. Oxford University Press.

Matthen, Mohan. 2012. "How to Be Sure: Sensory Exploration and Empirical Certainty*." Philosophy and Phenomenological Research. LLC.

Pritchard, Duncan. (2008). "McDowellian Neo-mooreanism." In Disjunctivism: Perception, Action, Knowledge, ed. Fiona Macpherson and Adrian Haddock. Oxford University Press.

Snowdon, Paul F. (1980). Perception, Vision, and Causation. Proceedings of the Aristotelian Society 81:175-92.

— (1990). The Objects of Perceptual Experience. Proceedings of the Aristotelian Society 64:121-50.

— (2008). Hinton and the Origins of Disjunctivism. In Adrian Haddock & Fiona Macpherson (eds.), Disjunctivism: Perception, Action, Knowledge. Oxford University Press.

— (2005). "The Formulation of Disjunctivism: A Response to Fish." Proceedings of the Aristotelian Society 105 (1): 129–141.

— (2009). "McDowell on Skepticism, Disjunctivism, and Transcendental Arguments." Philosophical Topics 37 (1): 133–152.

Soteriou, Matthew. 2009. "The Disjunctive Theory of Perception." In Stanford Encyclopedia of Philosophy (Fall 2009 Edition), ed. Edward Zalta.

Thau, Michael. (2004). "What Is Disjunctivism?" Philosophical Studies 120 (1-3): 193–253.

Travis, Charles. (2004). "The Silence of the Senses." Mind 113 (449) (January 1): 57–94.

— (2005). "Frege, Father of Disjunctivism." Philosophical Topics 33 (1): 307–334.

[* Pagination refers to the reprint in the Byrne-Logue volume]

SELFHOOD AS GRAMMATICAL RESPONSIBILITY BETWEEN THE WILL-TO-UNDERSTAND AND THE WILL-TO-EXPLAIN

Paulo Jesus[1]

SUMMARY

Reasons and causes typify two language games or grammars tending to incommensurability. These grammars institute a qualitative discrimination between the self-efficacy of being someone and the symmetric, selfless, efficacy of being something. Reasons can behave as causes, although it is fully absurd to interpret reasons as causes and vice-versa. Yet, in order for reasons to possess causal efficacy, one must assume: first, a monist ontology warranting the communication of dynamic efficiency between two chains of phenomena, that is to say, the chain of intentional representations and that of body movements so that a deep ontological homogeneity may coexist with a surface heterogeneity; and, second, a self-alert phenomenology which recognizes the peculiar "I feel" that must be able to accompany "my acting" and "my making happen", while acknowledging, however, the validity of an invincible metaphysical uncertainty. Selfhood emerges here as an unstable but unifying process of meaning-construction.

[1] Research funded by a post-doctoral grant (FCT/MCTES).

REASONS AND CAUSES AS LIFE STRATEGIES

In line with Wittgenstein's seminars (Wittgenstein, 1958) the classical work by G. E. M. Anscombe (1957) illustrates vehemently the incommensurability thesis which implies the irreducibility and heterogeneity between reasons and causes. This thesis rejects the possibility of the identity of reasons and causes, considering it as alogical. For reason and cause would be incompossible logical functions, semiotic processes with non-coordinatable normative principles. Reasons and causes would be parallel "language games" without tangency points, without mutual translation, given their lack of analogical grammars. They would be absolutely heteroclite tools without any common criteria of truth, each of them having its peculiar cognitive efficacy. From this standpoint, the fundamental option for the grammar of "reason" or for the grammar of "cause" is not grounded directly in the ontology but rather in the cognitive and symbolic practices. The same phenomenon may be described either in terms of "underlying reasons" (becoming thereby an "action") or in terms of "efficient cause" (becoming then an "event"), because each one of these descriptions belongs to *sui generis* modes of interpreting a phenomenon (either as dependent or independent with regard to a self-conscious agent) and of relating it with specific types of practices (*desire* or *conative practice* which aims to produce something successful in the world, and *belief* or *cognitive practice* that seeks for representational accuracy). Such practices involve different evaluative canons (one being performative and the other properly descriptive) as well as "symmetric directions of fit" (one subordinates the world-to-the-word and the other the word-to-the-world) (Anscombe, 1957, p. 56; Searle, 1985). The production of selfhood or subjectivity appears as the key effect of a relative instability in the meaning of enacted signs. The cogency of Wittgenstein-Anscombe's proposal imposes itself by sacrificing entirely the ontology of action—or rather by abstracting from it with an attitude of ontological agnosticism (Wittgenstein, 1953, p. 195). In fact, under this angle, the concept of *meaning* is absolutely neutral or indifferent towards ontology; and, hence, follows tacitly the crucial inference that ontology, in and by itself, is amorphous and non-constraining. Every phenomenon may be "action" or "event", provided the necessary and sufficient obedience to the principle of semiotic non-contradiction, that is to say: "A phenomenon may be simultaneously or successively action and/or event if it is integrated in each one of both language games by different semiotic players in the same moment, or by the same player in different moments". By definition, a language game is efficacious only if the players follow and enact a *Gestalt* of rules that suspends all other possible rules, and thus creates a space of

inalienable symbolic sovereignty with onto-phenomenological effects. In this sense, the so-called pre-linguistic purity or indomitability of the "mode or process of being" is reduced to silence: indefinite, unintelligible and inefficient or inert silence. To evoke a striking Wittgensteinian example, one might say that according to this semiotic constructivism nothing *a priori* in a toothache determines its possible meaning, because there is no *universal ontology for a toothache* but only *contingent grammars* that rule the construal of a toothache as a particular object and quality of possible experience. One should assert, in the last analysis, that practical semiotics decides all meaning, including the meaning of senses, the meaning of sensoriality and all tonalities of *qualia*, like pain and pleasure. To suffer from pain in general and from a toothache in particular means to play a game that makes me play it. Something is "painful" only within a determined symbolic game that consists in the "institution of a painful meaning as sense" or a "meaningful system of nociperception". Grammar evolves through embodied co-constructed learning and along epigenetic paths that define the developmental compossibilities of meaning and its embodiment.

Likewise, the self emerges as a grammatical competence, mainly self-narrative, defining a field of intelligible action, which can be termed "moral personality" or "grammatical responsibility". A rigorous linguistic turn would present itself as a non-ontological constructivism, a semiosis that neglects the possibility of *onto- or bio- or eco-semiotic nerves* capable of being determinant endogenous forces on the epigenesis and continuous restructuring of grammars. A strict symbolic autonomy would have an autophagic tendency and would claim for an exception regime of self-determination and over-determination, assuming its primacy over the realms of phenomenology and ontology. The living grammar is, however, a performing art and produces by itself all possible onto-phenomenological constellations. Semiogenesis and ontogenesis must merge perfectly so that embodied signs may produce what they signify. Thus, efficacious *semiosis* unfolds itself as a unifying experience of *auto-poiesis* and *auto-energeia* (or *self-creation* and *self-actualization*).

THE GRAMMAR OF REASONS AND CAUSES: PATHOLOGIES AND THERAPIES

The hermeneutic oscillation between reasons and causes denote a metaphysical confusion which insinuates itself continuously into the relationships between subjects and verbs. It is legitimate to ask with Wittgenstein whether such confusion generates true philosophical

questions or mere grammatical pathologies whose therapy would consist in reestablishing functional affective bonds tying nouns/pronouns and verbs. All conceptual confusions, made manifest in the reciprocal ambiguity between reason and cause, would reside in *practical* confusions which are "grammatical" or "logical" performances leading to nonsense, confusions between games or between tools or between expressive symbols. The differential practice of the language games here at issue—"to give reason of an action" and "identify the cause of an effect"—is described by Wittgenstein as follows:

> Giving a reason for something one did or said means showing a *way* which leads to this action. In some cases it means telling the way which one has gone oneself; in others it means describing a way which leads there and is in accordance with certain accepted rules. […]
> At this point, however, another confusion sets in, that between reason and cause. One is led into this confusion by the ambiguous use of the word "why". Thus when the chain of reasons has come to an end and still the question "why?" is asked, one is inclined to give a cause instead of a reason. […]
> The double use of the word "why", asking for the cause and asking for the motive, together with the idea that we can know, and not only conjecture, our motives, gives rise to the confusion that a motive is a cause of which we are immediately aware, a cause 'seen from the inside', or a cause experienced. – Giving a reason is like giving a calculation by which you have arrived at a certain result.[2]

The learning of a grammar warrants the regulation and preservation of meaning. The core of any grammatical learning is not, however, strictly linguistic, but rather behavioral and practical. Even if a particular language has no words to say "I", "reason/intention" and "cause/force", it is likely to offer, in its many *life forms*, a repertory of production and comprehension codes that rule the functioning of bodily or symbolic expressions capable of accompanying and meaning a unique logic of agency. Obviously, it is possible that the interpreter of a non-linguistic expression makes a false inference relying on an over-interpretation which results in a "projection of intentionality" on a course of phenomena, whose chaining

[2] Wittgenstein, 1958, p. 14-15.

and sequencing was merely causal. The understanding of any expression requires the understanding of a life form. That is why the jocose formula—"if a lion could talk we could not understand him"—carries a deeper truth than first expected (Wittgenstein, 1953, p. 190). As is self-evident, such incomprehensibility proceeds not from the irrelevant impossibility of access to the lion's mental activity but from the fact that meaning resides in the logical texture of a life form that we, humans, *cannot* share entirely with lions. Therefore, a lion remains incomprehensible because of the logical idiosyncrasy of his life form. Understanding a life form is always a matter of degree, for it depends on the extent to which *my* life form shares the practical-logical processes of the *other's* life form. In a fundamental sense, the *life worlds* of different *life forms* must overlap in part; or else the simple recognition of a life form as *other's* differential mode of self-organization would be impossible. In *my* "grasping a native's point of view" (Malinowski, 1925), understanding a neurotic person (Jaspers, 1913), or guessing an animal behavior, there is always a part of shared and a part of non-shared processes of being and meaning, grounded in a partly common and partly unique *life world*; that which justifies the quest for an eco- and bio-semiotics. Whitehead's (1929) emphasis on creativity and concrescence, Husserl's (1970) concept of *Lebenswelt*, Merleau-Ponty's (1968) metaphor of a germinal "chiasm" in sensibility, and Jonas' (1966) idea of freedom or selfhood as intrinsic to every life form, all point to the birth of meaning in *pre-subjective wilderness*, as it were. By the same token, they all maintain that for any living being the reality of grammar meets the morphodynamics of life and encapsulates the "desire" for meaning and the "desire" for a selving self. Nature and Logos as well as life and grammar must be regarded as continually fusing and co-evolving processes—otherwise they vanish. In this sense, once embedded in an actual self-becoming process, reasons and causes give semiotic shape to life, and constitute life strategies, in which meaning and being converge into creative transitions or "actual occasions" (Stenner, 2008; Whitehead, 1929/1978, p. 211, *passim*). Grammatical constellations can be life strategies if they are instilled with autopoietic energy and if, therefore, produce themselves by producing what they signify. The touchstone of the actuality of a "grammatical" life strategy lies, however, in the self-transformational and self-transgressive force that converts "grammatology" into "experience", a field of emotional intensities, nexuses and vectors, that is, a future-centered organism in development.

LIFE, EXPRESSION, AND UNDERSTANDING: THE SEMIOTIC CYCLE

The surface syntactic privilege of the subject in most language games makes us hallucinate the omnipresence of "reasons" and postulate the primacy of the personal pronoun over the verb, as if the pronoun were the first force from which all language would follow and become a self-propelled stochastic process. Though pervasive as it may be, such symbolic primacy of the subject should not impede an alternative view on the order of dependence, namely an order centered on the prominent value of "action". This alternative order would entail the *syntactic declassification of the subject* and consider it as a simple "active verb complement" (Descombes, 2004; Tesnière, 1959). Thus, all syntactic relations might be reorganized by the category of action. Indeed, verbs are not ruled by subjects; verbs convoke subjects and these respond to them as complements. The conception of language under the perspective of a center of agent gravity proposes a more radical transformation, the transformation of the general interpretive semiotics in narrative semiotics. As narrative competence and performance, *my* subjectivity or agency is a *simple semiotic potentiality*, whose actualization depends on the development of a vital relationship with a concrete life form. The efficacious semiotics is bio-semiotics, accomplishing the preservation of a pattern of meaning which unifies a process and makes it recognizable as a life phenomenon for a living being. The narrative grammar produces the agent vitality of the narrative subject, whose essence lies in an autopoietic semiotic practice. It follows that the subject who expresses herself does not communicate any form of self-knowledge; she only actualizes a grammatical know-how. The essence of a narrative lies in its implying a *discursive flow*, because the emergence of meaning and "intelligence" requires the acting-out of productive imagination and discourse against formless matter, the acting-out of vital textures with their space-time and rhythm of compossible connections:

> 'After he had said this, he left her as he did the day before.'—Do I understand this sentence? Do I understand it just as I should if I heard it in the course of a narrative? If it were set down in isolation I should say, I don't know what it's about. But all the same I should know how this sentence might perhaps be used; I could myself invent a context for it. (A multitude of familiar paths lead off from these words in every direction.) (Wittgenstein, 1953, p. 121).

One must observe in the logical dissymmetry between "reason" and "cause" that a "reason" does not suppose an infinite chain of reasons which would be accomplished by a continuous narration or by a narrative in progressive expansion *ad infinitum*. In fact, "reasons" allow one to grasp and generate the discrete, the discontinuous and even the hiatus, for their intelligibility does not rely absolutely on a regulative ideal of all-encompassing unity or totality. There is no logical need for a perfectly unified texture of reasons so that every new "reason" may enjoy vital efficacy. A "reason" tends to appear in ephemeral and local actualizations. Yet, it can reveal a relatively lasting or global effect of self-cohesion under certain stabilizing discursive circumstances. Produced by that self-cohesion, the epistemic (non-conjectural) certainty is always formed against a background of non-reason, an ultimate horizon of certainty which is not generated but spontaneously given as a common soil for further belief and understanding (the Husserlian *Boden* or *Urglaube*). This soil provides *background practical certainties* (Wittgenstein, 1969) that reside in the dynamic architecture of *life forms* and constitute the pragmatic condition for narrative meaning. The *force* of a *reason* derives from its quality of narrative operator and from its linkage with an actual life form: "What has to be accepted, the given, is—so one could say—*forms of life*" (Wittgenstein, 1953, p. 190).

SELFHOOD AS HERMENEUTIC APPLICATION

Reason and *cause* furnish two generative matrices for practical self- and hetero-interpretation. On learning the possible uses of both semiotic tools, every interpreter becomes cognitively motivated to apply them to her "life". In this hermeneutic application, one can either *confound*, *distinguish* or *articulate* their difference and, by so doing, obtain various configurations of selfhood and agency.

One telling example of *confusion*, a symptom of psychopathology or of poetic spontaneity, can be found in the expression assigned to a dementially altered Nietzsche: "I apologize for the poor weather!", denoting a life form grounded in a peculiar ego-pananimism. This confusion contains the fundamental psychophysical belief that *my self* is a *force of nature* (or is embedded in the flowing of natural forces) which can "make rain and snow", and comprises a pseudo-agentic intentionality in the sense that its structure is essentially *pathetic*, that is to say, actions are interpreted as affections resulting from multi-determined webs and loops of events. At the same time, it must be emphasized that such *confusion* can assume many forms and has the merit of attacking solipsism and proposing an

ecological ground for any reason that must always proceed from previous actions and affections linked with certain habits of meaning-assignment: "I wish to hear Brahms *because* it rains"; "I wish to hear Mozart *because* the sun shines". As for the *distinction* between reason and cause, its disciplinary practice consists in establishing two parallel equations: on the one hand, "personality-as-agency" equated with the grammar of free reasons and, on the other hand, "nature-as-objectivity" equated with the grammar of efficient causes (deterministic and probabilistic alike). This distinction assumes an irreducible dualism, validated by cleft categories but invalidated by phenomenological analysis. Let us concentrate, consequently, on the most fruitful, complex and communal hypothesis which posits the *articulation* of explanation (*Erklären*) and understanding (*Verstehen*), be it vertical or parallel, horizontal or sequential-alternated.

This cognitive style of "articulation" reflects a motivating belief that fortifies the adherence to *self-efficacy*, according to which the increase of cognitive self-possession favors a proportional increase of self-control and self-transformation abilities. It would make possible the enjoyment of more power and, thereby, the attainment of more intentional sovereignty and causal hedonism. Thanks to the semiotic work of "articulation", which conjoins semantic dualism with syntactic monism, *my* self-interpretation oscillates *strategically* between the grammar of reasons and the grammar of causes, seeking for an optimal intelligibility within the system of I-phenomena, that "I, s/he or it which makes me do and makes happen in me". Some connections in this system belong to a series of facts causally explainable, while other connections are interwoven in narratively understandable biographies. One may conceive of some vital elements that are *more* intelligible through explanation or through understanding (elements that demonstrate a kind of preferential inclination to one or other interpretive grammar), whereas other elements seem to benefit from the same level of intelligibility in both grammars. Under a meta-interpretive angle, one recognizes that every subject develops her self-theory—having recourse to the double regime of explanation and understanding—in order to describe the paradigmatic transitions in her epistemology of self-interpretation. How can I *understand* or *explain* the fact that I consider this vital passage incomprehensible? Why do I regard this event X as being without reason but flowing inexorably from a knowable cause? Or instead: Why do I believe that *now in face of Y* to interpret myself through self-explanation *makes not enough sense*? Why do I find compelling the recall of those emotions synchronous with episode Z as an *epistemic warrant* of the feeling of self-understanding? These questions are always implicitly or explicitly at stake in the process of symbolic self-interpretation.

In the permanent strategic *shift* between "explanation" and "understanding", accomplished by *my* self, if competent in both methods, it is expectable that such reinforcement of selfhood translates into semiotic transgression and invention which produce a new kind of valuable "biosemiotic diversity". Equally, selfhood lives by symbols that must be identifiable and recognizable. So, the coefficient of transgression goes hand in hand with the symmetric coefficient of semiotic preservation and conservation which enhances "biosemiotic diversity" with "biosemiotic compatibility" (patent in the "symbiotic" relationships between users or inhabitants of the same symbolic ecosystem). The "articulation" allows for a *strategic appropriation* of the powers of each game which can destabilize the pre-established games in order to test their elasticity or to perform and rehearse any novelty opening up new styles of playing or even new games. The practice of multiple games expands the player who embodies, in the last analysis, the concept of incomputable force (or indomitable agentic patterns of possibilities). Nevertheless, the strategic and metacognitive expertise does not elucidate entirely the articulation between different games. There is also an unintended *pathetic and practical dimension* at play that must be highlighted. The "player" moves typically from "reasons" to "causes" under various affective dispositions which express themselves semiotically, for instance in the ambivalence of reasons, in the frustration of the coherence expectancy between reasons and actions or in the lack of reasons (Wittgenstein, 1958, p. 88). At any rate, whenever a reason shows a high degree of inertia, it is vanquished by the dynamics of the cognitive habit of perceiving the ascending genealogy or archeology of causes. The uncovering of causes appears as a last semiotic resort that might signify an experience of learned self-loss. Another possibility not to be neglected is the following: the "player" believes in the ontological and epistemological value of causality as the most powerful binder, the cement of the universe. So, she only retains from reasons their predictive power regarding action, and hallucinates a quasi-mimetic relationship between reason and cause as the very essence of a reason, despite the singularity and the non-necessity of the causal force of reasons. The player strategies move from reasons to causes whenever the predictive power of "reasons alone" diminishes, as is noticeable in the case of reasons reconstructed through narrative retrospection or through moral reassessment of action (Freeman, 2009). The linkage of causes and reasons—and the shift between them—respond to a rational passion for unity and continuity, as conditions of intelligibility. Semiotic invention is the key operation that converts simple intelligibility into *self*-relevant "truth", whose prime mode of constructive expression consists in self-storying.

However, in contrast with the empirical and proof-oriented character of causal knowledge, the function of narrative self-interpretation is axiological, ethical and aesthetic, pursuing and prosecuting the "good form", the pregnant Gestalt, which achieves a desirable *symbolic self-refiguration*. A life story is a symbol which produces what it means. Though ephemeral, its psychological truth is absolute and constitutes the strongest mode of efficiency and efficacy of practical reason. The truth of *my* life story lies in its free vectorial form as a dynamic blazing of paths, a meta-stable embodied inscription, which is always situated in an inescapable onto-ethical horizon, and presents itself as a proactive quest and self-projective orientation. Being one and multiple, *I cannot* unify *who I am*, *I cannot* identify myself as agent, unless *I know how* to structure the process of *my* life as a narrative development toward a higher good, that is, as a "moral space" (Taylor, 1989).

The desirable "good narrative form" constitutes a cultural *prefiguration*, a meta-narrative semantic schema which provides the canon of all narrative (emotionally constructive) configuration and the criterion of a refiguring self-assessment of lived life (Ricoeur, 1983, pp. 105-162). With the autobiographies that give shape to a cultural meta-narrative of "agentic self", like Augustine's, Rousseau's or Goethe's, with the lives of great self-heuristic and self-zetetic characters, like Ulysses or Abraham, Don Quijote or Joseph K., that add substance and paradox to the unstable construal of selfhood, and with all daily micro-narratives that punctuate social co-ordinations, every self-interpreter composes a morphological spectrum of good, desirable, narrative forms. These bio-semiotic formations expose the unity of life and grammar, by nurturing every nascent self with a cultural library of myth-biographical figures which typify the narrative possibilities of a meaningful life; a "good life" being thus a life worth telling.

Doubtlessly, there is no "agent" without autobiographical awareness, but an autobiography comprises inherently a "self-ideology" or "myth-biography" on the optimal narrative sequence. And hence follows an organic body of form and matter, cognition and memory, poiesis and mimesis. The six elements of tragedy, expounded in Aristotle's Poetics and redescribed by Burke (1945) and Ricoeur (1983), show a *stabilizing systemic permanence*, while new semiotic configurations, explored by life and art, are massively perceived as de-structuring, abnormal and teratological. The "incredulity towards meta-narratives" (Lyotard, 1979) has a deep self-interpretive impact on the (de)valuation of certain biographical configurations. Take, for example, the narrator of *Anna Karenina* by L. Tolstoi who confesses his dogmatic belief in a certain Platonism regarding the crystallized *eidos* of a "good narrative/life" when he asserts that: "All happy families are alike;

every unhappy family is unhappy in its own way". In line with this platonic reasoning, every self should find her only mode of composing a perfect texture, a self-Gestalt, or else loose herself irreversibly in myriad modes of tearing the self-text. The figures of fracture and crisis become, evidently, the most significant operators of narrative opening and closure. On perceiving a fracture or a crisis, the narrative intelligence suffers a traumatic shock and paralyzes. Intelligence becomes a *pathos*, hostage of self-skepticism and self-irony; it is the *kairological pathos* which invites one to the acting out of narrative self-rewriting in face of the possibility of nonsense. *A contrario*, "*Croire que je peux, c'est déjà être capable*" (Ricœur, 2001, p. 90).

HOMO SAPIENS FABULANS: ONTO-PHENOMENOLOGY OF ACTION

The intelligibility of life is not a given, but a constructive labor, a vital task, that seems to be rooted in a "drive" (*Trieb*, *conatus*), naturally *explainable* (Damasio, 1999; Gazzaniga, 2006), towards narrative self-understanding which is reactivated by any catastrophe, abrupt qualitative change, able to disfigure one's narrative self-image, and threatens to destroy the writing studio itself. If at any point the world is no longer inhabitable as an "intelligible fable", then *homo sapiens fabulans* fragments and abandons the poetic endeavor (MacIntyre, 1981). Semiosis without poiesis brings about a monologue of repeated, voiceless, disembodied, signs. Only poiesis, as a self-performing art, can transform semiotics into bio-semiotics. Pure semiosis is confined to the naked corpses of signs, and the passage from semiotics to effective semantics is only launched by the living uneasiness of a self-*experiencing and* self-*experimenting I*, whose bodily intensities and imaginative connections merge together making drafts out of drafts, and composing a temporal landscape. Absolute nonsense, tangible in disconnected atoms of now-images and in lifeless signifiers, destroys the possibility of selfhood, for it jeopardizes all possibility of an onto-phenomenological solidarity within the moving triad: event/action, actor/character and author/writer. With the interruption of a unifying drafting labor, the "multiphrenic self" (Gergen, 1991) is no longer a poetic polyphony and regresses to the barbarian age of selfless *inarticulacy* and mute *process*, age of agraphy and alexia. The experience of the blind mechanics of tragedy and randomness, assigned to an exogenous increasingly unpredictable causality, may feed the (para-suicidal) belief in a meta-narrative of nonsense and self-dispossession. Moreover, it can dissolve the narrative competence, thus being selfhood confined to an amorphous, speechless, discontinuous *thisness*, to the passivity and silence

of a blank neutrum, to the brute ontology of events. This brute *neutrum* evokes Merleau-Ponty's (1964) meta-phenomenological concept of *wild being* and *wild meaning*, redefined in a narrative vein by L. Tengelyi (2005, p. 29-30) to signify the *continuous emergence of a dispossessed meaning* which structures and de-structures the always fragile narrative unification of a life. Such dispossession means not only that action and narration begin always *in medias res*, and therefore without a truly original spontaneity, but also that the *course of action* is permeated with events that constrain agents to answer them. Meaning-construction would be virtually infinite for infinite "minds" and "texts", but it is actually finite, because every actual, conditioned, *selving process* possesses limited poietic energy and limited capacity of transcendence in order to (re)constitute her objectified signs and her lived instants. In other words, something cannot mean *everything* (see Eco, 1994).

Once engaged in a narrative performance, the *telling I*—and here *telling* does not exceed the *tale*—must decide what is (in)comprehensible and/or (un)explainable and cannot dodge this task of self-epistemological decision-making and meaning-construction (Velleman, 2009, p. 205). Continuously on the brink of gross performative contradictions, the *telling-and-the-tale-I* must also decide whether she believes or not in the power of agency and how to live the consequences of that (un) belief. Feeling, imagination and belief generate hybrid truthmakers in the phenomenological production of selfhood. The phenomenology of self-efficacy conjoins those *qualia* and maintains the conviction of being cause or effect, active or passive principle, although this phenomenology is patently fallible, there being two possible major flaws: 1. to feel that my *intentional reason* is the force or efficient cause of a certain somatic, motor or physical effect, when such is not the case, due to the ignorance of the true exogenous cause that has simultaneously provoked effects on me and on the contiguous space; 2. not to feel the causal objective force of my reason/action, when it is the case, due to the lack of a conscious representation of the causal link (Wegner, 2002). Illusion of control and selfless automatisms constitute two usual fallacies, but the fundamental fallacy consists in believing in the ultimate proof-value of phenomenological data, for these data have the onto-epistemological status of ambiguous signs that require close interpretation. The evidence of their presence and intensity dissimulates their congenital ambiguity. In a surface semantics they have the value of an irrefutable truthmaker deixis (*index veri et sui*). Yet, for a deep semantics, at the level of cognition which infers and assigns relational functions, those *qualia* demand great interpretative discipline. As a matter of fact, those contingent elements of sensation cannot aspire to the

high status of pure transcendental elements of which one could *a priori* affirm that this *I feel* accompanies necessarily the performance of a self-determined causal power that initiates a new series of events in the world. Instead, they are *impure* elements that can accompany or not accompany the self's action. Furthermore, they may result from learning processes and form an undifferentiated complex of sensations, emotions, imaginations, and beliefs.

The most primitive layer of this "I feel", necessary but insufficient to infer *my* "I do (and make happen)", resides in the awareness or feeling of bodily effort. In this regard, there is a multi-secular inspirational strand of thought which values the positivity of somatic self-affection, whose key concepts include namely: Spinoza's (1677) "effort of being or persevering in one's being (*conatus essendi vel in suo esse perseverandi*)", Maine de Biran's (1807) "feeling of effort (*sens de l'effort*)" and Merleau-Ponty's (1945) "synthesis of bodily awareness (*synthèse du corps propre*)". These concepts work out a constitutive unity between self/hetero-efficacy and self/hetero-determination which demonstrates the nonsense of believing in any *causa sui* taken as absolute spontaneity. Reason-as-cause can only be felt in a very unstable way as a "somatic marker" (Damasio, 1994), an "authenticating feeling" of authorship (Wegner, 2002, pp. 326-327). In sum, the passage from "I feel" to "I do" and "I make happen" is possible but uncertain. Other passages are involved in this labyrinth of discontinuousness which encompasses the grammar, the phenomenology and the ontology of action.

In the grammar, there is no licit passage from an understanding to an explaining self-interpretation. Both semiotic games as such are disjunctive: reason cannot signify cause. In the phenomenological field, that passage exists; it appears to be there but—like all epiphanies—contains a hallucinatory structure. In the absence of metaphysical certainty, the agent enjoys a moral certainty of various degrees. The quality of this moral certainty depends, firstly, on the subjectively or intersubjectively constraining force of *my* psychosomatic evidence, and, secondly, on the subjective and intersubjective quality of *my* narrative co-production. The phenomenological ambiguity comes from the subtle chiasms between action and affection, between selfhood-otherness-thinghood, and between meaning and its epigenetic ecology. Finitude imposes to *my* selving relationships some particular configurations of determination and co-determination that are hardly discernible or computable. There is neither absolute activity nor absolute responsibility, but simply relative activity and responsibility, for one cannot identify absolutely self-determining centers of agency. Whenever "something" resists as incomprehensible in my experience, a detour through explanation may, then, be the best

way towards an enhanced self-understanding. However, in this case, if *I* master the distinction between the understandable and the explainable, then *I* am also responsible for the manner in which my responsibility is semiotically structured. A second-order responsibility can emerge here: I become responsible for conceiving myself as capable or incapable of being responsible.

At last, in ontology the passage must be possible. This paradoxical alliance between the apodictic and the problematic—"*must* be *possible*"—calls for a prudential stance, according to which the logical possibility of multiple disjoined and concurrent worlds (like, v.g., Leibniz' world of representations and world of motions) cannot be declared as nonsense *ex cathedra*. The hypothesis of a perfect ontological cohesion offers the highest degree of intelligibility. "Manyness in oneness" is a metaphysical landscape that appeals strongly to the desire of knowing as its final panorama. In fact, despite their semiotic irreducibility, reasons and causes may be valued as extrinsic denominations, and differential cognitive perspectives, to approach the same real efficacious forces, whose discrimination would lie solely in the contingent varieties of descriptive and interpretive constructions. Reasons and causes may signify differently, and yet merge entirely in one and the same ontological poiesis. Thus, Davidson's "anomalous monism" (2001, 2005) can be reconciled with the principle of deep dynamic continuity and affinity which is the bedrock of all intelligibility within the whole community of being and becoming.

To conclude, semiosis is the structural motion of signs, with their virtual infinity of possible patterns of motion and connection, such as the pattern of causes and the pattern of reasons. Poiesis activates semiosis for actual living purposes; therefore, transforms its geometry into a dynamic event, and its anatomy into a physiological process. When a subject becomes a competent player of diverse semiotic games, she becomes, by the same token, a competent self, that is to say, an autopoietic agent, who continuously instills energy into the fabric of language and, thus, recreates herself by recreating the texture of experience.

REFERENCES

Anscombe, G. E. M. (1957). *Intention*. Oxford: Blackwell.

Biran, M. De (1807/2005). *De l'aperception immédiate (Le mémoire de Berlin)*. Paris: LGF.

Burke, K. (1945). *A grammar of motives*. Berkeley: University of California Press.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam Publishing.

Damasio, A. (1999). *The feeling of what happens: Body, emotion and the making of consciousness*. London: Heinemann.

Davidson, D. (2001). *Essays on actions and events*, Oxford: Clarendon Press.

Davidson, D. (2005). *Truth, language, and history*. Oxford: Oxford University Press.

Descombes, V. (2004). *Le complément de sujet: Le fait d'agir de soi-même*. Paris: Gallimard.

Eco, U. (1994). *The limits of interpretation*. Bloomington, IN: Indiana University Press.

Freeman, M. (2009). *Hindsight: The promise and peril of looking backward*. Oxford: Oxford University Press.

Gazzaniga, M. (2006). *The ethical brain: The science of our moral dilemmas*. New York: Harper.

Gergen, K. (1991). *The saturated self*. New York: Basic Books.

Husserl, E. (1970). *The crisis of European sciences and transcendental philosophy*. Evanston, IL: Northwestern University Press.

Jaspers, K. (1913/1973). *Allgemeine Psychopathologie*. Berlin: Springer.

Jonas, H. (1966). *The phenomenon of life: Toward a philosophical biology*. Evanston, IL: Northwerstern University Press.

Lyotard, J.-F. (1979). *La condition postmoderne: Rapport sur le savoir*. Paris: Minuit.

MacIntyre, A. (1981). *After virtue: A study in moral theory*. Notre Dame, IN: Notre Dame University Press.

Malinowski, B. (1922/1984). *Argonauts of the western Pacific*. Long Grove, IL: Waveland.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Merleau-Ponty, M. (1964). *Le visible et l'invisible*. Paris: Gallimard.

Ricœur, P. (1983). *Temps et récit 1: L'intrigue et le récit historique*. Paris: Seuil.

Ricœur, P. (2001). *Le juste 2*. Paris: Esprit.

Searle, J.R. (1985). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Spinoza, B. (1677/1985). Ethics. In E. Curley (Ed.), *The Collected Writings of Spinoza (vol. 1)*. Princeton: Princeton University Press.

Stenner, P. (2008). A. N. Whitehead and subjectivity, *Subjectivity*, *22*, 90-109.

Taylor, C. (1989). *Sources of the self: The making of modern identity*. Cambridge, MA: Harvard University Press.

Tengelyi, L. (2005). *L'histoire d'une vie et sa région sauvage*. Grenoble: Millon.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.

Velleman, J. D. (2009). *How we get along*. Cambridge: Cambridge University Press.

Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Whitehead, A.N. (1929/1978). *Process and reality: An essay in cosmology* (corrected edition). New York: The Free Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

Wittgenstein, L. (1958). *The blue and brown books*. Oxford: Blackwell.

Wittgenstein, L. (1969). *On Certainty*. Oxford: Blackwell.

MOVEMENT AND INSTANTANEITY ON THE PROBLEM OF REFLEXIVITY

Clara Morando

One of the greatest concerns of Sartre was to finish once and for all with spiritualism in philosophy, i.e. the whole series of doctrines which consider as their theoretical primitive grounds all philosophical ideas concerning the unshakable stance of an interiority, of an irreducible subjectivity laying apparently in a *"vrai moi"*, which would univocally constitute the only certainty we are able to reach. This very direction of thought started with Cartesian *Meditations on First Philosophy*, attaining full-fledged amplitude in the endings of the nineteenth century and the beginnings of the twentieth, namely with Brunschvicg and Bergson. What Sartre aspires the most is precisely, in a broadest sense, to dissipate the illusion of substantialism attached to the Cartesian *cogito* assumptions

SUMMARY OF *THE TRANSCENDENCE OF THE EGO*

The Transcendence of the Ego's (1934-36)[1] core-thesis is one of the most controversial theoretical options in contemporary philosophy. Sartre initiates here a demanding battle against the main premises from both Descartes and Kant, but above all from Husserl, namely about a central unifier in consciousness called *Ego* or *I*. Sartrean philosophical exigencies claim, instead, that there is no self *in* consciousness as the subject of our conscious acts, «*neither formally nor materially*» (*TE*, 1). More: the self «*is outside, in the world, it is a being in the world, like the Ego [self] of another*» (*TE*, 1). One of the perplexities Sartre deals with has to do with the still

[1] From now on, we will make reference to this Sartre's œuvre by TE.

unexplained fact that Husserl, during the course of his life, changed his mind a few times about the very notion of consciousness[2]. In line with Husserl's ground-breaking view about consciousness, defined as being *intentional*, the notion of a unifying self is rendered here superfluous (*TE*, 7), since as the defining feature of consciousness is its being consciousness *of* something, whenever she is consciousness *of* something, the playing role of being conscious *of* something accomplishes itself the unifying activity.

Thus, the unifying principle of my different consciousnesses by which, for example, I perceive a chair, is the unity of the chair itself as an object of perception, not something *"in me"* that unifies these different consciousnesses (*TE*, 6). What is more, the unity of the object allowing for the unity of consciousness renders a unified self possible rather than the other way round (*TE*, 7).

Within this primordial unity of the object for consciousness, but not *in* consciousness (the first achievement in terms of onto-phenomenological precedence, even when relating to the unity of consciousness) we are not able to find here any *egological* stance that would pull the strings, exteriorly, to centralize the very unity as unity. To put inside consciousness a unifier of consciousness would be equivalent to open a regression till infinite. So, the unity of consciousness is not accomplished by anyone; better, is solely and just intuited by objects consciousness is conscious of[3].

So, one of the major claims Sartre endorses, is that the first and most fundamental mode by which consciousness is conscious of itself is not reflective, but *unreflective.* According to him, again, by being conscious of an object, consciousness is at the same time and by the same stroke conscious *of itself* (*TE*, 7-8). This very idea may seem to be an unintuitive one at first sight. When I am conscious of an object, I am apparently conscious of the object, not of myself. However the secret here lies on the

[2] Although we can find in Husserl's complete work a recurrent use of determined concepts as envisaged as being the key-ideas of his way of doing philosophy, that doesn´t mean that those same concepts have not been subject to various nuances of meaning. The truth was that, around 1905, Husserl began to describe his phenomenology on a basis of a transcendental turn, due also to a new appointment of the Kantian ideas. He realized, indeed, that he was severely misled on the treatment made to the ego in the Logical Investigations, as one seriously inadequate. Cf. Moran, Dermot, Introduction to Phenomenology, London / New York, Routledge, 2000, p. 77.

[3] In a passage of Being and Nothingness, 172; 212-213, we find a clear attempt to show how important is intuitive consciousness when facing directly the world, the in-itself: «There is only intuitive knowledge. Deduction and discursive argument, incorrectlty called examples of knowing, are only instruments that lead to intuition. When intuition is reached, methods utilized to attain it are effaced before it... If someone asks for a definition of intuition, Husserl will reply, in agreement with the majority of philosophers, that it is the presence of the thing (Sache) "in person" to consciousness... But we have established that the in-itself can never by itself be presence. Being-present, in fact, is an ekstatic mode of being of the for-itself. We are then compelled to reverse the terms of our definition: intuition is the presence of consciousness to the thing (l'intuition est la présence de la conscience à la chose).»

fact that the level of the unreflective has its own intelligibility too, and the prima facie implausibility we face results mainly from mistakenly assuming that only in *reflection* can consciousness be conscious of itself. We should, then, not see in reflection the only and the original mode in which this can happen; the reflexive soil of philosophy is a derived one, not an original one[4], stressing so the importance of phenomenology in describing such primal territory: of the unreflective.

Summing up. Against the view that reflection is the only mode in which consciousness is simultaneously conscious of itself, Sartre argues that unreflective consciousness of an object is conscious of itself too. In being conscious of the object, consciousness is simultaneously conscious of itself as being conscious of the object. Sartre grants in this case, consciousness is non-

-positionally conscious of itself: it does not take itself as the object of consciousness. This very circumstance does not make consciousness less self-conscious (*TE*, 8, 11).

Sartre argues then that unreflective consciousness takes precedence over reflective consciousness: unreflective consciousness does not require a reflective consciousness to be conscious of itself, whereas reflective consciousness does require a reflected consciousness to be reflective where that reflected consciousness is non-positionally conscious of itself (*TE*, 19).

The defense of the non-primacy of reflective consciousness, grounding his attack to all philosophies lying on a second-degree perspective, is what permits Sartre to destroy the conception of a self in consciousness. According to the author, both Descartes and Husserl think that there is a self as the subject of *any* consciousness-of-something. Moreover, Descartes and Husserl take the *cogito*, that is, the reflective grasp of consciousness by itself (typically in the form of the *"I think"*), as the *de facto* proof that there is a self in consciousness, since in the *cogito* the self never fails to appear (*TE*, 9). Sartre responds that although what is directly intuited in the cogito enjoys a privileged certainty, still the *cogito*, structurally considered, includes two consciousnesses, a reflecting and a reflected one (*TE*, 10). It is to the first one – the reflecting consciousness – we attribute the epithet of the unreflected movement of reflexivity in general – as that mark which gathers commonly the unreflected and the reflecting activities. The non-positional aspect of the reflecting activity is the same as the non-positionality of the unreflected. There is a primitive layer conscious non-positionally of itself that overcomes everything else, as absent of an

[4] Cf. Bergoffen, Debra, The Everlasting Soil of the Reflexive, p. 33.

egological centralizer, verging then into a paradoxical unknown.

Better: the reflecting consciousness is positionally conscious of the reflected consciousness, and non-positionally conscious of itself (*TE*, 10-11). But since the reflecting consciousness is non-positionally conscious of itself, the reflecting consciousness is on its turn unreflected, and thus lacks a self. Although the Cartesian and Husserlian *cogito* are necessarily linked to the appearing of a self, this one is a derived stance and not an original one, because the self only appears in the reflected consciousness, not in the reflecting one. The unreflected/reflecting are the two sides of the same coin, which surprisingly allow, by their own activity, the constitution of the reflected. There is no self neither in the unreflected nor in the reflecting consciousness of objects, and these two modes of consciousness can only be applied whenever they are dealing with objects. I repeat then: consciousness is always conscious *of* something[5]…

The self only and just only exists in reflection. Sartre agrees here, in this precise point, with Descartes and Husserl. The difference is that these last two didn't realize the reflexive was an arriving point, not an initial one[6]. So: where does the self "come from", if there is no self in unreflective consciousness? Sartre's suggestion is that the self does not simply happen to appear in reflection. Rather, reflection itself makes the self appear (*TE*, 11). We should not think of the self as a permanent entity, of which we are not conscious in unreflective consciousness and of which we become conscious in reflective consciousness. In line with these considerations, the self is a volatile entity that reflection produces as much as intuits. The self is in this very aspect looked like with Kantian transcendental unity of apperception in the sense it also accompanies all its representations, although in Sartre it surpasses a mere logical function.

It is not true, as is for many commentators, that either there is or there is not a self unqualifiedly. Sartre unambiguously states that the self appears or is intuitable in reflection (*TE*, 15), and only in reflective consciousness is there a self. The self is as much constituted as contemplated by reflection (*TE*, 34). There is a reflecting activity in reflection that has no self, and which constitutes in its turn the very self. Reflexivity is then a *Janus*-faced operation.

[5] *«Indeed the existence of consciousness is an absolute because consciousness is consciousness of self.» (TE,* 40, 90).

[6] Sartre, in order to make the superfluousness of the egological principle explicit, states that *«Like Husserl, we are persuaded that our psychic and psycho-physical* me (notre moi psychique et psycho-physique) *is a transcendent object which must fall before the* epoche. *But we raise the following question: is not this psychic and psycho-physical* me *enough? Need one double it with a transcendental* I (un *Je* transcendantal), *a structure of absolute consciousness?»* (*TE,* 36; 87-88).

In unreflective consciousness it is not the case that the self simply fails to appear. Rather, Sartre holds, there is no self at all. The self is not a subject in consciousness, but just an object for it, and only in reflection.

HOW DOES, THEN, THE SELF APPEAR QUA OBJECT IN REFLECTION?

The self does not appear in reflected consciousness all at once, in an apodictic and adequate fashion. Rather, the self appears through reflected consciousness and by profiles, in a non-apodictic, inadequate fashion (*TE*, 15, 16). This means that we intuit our own self in reflection roughly in the same way we perceive a physical object in perception –, incompletely and by successive intuitions of different aspects. According to Sartre, is neither reflected consciousness itself, nor an individual reflected consciousness, nor a real set of reflected consciousnesses. Instead, the self is the ideal unity of all our (potentially infinite) reflected consciousnesses (*TE*, 20-21). Finally, the self does not directly unify our reflected consciousness – the self unifies the states, actions and qualities of reflected consciousness (*TE*, 21). States and actions, in their turn, directly unify our reflected consciousnesses.

Sartre does not explicitly say what a state is supposed to be. But from his examples, notably hatred and love, plus some explanatory hints, one can gather that Sartre has in mind something like, classically speaking, passions or, in contemporary terms, emotions. The key feature of the states is that they are inert. We suffer them, instead of choosing them. They are psychical *passivity*.

The radical difference between consciousness, on the one hand, and states, actions, qualities, and the self, on the other, is captured in the fundamental opposition between consciousness and the psychical (*TE*, 28). Consciousness is impersonal (selfless), intentional spontaneity, directed "outwards" by necessity, and conscious of itself either non-positionally or positionally. The psychical, on the contrary, is a non-intentional pseudo-spontaneity, an object for consciousness, and never conscious of itself. The self is, thus, the synthetic totality of the psychical.

THE PROBLEM OF REFLEXIVITY ON DESCARTES

Descartes assures in the First Replies to the objections raised against his *Meditations on First Philosophy* that «*there can be nothing within me (as a thinking thing) of which I am not in some way aware*». In the Fourth Replies he reiterates this view: «*The fact that there can be nothing in the mind, in so far as it is a thinking thing, of which it is not aware… seems to me*

to be self-evident… We cannot have any thought of which we are not aware at the very moment when it is in us». In defining thought in the Second Replies, Descartes says: «*I use this term to include everything that is within us in such a way that we are immediately aware of it. Thus all the operations of the will, the intellect, the imagination and the senses are thoughts*». In article 19 of part I of *The passions of the soul* Descartes states «*that we cannot will anything without thereby perceiving that we are willing it*». He makes this same claim in his letter to Mersenne, 28th January 1641: «*For we cannot will anything without knowing that we will it." In the Fifth Replies he makes the same claim: "For does anyone who understands something not perceive that he does so?*»[7].

Statements like these can be seen as evidence that for Descartes all thought is self-conscious. At least, they turn explicit his adherence to the belief that all consciousness of anything involves consciousness of the self as being conscious of that thing… (For Descartes) all consciousness is self-consciousness.

What means precisely that claim for Descartes?

> 1) On one reading Descartes is interpreted as maintaining that every act of consciousness is accompanied by a second act of consciousness. The first act of consciousness has for its object a table –, and the second has the first act of consciousness for its object;
>
> 2) There is another way to read Descartes' claim that links it directly to Sartre's claim in *Being and Nothingness* that all consciousness is self-_consciousness. On this reading, Descartes is seen as holding that every act of consciousness is self-consciousness. In here, for Descartes, perception and consciousness of perception are one and the same. In the case where what one is conscious of is an external object, one's consciousness of oneself as conscious of a particular external object is not distinct from one's consciousness of that external object. Descartes would here to identify consciousness with self-consciousness. Although, we have to acknowledge that Descartes never developed the implications of this very thesis.

[7] Descartes, René, *Meditations on first philosophy*

Perhaps Sartre and Descartes share this last important thesis about the self-consciousness of consciousness, but still remain important differences between their analyses of consciousness.

For Descartes, at least on a traditional reading, what I am conscious of are ideas and thoughts in my mind. For Sartre what I am conscious of are not objects 'in the mind'[8], because as we know he denies there is a mind in the sense of a container full of contents. Rather, I am conscious of objects in the world or states of my own consciousness. And nobody denies that Descartes believed that the subject of consciousness is an immaterial substance. Sartre denies the existence of such a self. But it remains true that Sartre is heir to one of the most basic claims of Descartes' philosophy: that all consciousness is self-consciousness.

THE DYNAMICS BETWEEN REFLEXIVITY AND THE UNREFLECTIVE

One of the main questions arising from the way Sartre perspectives *transcendental consciousness* has to do with the fact that to know something about it is necessary to, inside the theory's basic assumptions, to be pulled reflectively against it, in order to grasp its contrasting differentiation: so as to achieve the truth about the essential activities of consciousness in the primitive *locus* of the *instant*, we are (reflexively) obliged to open a distancing breach towards that very same original spring of pure spontaneities. The sign of 'instantaneity' (meaning here the operational vector of how consciousness relates intentionally to the world) does not cope in any way with a sort of duration required by the traditional Cartesian reflexivity[9], on its own movement which makes to coincide the act of thinking with a thinker called 'Ego'. There is an hiatus between what the thinker is as a thinking nature, irreducible to a third person approach, trying ultimately to define him objectively but missing at the same stroke the subjective core of the *I*-experience, and the thoughts maintained by the very same thinker, even the one asserting consciously (or reflexively) himself as a 'thinking thing'. The third person perspective can be said, then, to correspond in terms of "natural effects" to what is the common

[8] «All is therefore clear and lucid in consciousness: the object with its characteristic opacity is before consciousness, but consciousness is purely and simply consciousness of being consciousness of that object.» (TE, 40, 90).

[9] Attaining to this idea of a pure and paralyzing *instantaneism*, we are called to note the following quotation about Descartes from Sartre's *Les Carnets de la drôle de guerre: Novembre 1939 – Mars 1940*, Paris, Gallimard, 1983: *«Descartes by refusing intermediaries between thought and extension, displays a catastrophic and revolutionary cast of mind: he cuts and slashes, leaving to others the task of re-stitching.».*

tendency of our consciousness to function in a second-_degree layer (as a reflexive one, as a bending operation of consciousness over consciousness), when we aspire to grasp the nucleus of our first person experience, but without being reflexively aware of the fact that the objectified I and its subjective experience are things totally diverse. The so-called explanatory gap between the first and third person points of view has truly a more deep-rooted motive, not just the one we find theorized within the main objectives of exact Sciences, but also the one which relates to the tendency of consciousness to phenomenically create the illusion of a solid *I*[10] we can phenomenologically grasp by the study of the (reflexive) *cogito*.

If the first strand of the intentional dynamism which characterizes pre-personal consciousness in Sartre cannot be totally unveiled by the traditional modern mechanisms of attesting about a pre-predicative subjectivity, namely the Cartesian one (with less strength the Kantian one by obvious although complex reasons), which rely on a duplicating strategy of consciousness within herself, in order to be able of re-lying on herself, to get to know herself. There's then a clear signal that reflexivity has on its grounds something that is not itself reflexive. Of course, to Descartes (even to Husserl), this was not acceptable, as it would mean the opening of a sort of an unconscious realm, which would seem to install a contradiction within the very definition of consciousness as equivalent to intentionality: consciousness wouldn't be, then, always consciousness *of* something.

The articulation between the markers of the Cartesian *cogito* and the one that relates to an obscure and thrilling immediacy of the activity of consciousness, when operating in a pre-predicative manner, when not establishing a vivid scission between the pole of an object and the pole of a subject, can lead precipitately to the assumption that Descartes has in some way anticipated the level of the pre-reflexive. We assume here that that very conclusion contains itself the metaphor of a power traditionally given to the instance of the *cogito ergo sum,* as being a primal and irreducible strand of the mental, which does not fully correspond to the plane of accuracy we want to achieve. If on one hand Descartes, confusedly and indirectly, inserts a certain amount of ambiguity in the treatment of the *cogito*, claiming that a full-fledged elucidation of its main traits is not possible due to a vertiginous immediacy, to a type of instantaneity embedded on and connected to its performing actualization (the exclusive act of thinking *I*), which goes onto a description of the thinking-act basic core as to be dramatically close to the sensible stratus, to matter, on the other hand, he

[10] *«Thus the consciousness that says* 'I think' *is precisely not the consciousness which thinks. Or, rather, it is not its own thought which it posits by this thetic act.»* (*TE*, 45).

also tends to voluntarily ignore the real significance of that undeniable proximity between what is lived and what is thought[11].

Returning to the Cartesian stance and to its inability to explain a sort of a hidden layer, i.e. infrastructural to the performance of an Ego, of a *cogito,* within the instant of an active thought of an *I*, we assume now, then, that it can be found here one of the reasons that always impeded Sartre to recognize truly a completeness in the marker of consciousness as Descartes described her: the assumption of the *cogito* as strictly reflexive, even though there are on it signals of an impassable weakness inside its reflexive force, pointing out to a still non-theorized vector of that very same *cogito*, precisely the moment when the reflexive I (the *ego cogito*) is produced within the movement, apparently instantaneous, of reflexivity. The problem here lies mainly not on the constitution of the object-pole Ego, or of a thing in general, but corresponds to the reflexive movement of that same constituting operations (of objects, I repeat, whatever they are, the Ego included), which is not itself reflexive on its grounds. If it were, the thinking-act, the very strand of thought as that which permits the grasping of a 'world', would be impossible, it would be paralyzed or, on the contrary, accelerated till the vertigo of destruction, and, at least, it would have no connecting roots with the sensible, with matter, again, with the 'world'. Reflexivity cannot be, despite Cartesian intents, an infinite mirroring labor as something suspended in the ethereal plane of a worldliness thought, and which faces also the great risk of a permanent dissolution as an irreversible splitting of the pieces belonging to an unbounded puzzle: the puzzle of consciousness. As I said above, this does not mean that for Sartre to explain the pre-reflexive trait belonging to the movement of reflexivity he needs to ultimately endorse the necessity of an unconscious territory[12].

It conveys to our exegesis on the present case of the bipartition pre-reflexivity/reflexivity to stress the fact that Sartre forged the idea of a pre-predicative consciousness basically on a work of Husserl – *Zeitbewusstseins*–, retaining, principally, its conception of a self-unifying consciousness throughout time, on a very Kantian-fashioned insight, which puts on evidence the lacking of fluidity, or of *poietic* movement, (un) detained by the Cartesian *cogito.* This one does not possess the resources

[11] Of course, we already know, Kant will after, in a systematic and within a *criticist* picture of the problem about the subject (i.e. of subjectivity, in its epistemic and epistemological senses), to enhance the architecture of a transcendental dynamism of consciousness which in some sort, through the potential of an *a priori* synthetic, gathers brilliantly sensibility and understanding, escaping from the menace of a confusing linkage extension-thought which fails to give us its deciphering-key.

[12] Sartre states that the assumption of an unconscious would be a non-sense, since pre-reflexive consciousness, as Merleau-Ponty claimed, is consciousness *de part à part,* is absolute consciousness.

to evade himself from the danger of intermittence, i.e. only just whenever pronounced he asserts its very truthfulness through the strategy of playing a sort of arithmetic's on duration, of cutting thin and with artificial rigor units of time, inter and intra-separated, that correspond ultimately to what is called *cogito ego cogitatum*. On Sartre, otherwise, we can see, although in a more sophisticated way than in Husserl, the defense of an "internal consciousness of time", to then unveil the non-thetic consciousness For-itself, and after to mark with the value of a leading-point the existence of an irreflective potency on consciousness. We stress here the fact that Sartre himself was inspired in several passages of Husserl, namely the ones found in the § 39 of *Ideen* and in the Ninth Supplement of *Leçons*.

In the Husserlian conception of consciousness, as being almost the same shared by Sartre on what concerns the general topic of intentionality, referenced above, there are two types through which consciousness relates to something, whatever that something is, which is in itself not consciousness, and they are mainly the longitudinal and the transversal ones. What counts, then, for our purposes of deciphering pre-reflexivity in a Sartrean fashioned-way, just in order to evidence the absence of theoretical concerns in the Cartesian picture, has to do only with the longitudinal intentional stance Husserl refers to (§ 39, Leçons / Ire, 26, 192). I.e. the retention of the retention, as a way intentional flux has to grasp him by himself in a non-thetic modality, that which will after open to the very possibility of a/the *cogito* event. There is no need, indeed, to pose in this primal level – the longitudinal one –, a second degree consciousness, a reflexive intentional derivation that would duplicate, by partitioning what should be undoubtedly *unum* – the intentional flux of consciousness –, and then picking up on one of the fractioned parts of that primeval intentionality in order to reify it under the image of a unifying object – in this particular case, an Ego: an egological objective pole whose main function is to set the course of an artificial and exteriorizing conjunction of what is left (of what has not been used to build the image of an Ego) from that first level of consciousness. Relying on this strategy to unify consciousness from that which is just a product of consciousness, and so only a small part of it, a reduced consciousness from a latter consciousness, stipulates an always remaining blind spot within the intentional surface of consciousness. Consciousness can never fully objectivate/objectify herself since she is, for Sartre, a pure intentional movement, and movement as movement cannot self-elucidate by stopping the very mechanism of movement: it would rather be something else. Consciousness is, then, only able to "objectificate" a part of herself, which does not still minimally correspond to its true nature, being just no more than a degraded image of

what really counts here – consciousness as (entirely) consciousness.

So, the self-apparition of the flux in the Husserlian phenomenology does not require a second consciousness which would be itself equivalent to a reductive re-petition of one layer of consciousness over another, or even, as previously shown, to a sort of artificial scissoring in itself, a reflexive one, that falsely pulls away the problem of reflexivity.

Summing up: Sartre retains from Husserl, rigorously, that the unity of the temporal objects rests on the very unity of the immanent subjective apprehensions, the entire picture of them implying by its turn the self-unification of consciousness. What Sartre precisely gathers to or makes explicit in Husserl's conception of a self-unifying intentionality (only theorized in the Husserl of the *LI* and of *Zeitbewusstseins,* the so-called first one) has simply to do with the fact that that unitary consciousness is since the very beginning (always) pre-reflexively self-aware (*de* soi), without being then doubled and crystallized in a detachable and namable object of (conscious) attention by the very movement of reflexivity. There is so an invisible moving vector which crosses thought, and this one has been yet pointed out, although scarcely, by Husserl when he stresses the importance of the livings, when he figures out that on the instant of directing consciousness towards world there is a vivid and a (non-conceptual) intuition about a silent potential force lying on the dynamics of reflexivity: the very act of it…

Sartre, on the other hand, just puts as fundamental the primitive intentional movement of consciousness, without submitting, as Husserl does, pre-reflexivity to the horizon of a phenomenological reflected product, this one emerging from that very reflexivity with the status of a mere derivation, but managing, despite all critics, to illustrate reflexivity as a Janus-faced potential containing the hypothetical virtue of turning the non-thetic into what we call a knowledge (a thetic object of something). The pre-reflexive *cogito* is, indeed, a necessary condition for the Cartesian *cogito,* and the way of being taken by (transcendental) consciousness is an intentional one, that can also here be named as a reflexive one, just in the sense that whether reflexivity or intentionality represent basically a mirroring activity, i.e. an energy or a power of awareness of being glued to worldly objects without the ontological "obligation" to open from them a reflexive distance which safely allows to identify, to nominate, to organize. The very operation of a reflected object forged through reflecting consciousness does not entail, then, a limitless second order though, since that would be onto-logically impossible. The reflecting act is itself a trace of a pre-reflexive stance, non-objective, archi-phenomenal, anonymous and autonomous. Reflexivity envisaged as an operation of thought and not as

something static, definite, finished under the rigid mode of *the* (reflected) object, concerns now to an inversion from the cannon attached to its very description, i.e. representing then, I repeat, an absolute differential to what is traditionally seen as its equivalent: the *cogito ergo sum*.

REFLECTING THE NON-REFLECTED

In line with all these considerations about the dyad pre-reflexivity/reflexivity, or better, about the non-reflexive nature of reflexivity and its assumption as an intentional reality, empty of contents on Sartre's onto-phenomenology, we dare into questioning the possibility or the attempt to theorize what is by definition translucid, "agitated" as an unpredictable wind, attaining 'things' equally as a free power, which main goal is to exercise power independently of on what it is exercised. It seems to exist, here, in this absolute or in-human omnipotence of primal consciousness, a sort of superfluity, a useless manifestation of infinite potentialities, turning the advent of an *I* (of an *Ego*) something gratuitous, almost as an unexpected accident. Is then transcendental consciousness who decides about a kind of egological advenance and that has not absolutely nothing to do with us, as beings supposedly self-aware of our own selves…

Insisting on the problem: how is possible to a translucid consciousness, to a basic movement of intentionality towards what is not intentional, to grasp (to grab) entirely the force of its own significance? If reflexivity is simultaneously pre-reflexive it means that she is always an incomplete task of awareness. At last: how can we 'see' the invisible? How could Descartes declare the primacy of the *cogito,* the implacable absoluteness of the *I think*, *therefore I exist*, if the ray of the 'thinking' possesses for itself, in its own act of segmentation, of realization, mysterious territories inhabiting in itself. If the very act of thinking were entirely self-deciphered it would simply not exist. When Kant affirms that Descartes' *cogito ergo sum* is a tautological proposition, in the *Critique*, he stresses not only that the thinking is *qua* a logical form of existence (and only a logical one – a logical function of transcendental synthesis, to be more accurate), without the necessity of implying concomitantly an egological existence – i.e. an object-*I* which thinks –, but he also focus attention on the specter of the indeterminacy that haunts the activity of thought in its very course.

When I say I think, therefore I exist, the I that become aware of the equivalence thought/existence is not the same as the one who activates and practices this particular act of thinking. The problem of the transcendental and the empirical… Reflexivity implies a chimerical gathering or, on the

contrary, a vanishing frontiers strike between the transcendental and the empirical realms.

THE BINOMIAL TRANSCENDING-ACT / TRANSCENDED SUBJECT-OBJECT

What calls attention the most, inside the problematic of reflexivity, as treated over the centuries as an ever non-surpassed topic by a wide number of philosophers and philosophies, is the fact that when studying this very issue we are repeatedly lead to the necessity of using visual metaphors. This recurrent sort of obligation seems to be due to an always-activating circularity within the pole-terms traditionally named subject and object.

When arrived at the Sartrean singular picture of subjectivity, there is a clear presupposition about what we could nominate (someway Hegelianly) a (circular) dialectic put up to movement by a selfless transcending consciousness power, that in a perpetual register of laboring, of functioning (and maybe a functional-ist one too), breaks instantly the contours of everything apparently absorbed into the rigidifying mode of an objectified existence – the subject included –, i.e. through an unstoppable dynamical blockage to every reflected object to be reflected once and for all – whatever it is –, it lacks the possibility of an enduring and sustainable "ontology" of things, impeded because the always-transcending consciousness way-of-being lies on a never-ending reflecting tendency which cannot itself be reflected. It resists to the light of consciousness conscious (or reflexive) acts an ever-cleared penumbra within what is its own activity whatever the pursued (or illuminated) objects, inclusively the object of herself only grasped by herself as an object. Objects in general, and as we know now better, also the specific object corresponding to an Ego, are inexorably constant de-actualizations onto the vast tissue of the 'world', an ever-finished one, lying on here the explanation why psychically we are, as derived objects to primitively non-objects, never self-guaranteed in terms of emotions, feelings, actions, qualities, states, etc., i.e. in terms of mere reifying psychological theoretical descriptions. It exist then a mysterious power of consciousness (in the sense of being cognitively unknown for us) connected to the very movement of its reflecting nature, and it is during its (temporal) realizations, whenever she "functions", in the spontaneous reflecting movements of its own, binomially pre-reflexive/reflexive, that an unpredictable margin of acting-consciousnesses constitute their deep-

rooted sovereignty[13].

Returning to the question of an always transcending pole of consciousness, the reflecting one, that which inclusively turns itself into a transcended object while being a reflexive transcending operation, as we mentioned above, the main difficulty to be stressed here has to do, firstly, and as already suggested before, also, with the onto-logical inability consciousness suffers to surprise herself in the heart of its own existence – when existing, when being conscious of something as something, precisely.

For achieving a complete voyage within circularity of consciousness by consciousness, to grasp vividly that very circularity by surrounding the circular, but without detaching ourselves from what is surrounded in that such strange and inhuman path of circularity by circularity, of consciousness, it would be necessary to stop over and over the primal mode of reflexivity to operate. I.e. it would be necessary to impede her to leave a trace of limitless reflected objects, inclusively of the reflected object which in a non-adequate fashion, although apodictic, corresponds entirely to the very consciousness (as an object to a non-object)[14]. The conceptual is then the reverse of the non-conceptual; what this means is that the fading of such a distinction would mean the end of a vital tension between the vectors of immanence and transcendence, the destruction once and for all of the possibility of reflexive consciousness.

That is why maintaining the symbols of an instant (or instantaneous) immanency and, at the same time, of a fluctuating and always-changing transcendence, somewhat delayed to that first mentioned strand of immanence, represents an unique key to understand, even though in a minimal mode, the onto-phenomenological vision of an all-clearing consciousness, as being the mother-soil from which everything else can be elucidated, despite its non-linear characteristics. Reference to oriental approaches to consciousness.

[13] We could serve ourselves from the possible metaphor quantum mechanics gives us and dare to establish a free analogy between the basic indeterminacy we find in the realm of micro-physics, where laws of time and space are neither Aristotelic nor Euclidean, and the indeterminacy we therefore conclude to exist in the invisible realm of pure spontaneities emerging (as non-contents) in the surface of transcendental consciousness.

[14] In the disturbance of this bi-univocal circuits of reflexivity/pre-reflexivity lay several triggering reasons, in terms of phenomenological description, for what we call psychopathologies, namely schizophrenia.

THE TENSION IMMANENCE-TRANSCENDENCE AS SIGN OF A CONSTANTLY OPENING SUBJECTIVITY

Insisting on the ability of consciousness to suddenly grasp herself during the very act of thinking, and the incompleteness we know the same task endows, it is important now to stress that:

> 1) when we consciously try to do so we are necessarily driven to the "natural strategy" of posing an image;
> 2) as such is so, then, is absolutely required a sort of "unnatural" and impossible *de facto* concomitance between the act of grasping consciousness thinking (merely without a subject) and the subject who thinks that that very same consciousness is thinking;
>
> 3) and, finally, c) that all of this methodologically discerned steps are operated in such a brief instant, that we could say it corresponds to the figure of a mathematical point of consciousness, and as minimally a point so condensed, conceptually so abstract, that is unviable to discriminate any fluctuation, any movement, whatever, on it. The maximum of a synthesis implied by the sudden activation of a *cogito* in time corresponds, analogically speaking, to the figure of a mathematical point, and this very metaphor offers us a hint on how the paradox of self-consciousness can be articulated.

Most of all, the subject has to see himself "looking at something" likewise in a mirror, which by its turn mirrors some image, a mirror which is in itself instantaneous and immobile. At that very moment, it happens a sort of sudden closing-up of consciousness on itself and on the object which was ob-jected to a consciousness, via an all-illuminating (or obfuscating) operation, in order to "visualize" such an image, trying not to loose its vividness and spasmodic nature. Suggesting since the very beginning of the essay that this particular self-performance is rather devoted to failure, the dream of an interior lucidness entails ultimately the nightmare of a permanent unsearchableness of the intimacy consciousness should enhance with herself: «Men is a useless passion» (*BN,* 670). The specter of an invading exteriority menaces, then, at every moment, a pseudo-interiority which the heralds of egological subjectivity always defended vehemently.

So: one of the problems of the Cartesian *cogito ergo sum* has precisely to do with its dooming instantaneity, since he is true every time it is

cogitated, but also unreachable in practical terms by the grasping activity of consciousness. Posing an Ego as the controller of that very *cogito* simply puts away the problem without truly solving it.

When the *cogito* is doubled in a sort of reification of its own traces, by sketching on him, then, the figure of an exterior mathematical point, a kind of point de fugue kept at a safe distance from us, fabricating its identifiableness, comprehensibility and objectivity, then, the movement and rapidness that should characterize him turn to be a dead instantaneity, not sustaining a real connectivity within its very instantiations. If this were so with the notion of transcendental consciousness Sartre has in mind, when he talks about pure spontaneities, of each one of the pure moments (or movements) of intentional consciousness, then, they would have no linkage with the other ones, they would exhibit an absolute value on their own that would correspond to an absolute lack of it, since useless, unconnected with a wider meaning as the one a 'world' would possess.

Do not forget that for Sartrean phenomenology consciousness gets its unifying stance not by itself, not through internal triggering of a unifying machinery inside her, but it attains unity by what is not her, by objects (of consciousness) that are not *in* consciousness, even the object *Ego*. If consciousness were not empty of contents, translucid like an all-free intentional wind, it would bear the same problem as, for instance, the Cartesian *cogito*: it would be dispersed over and over on a never-ending multiplicity of instantaneous objectified intentional moments, and lacking the possibility of gathering them onto the figure of a unity. Of course, in Descartes, we do not mention the word 'intentionality', since consciousness is turned absolutely to itself, as we see by the instrument of a hyperbolic doubt which main function is to detach greatly the autonomy or self-subsistence of that very same consciousness; in Sartre, otherwise, transcendental consciousness is rather a sort of nothingness, a mirroring primal surface detaining its own movement or dynamics, created *ex-nihilo,* glued to things which are different from her, accompanying them by making of them objects for a consciousness. Consciousness is here turned outwards, instead of being self-absorbed. However, this does not mean that the problem of the dispersion of conscious consciousness has been entirely step aside in terms of the present explanation. Reflexivity on Sartrean analysis just escapes from dispersion not only because it endorses the conception of a self-unifying consciousness from the first Husserl, the one of the LI and *Zeitbewusstseins,* to which we referred to in the beginning of this article, but also because it sustains a *Janus*-faced nature as being simultaneously pre-reflexive/reflexive. It is the pre-reflexive level (even though is not correct in the particular context we are to use the word

'level') which opens the possibility of a flowing-work of consciousness; it is, then, the unreflected reflecting activity within reflexivity that allows phenomenological unity of consciousness. What emerge from here, thereafter, are the discontinuities of the psychical (states, actions, qualities and the Ego, ultimately), a very different and derived strand linked more often, i.e. in a traditional picture, to what we call subjectivity in a broader sense.

To ascertain strongly the discontinuous marker of the reflected objects of consciousness (not the objects themselves which consciousness reflects, and also the very act of unreflected reflecting), we could say that Descartes poses his theoretical achievements in a very high level, the level of the (reflected) reflexive seen erroneously as a leading point from which everything else must succeed, whereas Sartre stresses out that the true soil of philosophy is a pre-reflexive one, not the one on which we find already constituted under the mode of 'things' what is definitively a 'non-thing'. Indeed, he establishes and defends the thesis that the ego is neither formally nor materially in consciousness; thus concluding that nothing can be accepted as forming an intrinsic nexus of consciousness which either directs or determines its projection towards objects or its intentional act in any manner (*TE*, 31).

The same critique can be done to Husserl, mainly in the sense that posing an Ego (as a *monad*), (and we are referring now to the so-called second one – the Husserl of *Ideen,* of the *Cartesian Meditations*), seems to be a way of installing an already sophisticated treatment of what is essentially anonymous under egological imposing, and to suspend / to paralyze, again, in a sort of aseptically environment, what is supposed to be fluid, non-breakable or partitioned into several instants between them disconnected. In fact, Sartre arrives, on its turn, at the radical notion of consciousness as nothingness (after radicalizing Husserl's reduction), and establishes the primary notion of a formless consciousness which, in spite of its formlessness, or nothingness, is also intentional (*TE*, 38).

It must there exist, then, a micro-continuity in transcendental consciousness to ensure apparent macro-discontinuities (the psychological reified elements); otherwise, the last ones would be suspended in the vacuum of the mental and would have no *raison d'être.* So: the psychologist's error in confusing reflective experience with consciousness and unreflected experience with the unconscious is merely an artificial creation of unnecessary psychological dualism structured wholly by consciousness (*TE*, 55-56).

Consciousness is constituted whether as a pre-reflexive reflecting activity and as a territory full of reflected objects, rendering its per-

formance to the first one, having then just a borrowed vividness from the spontaneous purities transcendental consciousness perpetrates. Also the hypothesis of the existence of an unconscious which would reign over consciousness, raising the question about consciousness's incompleteness and opaqueness, seems to Sartre a non-sense simply because consciousness, in order to be what is, needs to be consciousness *de part à part.* Posing an unconscious consciousness would represent for him an absurdity[15].

Consciousness as a transcendental dynamism, in the sense of being original and uncreated, endows an absolute immanent movement of intentionality (towards *worldly* objects); however, at the same time, she re-creates a full plan of transcendences (the psychic) which do not share the primal *kinetics* of pure spontaneities and are then doomed to be just a fragment, an instantaneous and falsely immobilized fragment, of something unnamed and *really* unknown for us. From pure immanence irrupts then a territory of pseudo-transcendences, the Ego included, that are, nevertheless, necessary to survival, i.e. to placate the anguish consciousness feels for the fact of being absent of an Ego. The Ego is then pulled up to a leading position of creation and unification of states, actions and qualities, when he is, only and just only, a simple product of an unbounded and non-egological power of consciousness. Not only the famous *Cogito ergo sum* of Descartes, but also the Husserlian monad of an *I*, would already correspond or suffer from the prejudice embedded on a view that only contemplates a second-degree layer of consciousness – the reflexive one –, without recognizing the fact that when studying the same consciousness is necessary to have in mind that the reflected, or in other cases, the cognitive (for instance), correspond to just one of the several stratus of consciousness.

A question remaining from all the topics discussed above, lies on the mutual (or non-mutual) implication between instantaneity and immobility, in and throughout consciousness. Is it possible for the *instantaneous* to be absolutely immobile, or does it imply, on the contrary, a sort of a minimal duration (*durée*) which introduces the possibility of *bare hiatuses* within the tissue of, for example, the very moment when the Cartesian *cogito* is pronounced, or when Sartrean transcendental consciousness self-protects from herself and inverts the order of creation by instantaneously cutting on the intentional surface the dead figures of the *psyche*? Can these last ones enjoy a proper dynamism instead of being fatally static and faded up?

[15] The tension between immanence and transcendence symbolizes in the Sartrean account of a-subjectivity the playing of an important antinomy, due to the twisting consciousness elaborates with its own instruments of self-deception (or of *bad faith*).

Sartre would vehemently give a negative answer to this question, however, it might prove useful if we regard carefully the great amount of abstraction required to think about notions as instantaneity and absolute immobility. These ideas are 'the ideas' we never get to fully accomplish, so paying special attention when using them to describe whatsoever (consciousness in this particular case) can represent, instead, a step forward.

BIBLIOGRAPHY

Breeur, Roland (2005), *Autour de Sartre. La conscience mise à nu,* Grenoble, Jérôme Millon.

Ey, Henri (1967), *La conciencia,* trad. Bartolomé Garcés, Madrid, Gredos.

Husserl, Edmund (1950), *Idées directrices pour une phénoménologie,* Paris, Gallimard.

— (1951) *Fenomenología de la conciencia del tiempo imanente* (*Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins*), Buenos Aires, Editorial Buenos Aires.

— (1953), *Méditations cartésiennes,* Paris, Vrin.

Poulette, Claude (2001), *Sartre ou les aventures du sujet. Essai sur les paradoxes de l'identité dans l'oeuvre philosophique du premier Sartre,* Paris, L'Harmattan.

Priest, Stephen (2000), *The Subject in Question. Sartre's Critique of Husserl in* The Transcendence of the Ego, London/New York, Routledge.

Reimão, Cassiano (2005), *Consciência, Dialéctica e Ética em J.-P. Sartre,* Lisboa, INCM.

Ruyer, Raymond (1966), *Paradoxes de la Conscience et limites de l'automatisme,* Paris, Albin Michel.

Sartre, Jean-Paul (2003), *La transcendance de l'Ego: esquisse d'une description phénoménologique,* Paris, Vrin [1934].

— (1983), *Les Carnets de la Drôle de Guerre,* Paris, Gallimard.

— (1960), *Esquisse d'une théorie des émotions,* Paris, Hermann [1939].

— (1949), *L'être et le néant: essai d'ontologie phénoménologique*, Paris, Gallimard [1943].

— (1971), *L'idiot de la famille: Gustave Flaubert de 1821 à 1857,* Paris, Gallimard

— (2005), *L'imaginaire: psychologie phénoménologique de l'imagination,* Paris, Gallimard [1940].

— (1936), *L'imagination,* Paris, Félix Alcan.

Zahavi, Dan (2008), *Subjectivity and Selfhood. Investigating the First-Person Perspective,* Cambridge (MA), MIT Press.

THINKING CLEARLY ABOUT MUSIC*

Vítor Guerreiro

ABSTRACT

In this article I argue against the arbitrariness of the concept of music and for an essentialist and naturalist framework, according to which music is a cross-cultural human phenomenon, defined by relational properties held together by uniform features of human nature. Building on Dickie's classification of theories of art *in natural kind theories and cultural-kind theories*, I argue for an enhanced natural-kind theory (which explains the institutional element), and use some developments in social ontology to show the inadequacy of an institutionalist approach to art and music.

Keywords: *Music, Art, Definition, Ontology, Social Kinds, Institutionalism, Proceduralism, Functionalism, Naturalism.*

* After MLAG's First Graduate Conference, this essay was published in teorema Vol. XXXI/3, 2012, pp. 25-47. The editors of this volume are very grateful to teorema's publishers for their permission to reprint here the final version.

What is music? Here is a question not easy to answer with anything truly insightful, as opposed to something true but trivial, such as "music is organized sound" or "music is sound evolving in time", or (worse) some inspiring and obviously false statement we're supposed to interpret figuratively. A notorious example of inadequate definition is Joseph Beuys' "Art is life, life is art" [Stachelhaus (1991)]. While this may express the insight that art and aesthetic concerns are pervasive in human life, not just on a few established art forms, and that we can find artistic elements in many activities that aren't standardly classified as art, it is hopeless as a definition.

The extant literature on the definition of music is scarce. Recent contributions are found in Jerrold Levinson (2011), Andrew Kania (2010), and Andy Hamilton (2007). I will not discuss these at length here, since my purpose is more methodological than definitional: I don't seek to provide a new definition of music but to enquire what *kind* of philosophical theory of music we should endorse, though at the end of the essay I clarify the connections between the ideas I develop here and one of the extant definitions of music.

The problem of defining music is independent from the problem of defining art. In principle, we don't need to know what art is in order to know what music is. Nevertheless, the same (kinds of) rival theories that seek to explain the nature of art can be brought to bear on the nature of music, though the arguments for them will differ. For instance, maybe the nature of music is best explained by a functionalist theory, examples of which are aesthetic theories (theories that rely on the notion of aesthetic properties or aesthetic experiences); or maybe it turns out the best definition is an institutional or an historical one.

THE FRAME THEORY AND THE PROJECT OF DEFINITION

Maybe the most widely accepted view on the nature of music (or at least one that fits well with the Zeitgeist, though not a default philosophical stance on the subject) is a kind of folk institutional or procedural theory: "Music is whatever a musician says it is." This is what we may call a "frame theory", following a witty remark by Frank Zappa:

> The most important thing in art is The Frame. For painting: literally; for other arts: figuratively — because, without this humble appliance, you can't know where The Art stops and The Real World begins. You have to put a 'box' around it because otherwise, what is that shit on the wall? If John Cage, for instance, says, "I'm putting a contact microphone on my throat, and I'm

> going to drink carrot juice, and that's my composition," then his gurgling qualifies as his composition because he put a frame around it and said so. "Take it or leave it, I now will this to be music." After that it's a matter of taste. Without the frameas- announced, it's a guy swallowing carrot juice. So, if music is the best, what is music? Anything can be music, but it doesn't become music until someone wills it to be music, and the audience listening to it decides to perceive it as music [Zappa & Occhiogrosso (1990), p. 140].

The "frame" here isn't essentially a material object but a *procedure* that may or may not be signaled through a material object. Make any kind of noise you want, or record any raw sound, natural or artificial, lacking rhythm, melody or harmony, or produce a silent performance, allowing people to hear fortuitous noises external to the performance, present it to an audience ("frame" it) and *voilà*: music. What worries me though is the following: *what does it mean* to say of something that it is music? What does "willing something to be music" mean? What is the *content* of such an intention? And if musicians have the power to turn anything into music by sheer force of will, what happens if different musicians *disagree* about the music-status of a particular case? Should we say that it is *and* it isn't music? Is it "the artworld" that has final say? But we can easily imagine examples of cross-cultural, cross-temporal and crossmodal artworld disagreement, not to mention simpler cases of institutional disagreement within the same culture, the same time and the same world. How are we to make sense of that?

Let me call your attention to a reasonably well-known statement by the composer Edgar Varèse, one of the pioneers of electronic music:

> Although this new music is being gradually accepted, there are still people who, while admitting that it is "interesting," say, "but is it music?" It is a question I am only too familiar with. Until quite recently I used to hear it so often in regard to my own works, that, as far back as the twenties, I decided to call my music "organized sound" and myself, not a musician, but "a worker in rhythms, frequencies, and intensities." Indeed, to stubbornly conditioned ears, anything new in music has always been called noise. But after all what is music but organized noises? And a composer, like all artists, is an organizer of disparate elements [Varèse e Choun Wen-chung (1966), p. 18]

Here Varèse seems to be making the suggestion that "music" is (*or should* be) a general term for "sound art", not restricted to sound events produced with traditional instruments and organized in tonal structures, though he speaks of "rhythms", which are a traditional ingredient of music. Any soundoriented activity with an artistic purpose and any sound event

produced by such activity would be, according to this, music. This would make the definition of music dependant on the definition of art: as if "music" referred more to the artistic medium of sound than to a specific way of working that medium. This is implausible in that there is more to the identity of an art form than the identity of its medium – different art forms can share the same media (e.g. jewelry and sculpture) So more must be said about the relationship between the concept of *music* and the concept of *sound art*, even if we further qualify the latter as non (primarily) linguistic sound art, to exclude cases of spoken poetry, drama, and the design of things like public water fountains, which include acoustic aspects. Even when all of these are excluded, there may still be more than one sound oriented activity subsumable under *sound art*.

Though apparently dismissing the question of whether a given sonic work is also a *musical* work, in the same article, a couple of paragraphs later, Varèse seems to unwittingly reintroduce that question:

> But, considering the fact that our electronic devices were never meant for making music, but for the sole purpose of measuring and analyzing sound, it is remarkable that what has already been achieved is musically valid [op. cit., p. 19].

This contrasts with what was said before, since it seems that in addition to being organized noises, some sound events are also "musically valid", a property which they can arguably lack, if there is any sense to the word "remarkable" in that sentence. And even if no one had "stubbornly conditioned ears" it would still make sense to ask in what being "musically valid" consists, which seems to me another way to phrase the question "what is music?" since there can be no "musical validity" if there is no objective difference between music and non-music (whether or not the concept of music is an "evaluative" concept) and no matter how vague around the edges that concept is. To be "musically valid" can only mean "to satisfy conditions for musichood".

Now, can a "frame theory" be a satisfactory theory of music's nature? Is the concept of music an arbitrary concept, one that applies to whatever we decide it applies? Is it the concept of a culture-bound reality, so that nothing can be music except in a culture that has some concept of music? Or does the concept of music we *seek* (whether or not that is the concept or concepts we *have*) actually pick out a cross-cultural, non-arbitrary human phenomenon, a universal human feature?

The reason an institutionalist or proceduralist approach has some *prima facie* plausibility is that, in a sense, we really decide what is art and

what is music, but not in a way that vindicates the metaphysics of the institutionalist or proceduralist (both "frame theorists"). That is, we can establish arbitrary rules about what counts as art and what counts as music in what context. We can create *art-institutions*, just as we can create all sorts of other institutions. We can also extend concepts beyond their original domain of application (Pierre Schaeffer's concept of *musique* concrète is one such example). But this still leaves us with the problem of why we have those institutions in the first place and what individuates them as art institutions.

The problem of the definition of music is often presented as a problem about the *concept* of music. However, we should clearly demarcate *concepts* or *representations* of reality and reality itself. Maybe this line tends to blur when it comes to social kinds because of a certain background belief that social kinds aren't "really" a part of the furniture of reality. We shall now try to unblur this line.

CONCEPTS FOR SOCIAL KINDS

We must be very careful when talking of *the concept of* music or *the concept of* art, as if it was any clearer what a *concept* is than what music and art themselves are. It is not obvious that we refer to things by "expressing concepts" with our words nor that concepts aren't just a philosophical invention. Here is a tentative view on how we arrive at concepts: we start by having coordinated noises (words) that refer to roughly the same things in virtue of perceived similarities that might prove misleading. In time, our discernment of relevant similarities becomes more and more fine-grained; we form provisional lists of properties that apparently all cases of *X* have in common, calling such lists "concepts of *X*", and as we abstract more and more aspects of the things referred to by the same coordinated sounds we come to realize, in some cases at least, that they in fact share a common nature, and with each addition or subtraction from our list we have a *reformed* conception of *X*. So "music" and "art" are such coordinated noises, by which we refer roughly to the same activities, objects and events. In time, we either discover that different things we refer to by the same coordinated noises have in fact relevant similarities or share a common nature or not; we either discover that those activities, objects and events are (relevantly) cross-culturally related to other activities, objects and events, or not. It is only in hindsight that we speak of *concepts of music and concepts of art*. So when the ethnomusicologist remarks that "they don't have our concept of music", either implying that they have a *different*

concept of *music* or that they don't have a concept of music at all, the appropriate answer is: how is that relevant? We can't assume without argument that just because people don't share the same concepts then their activities don't have relevant similarities nor share essential properties. Very often, different people abstract different aspects of the same reality and exaggerate the significance of the particular aspects on which they focus, creating the cognitive illusion of a radical, unbridgeable gap between "different concepts of *X*". But in fact, if such different concepts are concepts *of* the same thing at all, then there must be an overarching concept (i.e. a list of properties which we arrive at in hindsight) that includes both (whether or not we explicitly *have it*), no matter how they may differ, since otherwise we have no justification for calling them "concepts *of X*". The concept of H2O and the concept of *the stuff that fills lakes and runs from taps* have the same extension, but they are concepts of the same thing because water is what happens to fill lakes and run from taps. The concept of *water* is the overarching concept that includes both the concept of H2O and the concept of *the stuff that fills lakes and runs from taps*. These "different" concepts are in fact concepts of *different features of the same substance*. As noted by Sainsbury and Tye (2011), "A conception of *water* is a body of information concerning water. There is no such thing as the concept of water (various distinct concepts, like *the* concept *of* H2O and the concept *stuff that falls as rain*, have water as their referent, and so are concepts of water). By contrast, there is such a thing as the unique concept *water*."

Now, what is the overarching concept that binds all (actual or merely possible) culturally relative concepts of a social kind such as art or music? What feature (or features) must any culturally-relative concept of music have, if it is to be a concept of music at all? And what does "culturally relative concept of *X*" exactly mean (where *X* is a social-kind term)? As far as I can see, a culturally-relative concept of *X* is a restrictive concept of *X*, a concept that, in virtue of ignorance or chauvinism, excludes a subset of *X-variants* from its extension. From this I gather that a culturally-relative concept of *X* either collapses into a concept of a particular *X*-variant, or into a concept of a subset of *X*-variants, accompanied by unawareness that these are in fact *variants*, that they are cross-culturally related to other phenomena (think of different cultures unknowingly producing variants of the same board game). But then no such concept could have explanatory power to deal with crosscultural, cross-temporal and cross-modal scenarios where there are enough deep similarities between different things that in a more parochial context would not be considered tokens of the same *X*. Social phenomena can have relevant or deep similarities, even if they originate in different cultural contexts or from the actions of people who don't

share an overarching, crosscultural concept of such phenomena, the most striking examples being that of language and money: different cultures that don't have any concept of language and money can share the properties of having language and having money, and this state of affairs is compatible with their ignoring crucial facts about language and money.

If when thinking about the nature of a social phenomenon demanding explanation, people don't have an overarching, cross-cultural concept of it in mind, then they *should* have it, if they are to think correctly about the subject. There is otherwise no interest in the philosophical project of definition. The point of defining concepts such as *art* and *music* is not just to have a definition that is extensionally adequate, with special emphasis on recalcitrant cases of avant-garde works, as if accommodating such works and taking at face value artist's often hasty and ideologically motivated *statements* about art took precedence over understanding what it is that artists *do* when creating art, what audiences do when appreciating it, and why we came to have any conceptions of art at all. What we want to define, therefore, isn't the concepts we *happen* to *have* but the concepts we *should* have if we are to make sense of how the culturally-relative concepts connect with each other and of the nature of the phenomena in question.

In the philosophy of music in particular, we should be engaged with enquiring whether there is a usable concept of music such that a) it captures a subset of human sound oriented activities (independently of how different cultures divide sound-oriented activities) which b) constitute a cross-cultural, non-arbitrary human phenomenon, a universal human feature, c) that we can use to explain 1) why we have culturally relative conceptions of music at all, 2) what makes them conceptions of *music* instead of something else, and 3) why we are inclined to describe as "music" sound-oriented activities that may originate in cultures that lack "our" culturally-relative concept of music (sup-posing we have one and whatever it is) 4) why (primarily) non linguistic sound-oriented activities from the distant past or from an alien social background can still appeal to some of us, why this appeal seems independent of any procedural or institutional framework, with which we have nothing to do anyway. In other words, the role of a philosophy of music isn't to tell us how we already think about music but how we *should* think about music if we are to understand why we have any culturally-relative concepts of *music* at all. If our theory doesn't do that, then it's not a philosophical enquiry on music.

Social kinds raise complications that natural kinds don't, since social kinds don't exist independently of social beings and their representations of reality. Whereas the nature of things like water, silver and cadmium is mind-independent, it's not obvious, at the very least, that the same is true

of social kinds. We can be wrong about the nature of water and there are empirical discoveries (such as water's chemical structure) that can make us change our views. But what empirical discoveries could we make that would lead us to revise our concepts of art or music? Of course, every time we encounter a new work of art or become acquainted with artworks from a different culture, we learn something about the *extension* of the concept of art. But there is no empirical finding, other than acquaintance with the work itself, from which we learn that those things are art (that is, no artistic parallel to the empirical discovery of the chemical structure of water). We simply *recognize* those objects as art (or we don't). We can learn about the essence of art neither by chemically analyzing artworks nor by any such empirical scrutiny. Even though the recognition of art is a matter of experience, the essence of art must be captured, if only partially, through a *priori* reflection on our experience of what are thought to be central cases of artworks and how we already think about them.

In a particular case, we may have doubts concerning the artwork status of a given object, or we may be unaware that a certain object is a work of art. Someone may then call our attention to the work's aesthetic properties. But how do we know that having aesthetic properties is a part of the essence of art? How do we know whether that is necessary or sufficient for art? And how would we try to disprove such idea? Providing examples of works of art with no aesthetic properties will only be useful if we already have an idea of what a work of art is. Otherwise, how do we know that the proposed counter-example *is* a work of art and therefore a genuine counterexample? Moreover, if the existence of aesthetically dysfunctional or even anti-aesthetic artworks is compatible with an aesthetic theory of art, there is no way we can know that empirically. We have no alternative but to *think* about it. This doesn't mean that social kinds are any less objective or that the essence of a social kind isn't mind-independent (in the same sense that the nature of mind is mind-independent). There is confusion between the mind-dependence of facts about whether a particular thing counts as an instance of a social kind and the mind-dependent nature of the social kind itself. A confusion between something's being contingent upon the existence of social beings and having its nature determined by the subjective states of social beings. This mistake is easily dispelled: beliefs are contingent upon the existence of thinking beings, but what a belief is (what makes it different from other mental states) doesn't depend on the beliefs of thinking beings about the nature of beliefs. The most straightforward analogy with a social kind I can think of is with language: the existence of language is contingent upon the existence of social beings capable of having beliefs about their grunts and squiggles, but what language *is* (how it differs

from other social kinds) doesn't depend on our beliefs (or absence thereof) *about* language. One of the tasks of a philosophy of music is to determine whether, despite superficially appearing to be an arbitrary concept, the case of music turns out to be relevantly similar to the case of language.

NATURAL-KIND THEORIES OF ART AND CULTURAL-KIND THEORIES OF ART

To clarify what I'm aiming at, I'll use a classification of theories of art presented by George Dickie (1997) in his article "Art: Function of Procedure, Nature or Culture?"

In that article, Dickie divides theories of art into *natural-kind theories of art* (NKTA) and *cultural-kind theories of art* (CKTA). These notions will prove to be immensely helpful. Here is how he defines both types:

> NKTA: A natural-kind theory of art would be one in which it is claimed that art first emerged as a result of natural-kind activity and that art has continued to be created as a result of natural-kind behavior [Dickie (1997) p. 26].

> CKTA: The institutional theory of art, in either its earlier or its later version, is clearly a cultural-kind theory because it takes a cultural, institutional structure to be the necessary and sufficient matrix for works of art. [...] For the institu-tional theory, various natural-kind activities may show up in various artworks, but there is no reason to think that any one natural-kind activity is or needs to be present in every artwork [Ibid., pp. 27-28].

By "natural-kind activities" (NKA) and "natural-kind behavior" Dickie means those things that are spontaneously done by living organisms; activi- ties like "gathering food, stalking prey, eating, mating, building nests, constructing the elaborate courtship bowers that birds do, living solitarily and living in social groups" [Ibid., p. 25]. Cultural-kind activities (CKA) and cultural-kind behavior are characterized by not being genetically fixed. They are "particular ways of living together, particular ways of hunting, particular ways of raising food, rituals of eating and marriage" [Ibid.], etc.

There isn't a strict separation between NKAs and CKAs, though not all CKAs are NKAs. The relation is somewhat more complex. "Some cultural- kind activities are particular ways that, in one way or another, human beings have come to organize their natural-kind activities. Such activities are in some sense invented by the members of a particular group and are passed on by learning" [Ibid.].

A natural activity organized in multiple ways not biologically

predetermined is still a natural activity. Human CKAs comprise biologically non rigid activities that may be performed in a biologically rigid (narrowly in nate) way by other species (e.g. mating, stalking prey and gathering food), activities that are discovered, invented, passed on by learning (e.g. writing and the use of fire), and the creation of institutional reality (e.g. counting a line of stones as a territorial boundary, counting wampum shells as money, etc.). It's very important not to confuse natural-kind *activities* with what we usually call *natural kinds*: things like *water*, *silver*, and *willow tree*, things that are independent of any mental states or conscious activity. A natural-kind theory of art isn't a theory according to which *art* is a natural-kind in this sense. A natural-kind theory of art is a theory according to which the activity type *art-making* is a cross-cultural, non-arbitrary human phenomenon, independent of any *art concepts* that people may form or acquire (the same way *language* is independent of a language-concept and *depiction* is independent of a depiction concept).

By "conceptual dependence" I mean the property in virtue of which the fact that some object X counts as Y is dependent upon X's being *conceived* or *described* as Y. A classic example of conceptual dependence, given by Nelson Goodman (1983), is that of configurations of stars as constellations. A configuration of stars is only a constellation from the viewpoint of an earthly observer and under a shared description (the fact that a certain stellar configuration counts as the constellation of Orion the Hunter is also a *social* fact). Facts about what configurations of stars count as constellations are conceptually dependent facts. Though Goodman was making a case for a kind of constructivism (the belief that all facts are conceptually dependent), we don't have to embrace constructivism to accommodate conceptual relativity as a real phenomenon, since conceptual relativity is perfectly consistent with realism. Some facts can be conceptually dependant only because not all facts are. For there to be conceptually dependent facts such as the fact that X counts as constellation Y there must be conceptually independent facts: the fact that there are configurations of stars, the fact that some of these are visible to earthly observers as describing certain forms, the fact that there are earthly observers endowed with imagination (the ability to see hunters or giants in arbitrary stellar configurations) and capable of having shared beliefs, etc.

Some CKAs are NKAs but not all are. NKAs that are biologically rigid are not CKAs. CKAs that are conceptually dependant are not NKAs, though they are partly constituted by NKAs. CKAs that are conceptually independent are cross-cultural phenomena.

The notion of conceptual dependence allows us to make a relevant distinction between CKAs: those whose individuative properties include the property of being represented as the activity-type they are, and those that are individuated merely by their constitutive NKAs, independently of being thought under any description. Another way to put this is to say that

CKAs that are conceptually independent are *human universals*, that is, cross- cultural, non-arbitrary, biologically non-rigid human phenomena. Roughly, we have CKAs that are conceptually dependent and CKAs that are not conceptually dependent. What characterizes the former is that they involve at least one biologically non-rigid NKA – *language*, without which no object X can count as any Y in whatever context. Conceptually dependant CKAs are those that essentially involve the act of counting some X as some Y in a context.

These conceptual relations can be represented in the following diagram:

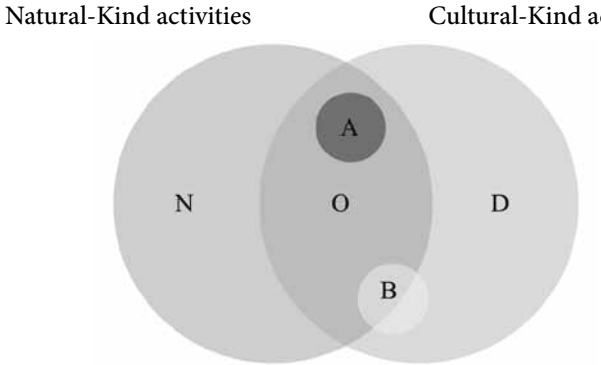

Fig. 1

N = Biologically rigid natural-kind activities.
D = Counting somo object (X) as embodying a finction (Y) in a context (C).
O = Cross-cultiral, biologically non-rigid activity-kind theory of art.
A = Activities cluster for art, according to a natural-kind theory of art.
B = Activities cluster for art, according to a cultural-kind theory of art.

We can illustrate this with a few examples: *speaking English* is like B in the above diagram. It satisfies two important conditions: 1) though speaking, like all our activities, has biological constraints, it is not biologically rigid, in the way that the cries and calls of many non human animals are. 2) There are no facts about what grunts and squiggles count as English utterances independently of there being a concept of *English*. In a world where people have no shared beliefs about what grunts and squiggles count as utterances of English, there are no utterances of English.

Satisfying condition 1 (lacking biological rigidity) is both necessary and sufficient for a particular NKA to be a CKA. Satisfying condition 2 (being conceptually dependent) is sufficient but not necessary to be

a CKA. CKAs that satisfy condition 2 are partly constituted by a NKA or cluster of NKAs, though no cluster of such activities is sufficient to individuate them as the activity type they are. They must include shared beliefs or representations about their component activities, for instance, the shared belief that these amount to *speaking English*. (Of course, a subgroup could speak English without being aware that's what they're doing.) Facts about what grunts and squiggles count as utterances of English are thus analogous to facts about what configurations of stars count as constellations. There is at least one NKA of which all activity-types located in the D area of the diagram in Fig. 1 depend: language. This is because all activity-types located in D have the same basic structure: *counting some object (X) as em bodying a function (Y) in a context (C)*, which is at bottom a linguistic operation. (Searle:1995; 1999; 2010) Counting grunts as utterances, counting pieces of metal as money, counting certain utterances as promises, counting certain graphic patterns as national flags, and so on. In other words, the D area in the diagram is where the creation of social and institutional reality is, the most basic institutional fact being that of language itself. In fact, the whole of D area should be seen as an "outgrowth" of the O area, specifically of our linguistic abilities. Combining this analysis with Dickie's classification of theories of art in NKTA and CKTA enables us to see how the metaphysics of society illuminates the metaphysics of art in general and music in particular, and what kind of theory of music's nature will have the most explanatory power.

Now, consider the type *imposing syntactic structure on physical events* (such as grunts and squiggles), which is a non-rigid NKA. We don't need a concept of syntax in order to divide grunts and squiggles into discrete, repeatable units that preserve their identity and perform different functions in different contexts and in order to perceive separate grunts and squiggles as tokens of the same type (for instance, in this article there are exactly 26 to kens of the word type *type* and 10 tokens of the word-type *token*, all of them separate spatiotemporal realities). The only thing required is that the relevant abilities are in place, that is, we need to have the right kind of brain. Doing this amounts to having linguistic behavior, without a *conception* or *description* of that behavior being necessary for the activity-type to count as *imposing syntactic structure on physical events*. In fact, we would have no descriptions and no articulate conceptions of things whatsoever if it weren't for this ability.

CKTA are not characterized by the trivial assertion that activities like art and music are cultural activities (they obviously are not biologically rigid NKAs and no plausible NKTA would assert they are) but the non-trivial assertion that no cluster of NKAs is sufficient for art. For a CKTA, a shared *conception* or *description* of the relevant NKAs *as art* (or as music), provided by a cultural or institutional matrix, is both necessary and sufficient for art (and music). The paradigmatic CKTA here is given to us by Danto:

> It is the role of artistic theories, these days as always, to make the artworld, and art, possible. It would, I should think, never have occurred to the painters of Lascaux that they were producing *art* on those walls. Not unless there were neolithic aestheticians [Danto (1964), p. 58].

According to this view, it's not something intrinsic to the activity of cave painters that makes their paintings art, but the separate cultural activity of counting things as art (which is embodied in a "frame": be it a procedure or an institutional background), whereas for a NKTA it is something intrinsic to the activity (but *not* to the objects produced by that activity) that makes the products of such activity artworks. Artworks, according to NKTA, have *functional essences*: no arrangement of physical stuff or concatenation of sounds is an artwork or a musical work in virtue of intrinsic properties (though the relevant properties may depend on some of the object's intrinsic properties), but because it has certain functions in virtue of a causal history that traces back to human intentional states. An example of this are aesthetic theories of art for which the essence of art lies in the intentional realization of aesthetic properties in artifacts [Zangwill (2007)]. For NKTA, the transition from a world without art to a world with art is achieved simply when cognitive agents strive to realize aesthetic properties by producing objects with the appropriate non-aesthetic properties on which the relevant aesthetic properties depend. In so far as the type *intentional exploration of aesthetic properties* doesn't require that cognitive agents have a *concept of the aesthetic* or a *concept of aesthetic properties*, the individuation of the type *artistic creation* requires only the resources of a NKTA. The fact that cave painters weren't *aware* that they were creating artworks in virtue of the absence of such a concept is no more relevant to the existence of art than not having a concept of *language* is relevant to having language.

For CKTA, the transition from a world without art to a world with art is achieved by institutional reality (whether or not it involves an "artworld"): shared representations about what counts as "art" and about the appropriate context in which countings of things as art are successful or felicitous (e.g. maybe John Cage can make it the case that gurgling carrot juice counts as music but I can't). Here arthood is also characterized functionally but the functions in question are of a whole different sort. This may sound odd, given the traditional contrast between functionalist theories of art and institutional theories of art, where functionalist theories belong in the NKTA side of the divide. To make sense of this we need a general characterization of functions.

FUNCTIONS, ARTIFACTS AND INSTITUTIONS

In *The Construction of Social Reality*, Searle (on whose analysis of the nature of institutions I draw here, as in the previous section) offers a taxonomy of social facts, including the assignment of functions. For our present purpose, I need only focus on two kinds of function assignment: *causal agentive functions* and *status-functions*. (Nothing here hinges on Searle's being right (e.g. against Millikan) about functions in nature generally.) Causal agentive functions and status-functions are both kinds of agentive function, that is, functions an object has in virtue of being *intended* to have them (they contrast with *non-agentive* functions, such as the heart's function of pumping blood, which it performs independently of anyone's intentions). Examples of causal agentive function are the functions of artifacts in general, such as the function of being a screwdriver. An artifact has its function in virtue of having the right physical powers (like the power to screw in other things) *and* in virtue of being intended to have that function. However, in order to discharge their functions, artifacts depend solely on their physical structure, not on shared representations about them (there need not be some agreement about screwdriver status for something to be a screwdriver, all that is needed is the intention to screw in things using the appropriate physical structure). Causal agentive functions are not language dependent. Yet, no arrangement of physical stuff is an artifact if no one *intended* it to have a certain function.

Status-functions are functions no object can perform in virtue of its physical structure alone. No arrangement of physical stuff is a territorial border unless it's collectively represented as a territorial border, no matter how physically effective it is in keeping people from crossing it. But even a line of stones with no physical power to keep people from crossing it can be a territorial border if it's collectively represented as a territorial border. And the fact that it is so represented causally impacts people's behavior. These functions differ from causal agentive functions in that they are language dependent. According to Searle, the basic structure of all status-functions is *X counts as Y in C*, i.e., some object (X) counts as embodying a function (Y) in a context (C). Counting X's as Y, through shared representation (what Searle calls *collective intentionality*), creates and assigns *power*, generating properties of X's that they can't have in virtue of physical structure alone. Status-functions are the backbone of all institutional reality.

We can now see more clearly the difference between traditional functionalist theories of art and institutional/proceduralist theories of art, that is, between theories that belong in the NKTA group and theories that belong in the CKTA group: while the former appeal to causal agentive functions, of the same kind involved in the individuation of artifacts generally, the latter appeal to a status-function. This is roughly how a cultural-kind theorist sees artwork status: *the snow shovel (X) counts*

as an artwork (Y) in the twentieth century artworld (C). And the same structure applies to music: *gurgling carrot juice (X) counts as music (Y) in the twentieth century artworld (C)*.

It's not the appeal to functions that distinguishes both families of theories, but the *kind* of functions appealed to, the former being language independent and the latter language dependent. It's just that traditionally in the philosophy of art, "functionalist" is a term reserved for theories that appeal to causal agentive functions of a specific kind, namely, *aesthetic* functions: the intention to realize an aesthetic property via certain non-aesthetic properties is analogous to the intention to screw in things using the appropriate physical structure; no linguistic articulation of the aesthetic property is needed, only the intention to produce a certain *kind* of experience *of* non- aesthetic properties. Of course, these intentions will become more complex and will integrate conventional aspects which are language-dependent. But at rock-bottom they're not language-dependent.

Now, status functions have an interesting property that will raise serious difficulties for the institutionalist about art, namely, *self-referentiality*. Consider the case of money: it's part of the definition of money "to be represented as money", since nothing can be money in virtue of its physical structure alone. This seems like a vicious regress, but actually it isn't:

> The word "money" marks one node in a whole network of practices, the practices of owning, buying, selling, earning, paying for services, paying off debts, etc. As long as the object is regarded as having that role in the practices, we do not actually need the word "money" in the definition of money, so there is no circularity or infinite regress. The word "money" functions as a placeholder for the linguistic articulation of all these practices. To believe something is money, one does not actually need the word "money". It is sufficient that one believes that the entities in question are media of exchange, repositories of value, payment for debts, salaries for services rendered, etc. And what goes for money goes for other institutional notions such as marriage, property, and speech acts such as promising, stating, ordering, etc [Searle (1995), p.52]

The self-referentiality of status-functions tells us something of key significance: although the existence in a given society of objects that embody a particular status-function depends on the activities of cognitive agents, the *nature* of that status-function isn't arbitrary and it doesn't depend on any beliefs people have about the nature of status-functions (two points: a) *what it is to be a status-function* is mind-independent, b) *what individuates a status function from other status-functions* is mind-independent). There can only be meaning in counting something as money because the nature of money is already established mind-independently by whatever it is to perform that role in that network of practices. There are objective restrictions (logical and metaphysical) on what can *function* as money once the status-function is assigned. The status-function itself

is only intelligible *because* of those objective restrictions. "It's money because I say so" means *nothing* in the absence of those mind-independent restrictions.

Likewise, if *art* and *music* really are status-functions, then "being represented as art" is part of the definition of art, and "being represented as music" is part of the definition of music (which is what the institutional theorist claims). But if they truly are status-functions then the terms "art" and "music", as they occur in the *definientia*, must be placeholders for the linguistic articulation of practices, relations and causal roles involved in the individuation of those specific status-functions. *Status-functions are not individuated from other status-functions by their linguistic descriptions*. So, if all we have to go is a linguistic description (or a procedural "frame"), we don't have a status-function, we have only the general form of a status-function.

This argument is different from traditional objections of circularity raised against the institutional theory. According to these objections, circularity shows that the concept of art can't be an institutional concept. And a conventional response is to deny that circularity poses a problem, based on the idea that artworld institutions can't be individuated from other institutions in an informative, non-circular way. By contrast, the objection I'm raising here, based on Searle's analysis of institutional facts, is that we can very well have (and we do have) an institutional concept of art, but that concept *presupposes* a more basic functional concept. Language allows us to assign functions no object can perform in virtue of its physical structure alone. No mere line of stones can physically keep people from crossing it, but it effectively functions as a territorial border if represented as such. We can have a stone wall, which is primarily an artifact with a causal agentive function of keeping people out (or in) but if people collectively represent it as a territorial border it impacts their behavior even if through time it's reduced to a mere line of stones. The same relation holds between a particular good that in an barter economy functions as a *de facto* medium of exchange, and an object that embodies the money status-function. Language extends power "at will", but *what* that power *is* isn't decided by us, any more than our use of concepts fixes the ontology of concepts. All institutional kinds, though immensely flexible and multiply realizable, bottom out in a network of practices, relations and causal roles. The conclusion this argument aims at isn't to remove the institutional concept of art but to say that any such concept presupposes a more basic explanation. At best, a CKTA collapses into a NKTA enhanced with an explanation of how the basic functional roles essential to central cases of art-works can be extended through language, analogously to the roles of the stone wall and the *de facto* medium of exchange in a barter economy.

This idea provides us a useful tool to think about recalcitrant cases of avant-garde art and indiscernible duplicates, so hastily taken to "refute"

more traditional (functionalist) aesthetic theories of art. The same goes for recalcitrant cases of music, such as silent pieces, *musique concrète*, pieces with no temporal structure, or that lack "basic musical properties" such as rhythm, melody or harmony. Supposing a basic functionalist account of art in terms of the intentional realization of aesthetic properties, and a basic functionalist account of music in terms of the intentional realization of rhythmic, melodic or harmonic properties (to be developed elsewhere), what we need is an explanation of how, in the context of both a NKTA and a NKTM (natural-kind theory of music), our ability to create institutional reality from linguistic operations widens the scope of objects that can belong to a particular artistic or musical tradition, e.g. *found objects* and *found sounds*, even if no possible or actual tradition could be entirely constituted by objects of that kind. To give an example used by Kania (2011), pp. 8-9: there can be blank canvases in a tradition of painting even if no tradition of blank canvases could ever be a tradition of painting. I think an enhanced NKTA could explain exactly *why* and *how* this comes to be. The institutionalist, on the other hand, takes it for granted and moves on from there, since he takes the procedural "framing" (the imposing of a status-function) to be the explanatory mechanism, and not a part of the *explanandum*.)

ENHANCING THE NKTA

Though only a NKTA gives a rock-bottom explanation of the existence of art and music, I believe that a strict disjunction between a natural-kind explanation and a cultural-kind explanation won't afford us the explanatory power required by a thoroughgoing metaphysics of art in general and music in particular. This is for the simple reason that, as language users, we can't help creating institutional reality out of our biologically non-rigid NKAs: "Given a language you can, so to speak, create institutional facts at will (that is the top-down part); but when you have a language, other social institutions will inevitably grow up out of language (this is the bottom-up part)" [Searle (2010), p. 63]. Even if a cluster of biologically non-rigid NKAs is sufficient to individuate the activity-types *music* and *art*, that is, even if at rock-bottom, art and music are conceptually independent natural-kind activities, as language users we will inevitably have institutional reality of an artistic and musical ilk; we will form concepts of *art* and *music* and we will inevitably extend those concepts beyond their original domain of application, with meaningful artistic and musical consequences (objects with no apparent aesthetic functions can be "secondary works", they can derive their art-status from the property of being *about* works that have recognizable aesthetic functions [Zangwil (2007), p. 70]. And we will have these institutional extensions of our natural abilities even in the absence

of any explicit words for "art" and "music". No complete philosophical theory of either art or music can leave out that portion of reality.

This means that the concepts of art and music we *seek* (not those we have) will both be two-layered concepts: they will have an element of "rock- bottom functionalism" (with causal agentive functions doing the explanatory work) and an element of status-function, explaining how the power of language to create institutions widens the scope of objects capable of art-status, in a way that renders such objects intelligible as members of an artistic tradition.

But what would a natural-kind theory of music look like? What activity-types would constitute a sufficient cluster for the (conceptually independent, cross-cultural) existence of music? Here is a rough (non-exhaustive) list of NKAs that might be included:

a. Dividing the pitch continuum into discrete, repeatable pitches and identifying separate tokens of the same pitch-type as "the same again".

b. Perceiving sounds an octave apart as "the same but higher" or "the same but lower".

c. Organizing pitches into divisions of the octave called "scales", on which melodies are based.

d. Perceiving certain beats in a grouping of beats as unaccented relatively to an accented beat in the same grouping.

e. Perceiving separate tokens of the same accented-unaccented beat pattern as "the same again".

f. Perceiving certain simultaneously-sounding pitch-aggregates as a "single entity" (chords) and identifying separate tokens of the same type as "the same again".

g. Imposing different syntactic functions on tokens of the same pitch-type or pitch-aggregate-type, according to their context (pitches pre- ceding and following it – tokens of the same pitch-type or pitch-aggregate-type sound consonant or dissonant according to context and can perform many different functions).

h. Forming auditory expectations.

i. Imaginatively perceiving movement in a sequence of sounds.

j. Recognizing "contour similarities" or isomorphisms between tonal movement and extramusical processes (e.g. the process of undergoing

an emotion, or a fluttering movement).

All items in the above list are prospective "musical universals", that is, types of mental activity that constitute the type *listening to sounds as music* and underlie any actual or counterfactual musical tradition, though no particular musical tradition needs to deploy *all* of them (a particular musical culture may lack what we call "harmony" or it may be constituted entirely by drones, drumming or rhythmic yelps). [For more on universals in music, see Brown & Jordania (2011); Stevens & Byron (2009); Patel (2008); Nettl: (2000), (2005)] As Searle has remarked concerning speech acts, the possibility that a particular tribe doesn't have promises is as relevant for a taxonomy of speech acts as the inexistence of tigers in the South Pole is relevant for a taxonomy of animal types [Searle (2006)]. Likewise, the fact that a particular culture lacks one or more items of the NKA cluster for music is of no metaphysical significance, no matter how interesting in other respects. As long as enough items in the list are present, there is still *music* in that culture; and should music be totally absent from a particular culture, that would be an interesting fact but it wouldn't dislodge a natural-kind theory of music, since a natural-kind theory is compatible with the idea that music, like reading and writing, is an invention that builds on existing brain functions and not a biological adaptation [Patel (2010)], though it is also arguably a very ancient and universal phenomenon (the oldest known artifacts capable of producing pitches are bone flutes dating back 35 000 to 40 000 years). A much more recent invention, chess, also builds on cognitive abilities that weren't naturally selected *for* chess. A particular culture's not having chess would be distressing if it were impossible, say, to teach a ten year old in that culture how to play chess. But it is possible, because the cognitive abilities on which chess depends are universal. Likewise, the ability to perform and appreciate music is all but restricted by parochial contingencies. The absence of anything remotely recognizable as music in a peculiar culture, were it to occur, would be no more significant than some cultures not having written language or chess, as long as it remained a natural possibility, say, to teach a ten year old in that culture to play a musical instrument. One would expect the ease with which cultural phenomena disseminate beyond their initial geographic boundaries (think of Anglo-Saxon pop music, movies, and videogames, but also, of course, writing, chess and phenomena such as the establishment of a *lingua franca* between linguistically separate communities) to temper enthusiasm in cultural particularism. The universality of a human feature need not rest upon narrow innateness or direct biological adaptation.

CONCLUSION

I said that the project of definition in the philosophy of art (and music) should aim at explanatory power and not just extensional adequacy or consistency with the way people *actually* think about the subject. I've been discussing the methodology of the philosophy or art and music rather than arguing for a particular definition of music. Though I can't pursue that task here, I don't wish to leave the reader without some suggestion of how we can pursue an enhanced NKTM and some material to think this question through.

As I've stated earlier, the extant philosophical literature on the definition of music is scarce. I don't have the space to engage in a full discussion of the more recently proposed definitions, but I want to say a few words about the definition that so far seems to me the most plausible and compatible with my proposal of an enhanced NKTM (by this I'm not implying, of course, that it is compatible *only* with my proposal). This is the disjunctive definition presented by Andrew Kania (2011), p. 12.

> Music is (1) any event intentionally produced or organized (2) to be heard, and (3) *either* (a) to have some basic musical feature, such as pitch or rhythm, *or* (b) to be listened to for such features.

A disjunctive definition fits well with the two-layered structure I proposed above: the first disjunct (condition *a*) will be explained in "rock-bottom functionalist" terms – a natural-kind theory that accounts for the special character of rhythmic, melodic and harmonic properties. This could be done in terms of "metaphorical perception" [Scruton (1983), (1997), (2009), Peacocke: 2009) or in a way that dispenses with aesthetic metaphors (Budd: (2003); Trivedi (2008), (2011), such as a theory of resemblance plus (spontaneous) imagination. These are theories that seek to explain what it is ("at the foundational level") to *perceive a sequence of sounds as music* in terms of "basic" musical properties or "musical fields of force" to use Scruton's expression. This explanation of why rhythmic, melodic and/or harmonic properties are essential to (central cases of) music will eliminate the apparent circularity in the definition above. I have no space here to engage in a discussion of such views, but the relevant point is that they both appeal to biologically non-rigid NKAs (some form of imaginative perception) and so are equally appropriate for a rock-bottom functionalist explanation of the first disjunct of our definition. They are equally compatible with the idea that music is defined by relational properties whose non-arbitrary clustering is guaranteed by two key features of our cognitive architecture: a) the ability to impose syntactic structure on physical events, such as sequences of sounds and squiggles on a surface and b) the ability to imaginatively explore isomorphisms between domains.

The second disjunct (condition *b*) expresses the element of status-function in musical works. An event which has none of the properties essential to central cases of music can nevertheless have music status in virtue of a certain *aboutness* relating it to basic musical properties. This isn't to say that such works have no *aesthetic* functions. They merely lack basic musical properties. But sounds can have aesthetic properties in the absence of basic musical properties, the same way the literary description of a sunset can have aesthetic properties which will differ from the aesthetic properties possessed by the painting of a sunset.

This definition also leaves room for works of sound art that a) have aesthetic properties, b) don't have basic musical properties and c) are not music because they were not intended to be listened to for basic musical properties. This way, the concept *sound art* subsumes: 1) central cases of music, 2) derivative musical works, 3) non-musical artistic sound works, which include the arts of language (spoken poetry and narrative, drama), and things like *soundscapes* and all sorts of sound design involving aesthetic properties. It also leaves room for examples of non-artistic music, thus providing a neat classification of all possible and actual sound-oriented activities with social functions that may or may not be primarily aesthetic.

There are fears that disjunctive definitions are *ad hoc*. The fear is that once we accept two disjunctively sufficient conditions for being *X* there is no reason we can't keep on adding disjuncts until we have a perfectly gerrymandered concept, whose referents have no common nature. So, if we have two disjunct conditions for art, we'll soon have a thousand disjuncts, and so on, to infinity. Setting aside the slippery-slope argument – to which Denis Dutton (2006), p. 375 has given a sharp reply: "A thousand or more ways of being art is still a long distance from an infinite number of ways to be art" – a disjunctive definition might still be *ad hoc*. I don't think this is the case here though. The disjunction mirrors our twofold ability to impose functions on objects: functions they perform in virtue of physical structure (causal agentive functions), and functions they perform in virtue of collective representation (status-functions). We can have complementary theories of both art and music in terms of functional essences: Central cases of musical events issue from the intentional realization of rhythmic, melodic or harmonic properties *via* the realization of certain sonic properties. Central cases of artworks issue from the intentional realization of aesthetic properties *via* the realization of certain non-aesthetic properties. Central cases of musical artworks issue from the intentional realization of aesthetic properties *via* the realization of rhythmic, melodic or harmonic properties (I leave open whether the latter properties can be aesthetically neutral). The element of status-function is not present only in recalcitrant cases (in the absence of basic musical properties), but is more or less ubiquitous. Consider the type *composing a string quartet in the classical Western tradition* or the type *performing in the tradition of the Persian radif*. Analogously to the type *speaking English*, this

is a conceptually dependent piece of reality, since no purely (intrinsic) sonic facts establish what counts as a *string quartet*, a *radif*, or any other conventional musical structure. Hence the appeal to status-functions isn't an *ad hoc* device to forcibly fit recalcitrant cases into our theory, since central cases are also embedded in institutional reality. What an enhanced NKTA / NKTM calls into question isn't the pervasiveness of institutional reality in concrete artistic practices, but the key proposition of cultural-kind theories, that the element of status-function is definitionally basic, that it is the mechanism by which art and music come into existence.*

NOTES

* For helpful suggestions, comments and criticisms, I wish to thank Aires Almeida, Andrew Kania, Joao Alberto Pinto, Nick Zangwill and the anonymous reviewers of this paper. Any mistakes or misconceptions it may have are my sole responsibility.

REFERENCES

BROWN, S. & JORDANIA, J., (2011), "Universals in the World's Musics", *Psychology of Music*, SAGE, pp. 1-20.

BUDD, M., (2008), "Musical Movement and Aesthetic Metaphors", in *Aesthetic Es- says*, New York, Oxford University Press, pp. 154-170.

— (2008a) "Understanding Music", in *Aesthetic Essays*, New York, Oxford University Press, pp. 122-141.

CARROLL, N. (ed.) (2000), *Theories of Art Today*, Wisconsin, University of Wiscon- sin Press.

CURRIE, G., (2000), "A Note on Art and Historical Concepts", *British Journal of Aesthetics*, Vol. 40, No. 1, pp. 186-190.

— (2009) "Art for Art's Sake in the Old Stone Age", *Postgraduate Journal of Aes- thetics*, Vol. 6, No. 1, pp. 1-23.

DANTO, A. (1964), "The Artworld", *The Journal of Philosophy*, Vol. 61, No. 19, pp. 571-584.

DAVIES, S. (1990), "Functional and Procedural Definitions of Art", *Journal of Aes- thetic Education*, Vol. 24, No. 2, pp. 99-106.

— (1997), "First Art and Art's Definition", *The Southern Journal of Philosophy* , Vol. 35, pp. 19-34.

— (2000), "Non-Western Art and Art's Definition", in Carrol, N. (ed.) (2000), pp. 199-216.

— (2011), *Musical Understandings*, New York, Oxford University Press.

DICKIE, G., (1997), "Art: Function or Procedure, Nature or Culture?", *The Journal of Aesthetics and Art Criticism*, Vol. 55, No. 1, pp. 19-28.

— (2000) "The Institutional Theory of Art", in Carroll, N. (ed.) (2000), pp. 93-108. DODD, J. (2012), "Defending the Discovery Model in the Ontology of Art: A Reply to Amie Thomasson on the *Qua* Problem", *British Journal of Aesthetics*, Vol. 52, No. 1, pp. 75-95.

DUTTON, D. (2000), "But They Don't Have Our Concept of Art", in *Theories of Art Today*, in Carroll, N. (ed.); pp. 207-240.

— (2006), "A Naturalist Definition of Art", *Journal of Aesthetics and Art Criticism*, Vol. 64, No. 3, pp. 367-377.

GAUT, B., (2000), "Art as a Cluster Concept", in Carroll, N. (ed.), pp. 25-44. GOODMAN, N. (1983), "Notes on the Well-Made World", *Erkenntnis*, Vol. 19, Nos 1-3, pp. 99-107.

HAMILTON, A., (2007), "The Concept of Music", in *Aesthetics and Music* (Chap. 2), New York, Continuum.

KANIA, A., (2010) "Silent Music", *Journal of Aesthetics and Art Criticism* , Vol. 68, No. 4, pp. 343-353.

— (2011), "Definition", in *The Routledge Com panion to Philosophy and Music* , New York, Taylor & Francis, pp. 3-13.

LEVINSON, J., (2011), "The Concept of Music", in *Music, Art and Metaphysics*, New York, Oxford University Press, 2011, pp. 267-278.

MCIVER-LOPES, D. (2007), "Art without 'Art'", *British Journal of*

Aesthetics¸ Vol. 47, No. 1, pp. 1-15.

MESKIN, A. (2009), "From Defining Art to Defining the Individual Arts: The Role of Theory in the Philosophies of Arts", in Thomson-Jones, K. & Stock, K. (eds.), pp. 125-149.

NETTL, B. (2001), "An Ethnomusicologist Contemplates Universals in Musical Sound and Musical Culture", in Wallin, N, Merker, B. & Brown, S. (eds.), *The Origins of Music*, Massachusetts, MIT Press, pp. 463-472.

— (2005) , *The Study of Ethnomusicology – 3 1 Issues and Concepts*, Chicago, Uni- versity of Illinois Press.

PATEL, A. (2005), "Music, Biological Evolution and the Brain" in M. Bailar et al. (ed.), *Emerging Disciplines*, Houston, Rice University Press, pp. 91-144.

— (2008), *Music, Language and the Brain*, New York, Oxford University Press. PEACOCKE, C. (200), "The Perception of Music – Sources of Significance", *British Journal of Aesthetics*, Vol. 49, No. 3, pp. 91-144.

SAINSBURY, R. & TYE, M. (2011), "An Originalist Theory of Concepts", *Aristotelian Society Supplementary Volume*, Vol. 85, No. 1, pp. 101-124.

SCRUTON, R. (1983), "Understanding Music", in *The Aesthetic Understanding: Essays in the Philosophy of Art and Culture*, New Yotk, Taylor & Francis, pp. 77-100.

— (1997), *The Aesthetics of Music*, New York, Oxford University Press.

— (2009), *Understanding Music – Philosophy a nd Interpretation*, Continuum Inter- national Publishing Group.

SEARLE, J. (1995), *The Construction of Social Reality*, New York, The Free Press.

— (1999) , *Mind, Language and Society - Philosophy in the Real World* , New York, Basic Books.

— (2006), "Social Ontology: Some Basic Principles", *Anthropological Theory*, Vol. 6, No. 1, pp. 12-29.

— (2010) , *Making the Social World*, New York, Oxford University Press. STACHELHAUS, H. (1991), *Joseph Beuys*, New York, Abbeville Press.

STECKER, R. (1986), "The End of an Institutional Theory of Art", *British Journal of Aesthetics*, Vol. 26, No. 2, pp. 134-42.

— (2000), "Is It Reasonable to Attempt to Define Art?", in Carroll (ed.), pp. 45-64.

— (1990), "The Boundaries of Art", *British Journal of Aesthet ics*, Vol. 30, No. 3, pp. 266-272.

— (1997), *Artworks – Definition, Meaning, Va lue*, Pennsylvania State University Press.

STEVENS, C. & BYRON, T. (2009), "Universals in Music Processing", in Hallam, S., Cross, I. & Taut, M. H. (eds.), *The Oxford Han dbook of Music Psychology*, New York, Oxford University Press, pp. 14-23.

THOMSON-JONES, K. & STOCK, K. (eds.) (2009), *New Waves in Aesthetics*, New York, Palgrave-Macmillan.

TRIVEDI, S., "Metaphors and Musical Expressiveness", in Thomson-Jones & Stock, (eds.) (2009), pp. 41-57.

— (2011) , "Music and Imagination", in *The Routledge Companion to Philosoph y and Music*, Taylor & Francis, pp. 113-122.

VARÈSE, E. & CHOU WEN-CHUNG (1996), "The Liberation of Sound", *Perspectives of New Music*, Vol. 5, No. 1, pp. 11-19.

ZANGWILL, N. (2007) , *Aesthetic Creation*, New York, Oxford University Press.

ZAPPA, F. & OCCHIOGROSSO (1990), P., *The Real Frank Zappa Book* , New York, Simon & Schuster.

SHUT UP AND LISTEN
RULES, OBSTACLES AND TOOLS OF PHILOSOPHICAL DIALOGUE

Tomás Magalhães Carneiro

> *Sapere Aude! [Dare to be wise!],*
> *It is so convenient to be immature.*
> *If I have a book to have understanding in place of me, a spiritual adviser to have a conscience for me (...) I need not make any efforts at all. I need not think, so long as I can pay; others will soon enough take the tiresome job over me.*
>
> *Immanuel Kant, "An answer to the question: What is Enlightement?"*

1 - DARE TO THINK

It is our firm belief that every philosophy teacher accepts this other well intended kantian motto. "You will not learn from me philosophy, but how to philosophize, not thoughts to repeat, but how to think." Besides from not knowing what exactly did Immanuel Kant do to put into practice this general pedagogical program this motto also has the problem that from being repeated to exhaustion it has become nothing more than an educational truism with almost no practical realization in our common philosophy classes.

But if its true that almost every philosophy teacher shares with Kant his noble intentions it´s fair to ask why is it so hard to move from them to more concrete pedagogical practices that turns "ON" our students thinking-switch and forces them to leave their current state of intellectual

anesthesia and total lack of "guts to think" by themselves? Why is that systematically those same teachers take has "good thinking" the mere repetition of ideas of others, the conformation to the syllabus and the accordance to their own will as teachers? These are the consequences of our educational system being totally centered in teaching our students "what to think" instead of "how to think". To change this situation we have to change our educational paradigm from a "teacher centered" approach to a "student centered approach". In this paper our proposal is that this can be done by allowing our students a safe place to talk to each other in the philosophy classroom. By "talk to each other" we don´t mean a kind of a chat between friends or the typical classroom debate that most of the times is just an organized way to make some noise or, at best, an alternative way for students to express themselves and for the teacher to evaluate them. What we aim in this paper is for a space for dialogue and critical thinking in the classroom and that is what we mean by "safe-place to talk", a structured and monitored type of conversation that aims to be, at the same time, rich and rigorous, creative and critical, serious and playful. By "talk to each other in the philosophy classroom" we mean to learn to philosophize trough Philosophical Dialogue with the guidance of a teacher who works no longer as a teacher but as a referee and a coach that allows his students to play a sort of a game called the "game of philosophy" and in doing so he strives to improve the way they think by improving the way they have meaningful conversations with each other, the way they dialogue in a philosophical manner.

Laziness and cowardice are the reasons why a large proportion of men (…) gladly remain immature for life (Kant, op. cit.). We once again read these words from Immanuel Kant, look around us and see that this "laziness and cowardice" are two predicates that we so often tend to associate with our students that we wonder if these predicates are not part of our students genetic pool, if they are not fundamental traits of their *student nature*.

We believe that this is hardly the case and that a simple visit to your local kindergarten will prove us right. Four and five years old children, who have just enrolled in the education industry generally show us the opposite of this poor intellectual panorama. They are usually brave enough to compromise themselves with there own thoughts and most of them are happy to raise their fingers in the air to try and test some new hypothesis about some wonderful new problem that just came up. *Laziness* and *cowardice* are not common predicates in four and five year olds students.

We continue our school-tour and visit other classrooms from primary to secondary education and notice that the fingers in the air start to rarify. If we pay attention to what happens inside these classrooms we see that increasingly less time is given for students ideas, problems and arguments and more time is given to teacher's ideas, problems and arguments. As we approach the secondary level we notice that the student's job becomes a simple task of learning and sometimes guessing the knowledge the teacher possesses and repeating it in the exams.

Conforming to authority and learning other peoples ideas, problems and arguments for almost twenty years is, in our view, a real but dreadful and boring site and it´s no wonder that in secondary and university level classrooms you can hardly find challenging and adventurous fingers in the air trying to "air out" some new and brilliant idea that "just came up".

Being a "Philosophy with Children" teacher working in several schools from pre-school to secondary education this impoverishment of our students critical and creative horizon is something that we sadly noticed too many times for the past several years. And for us this is a clear sign that we have to blame our educational system and the actual work of educators, teachers and parents that are more or less consciously engaged with that educational system, A system that promotes this *laziness and cowardice* in our students mindset and where conformism to authority is the name of the game. This "conformism to authority" can, and in fact does, lead to our students acritical dogmatism where the word of the teacher is the "absolute truth" but it can also lead to an acritical relativism where the different words of different teachers are all "absolut truths" even if they are contrary to each other. Actually this lazy attitude of "anything goes" is all too common to the average student that is incapable of thinking by himself, and this is the opposite attitude than the one teachers should value in a philosophy student for whom "socratic skepticism" should be the name of the game.

We are sure that no philosophy teacher believes that he should be teaching students to become dogmatic, acritical, in a word, *asocratic* human beings and we are sure that every philosophy teacher believes he is educating *socratic students* who strive to see beyond the limits of dogma and are comfortable with the fundamental uncertainty of our philosophy intellectual tradition. But a sort of a performative contradition arises when we see that, despite the efforts and intentions of our good teachers most of our students prize dogmas and secure bites of knowledge above uncertainty and problematic questions. And we suspect that this happens because he is thaught from an early age that uncertainty and problems are not going to

take him very far in the path of good grades and academic curriculum, the only criteria of excellence generally accepted.

But here is important to ask where can we find the real spirit of our signature, i.e., the essence of philosophical work? Do we find it in the syllabus and in the objective arguments and philosophical theories presented to our students in a more or less clear and systematic way by the more or less competent teacher? Or do we find it in the subjective anxiety and the inner struggle of the individual student that tries to make sense of the world around him and his place in it?

We believe this is where we can find the essence of philosophy and this is what we should be trying our students to get from practicing and studying philosophy, a sense of wonder to everything around him and an anxiety to think and struggle to find answers to their own philosophical problems. And this is possible only if we can find out how to motivate our students to that "inner struggle", if we are capable of creating in their spirits the same anxiety and "will to know" that the Great Philosophers we praise felt towards their own philosophical problems. Without conquering the problematic dimension of philosophy, without that anxiety and "will to know" our students won't endorse any effort in the difficult and demanding task of thinking for themselves and will prefer to listen to a lecture or read a book instead of thinking by their own heads. Useless to say the importance of attending lectures and reading good philosophy books but we can ´t let these activities replace the essence of any philosophical work, and that is to think critically and creatively about philosophical problems.

In our view the job of the philosophy teacher is, in the first place, to make sure that this anxiety to philosophize arises in his students. This anxiety is like a necessary condition to every philosophical work. After that necessary condition is assured in his classroom the teacher needs to guarantee that the philosophy process is done accordingly to the principles of the "game of philosophy" and for this other necessary conditions have to be met. Conditions like willingness to pursue the truth and avoid wishful thinking, knowledge of the appropriate philosophical tools for theories and argument assessment and an inner attitude of recognition of ones own limitations and availability to search for other opinions and points of view.

Although he may not want to recognize it a teacher who works along the current educational paradigm will have to focus his work on teaching "what to think" and not "how to think" and the final product of this is generally a pedagogical tragedy with most of our students ending up their

school years with an essentially *asocratic attitude* towards the school curriculum and toward their educational process and life in general. What we get from this widespread way of teaching our children are grown up adults with fundamental childish patterns of thought. Men and women who systematically think along with the herd and don´t see the point in becoming autonomous critical thinkers who constantly and systematically strive to improve their thought patterns that remain essentially the same since their primary school days: prone to wishful thinking, incapable to reason with logic and coherence and fundamentally self-centered and deaf to other persons perspectives and thoughts. The actual end product of decades of education is *asocratic* citizens, adult men and women who can hardly think strait, and you just have to look around and talk to your fellow citizens to realize that this is true and that we haven´t really learned nothing from what Socrates taught us 2500 years ago.

In contrast with this actual and sad picture the idealized *socratic student* (and citizen) would be a motivated philosophy student who seeks to know and to think along with the great minds of the past and present about the fundamental problems of philosophy that are also his own fundamental problems has a Human Being.

This idealized student is a sort of proto-philosopher that thinks from his own head and refuses the dogma and the assimilation of knowledge that comes only trough the teacher's authority.

We believe that this picture is not a farfetched utopia but a picture of the real potential of our students or of what our students can actually become. An image that is completely obscured by a pedagogical system that forces teachers to "teach to the test" and does not allow our students to fulfill their potentials or, at least, properly develop their critical and creative skills and embrace that *socratic atitude* towards life and learning.

It is necessary that the teacher finds this "potential student" in a compromise between the "ideal student", that student designed by the most progressive and optimistic pedagogical theories and the "real student", that ungrateful student that screams to us that the "ideal student" is a mirage and we might just give up on those unrealistic intentions that first led us to become teachers and educators. What our teachers are lacking is not good intentions, nor theoretical ambitions, what they are lacking are good exercises and well tested practical strategies for the teaching of philosophy that serve as bridges between those abstract pedagogical intentions and the concrete chairs and tables of our "all too real" classrooms.

It is here that we believe we can contribute in a way to the fulfilling of that "kantian ideal" professed by almost everyone but practiced by so very few.

In this article we present some philosophical dialogue-oriented strategies that, we hope, will allow any motivated teacher to help his students become better critical thinkers and, in doing so, help them become better philosophers with an actual and practical sense of intellectual critic, courage and stamina to get a grip on real philosophical problems, creating their own philosophical lines of thought, trying out their own arguments, looking for their own examples and counter-examples, inventing original mental experiments, working on concepts, etc. By doing all of this they will be actually using all the tools and techniques philosophers use in their work. Once again, it´s in the subjective struggle of a human being against a real philosophical problem that lays what we called the "spirit of philosophy" and to find that "spirit" is what we aim in a Philosophical Dialogue with our students.

Once again, we believe the methodology that best serves our interest of teaching "how to think" is not the lecture, the reading of philosophical texts nor the answering of textbook "philosophical quizzes". The methodology we propose here is actually the most ancient way of engaging in Philosophy, the Philosophical Dialogue in the way Socrates taught us 2500 years ago.

2 - DARE TO DIALOGUE

To see philosophy and its teaching as a dialogue among pupils is not to take away the teacher's responsibility to teach and the students responsibility to learn. To teach philosophy through dialogue is to look for a different approach to this teaching and apprenticeship. A *socratic* approach in which the teacher gives his students a "how" rather than a "what" and the student don´t get an answer from the teacher but find a question in its place. More than pass on to his students what they should think the teacher shows the pupils how to engage in critical dialogue with each other. It´s this dialogue among peers that will improve both their critical thinking skills and also their dialogical attitudes, that are at the same time social and philosophical attitudes. If students understand and learn how to maintain profound and meaningful dialogues in a group they will be also learning how to reason better and in a more profound and meaningful way. It´s likely that they will use these new acquired skills and attitudes even when they are alone reasoning by themselves.

In a Philosophical Dialogue the teacher should work as a teacher but as a referee showing his pupils the rules and basic moves of the "game of philosophy". These rules and basic moves allowed in a Philosophical Dialogue are nothing more than the social, logical and philosophical standards of reasoning and argumentation. That's why we believe that the regular practice of Philosophical Dialogue with our students will help them to understand and refine their own inner dialogue helping them to become more conscious of what they are doing when they think and, at the same time improving the way they think. Teaching our students to dialogue is, at the same time, teaching them how to think.

Even for those teachers interested (or compelled) to pass on to their students some philosophical content in the form of arguments and theories the use of Dialogue as a pedagogical tool can be a valuable instrument to allow them to identify with greater precision the problems students face when coming into contact with philosophy, this awkward signature with its strange problems.

In addition to conducting a dialogue with their students the teacher will have a privileged access not only to *what* his students think but, much more important and interesting than that, to *how* his students think. And if our teacher knows where to look he will certainly have a radically different image from that clean and polished image that students are expert at presenting him in a test or assessment form. Years of education have prepared our students to be experts in saying what other people want to ear. And this different image of our students will appear because in a Dialogue are not the end products of students reasonings and thoughts that comes to surface but, instead, what surfaces are the actual processes of formation of these reasonings and thoughts. All the attitudes, assumptions, feelings, dogmas, values, pressures, frustrations, likes and dislikes that structure our students mental landscape and enable them to think the way they think. All that underground work that is invisible to the teacher in an exam or in a normal lecture-style class will emerge and become explicit in a Dialogue. Facilitating a Philosophical Dialogue with a group allows the teacher to have a general overview on how his students really think and that will give him the opportunity to check if they have or lack the fundamental attitudes of the philosopher such as resilience and patience to engage in hard intellectual problems, courage to face the authority and humility and willingness to abandon their dearest beliefs if proven wrong or incoherent.

From the students point of view having Philosophical Dialogues with their peers will help them discover some substance and relevance in the philosophical problems presented to the group. Without these engagement

with others these problems will hardly deserve any real attention from students who will see them as too much distant from their interests, abstract and useless to deserve any kind of engagement apart from the periodic effort to get a good grade in an exam or approval from the teacher in the classroom inquisitions. And that, in our view, is taking away the personal and subjective character of every philosophical problem that needs to be lived and felt has a real problem to someone real in order not to become some mere utilitarian concern that reduces its spirit to some set of vague and lifeless concepts with no significance to student's lives. Philosophy should matter to them inside but also outside philosophy classrooms.

A Philosophical Dialogue in the classroom will most certainly allow students to spot a direct connection from the philosophical problems presented by the teacher to different persons (their classroom colleagues) who have different experiences, expectations and value assumptions and that live that problem in different ways.. This allows them to broaden their philosophical landscape and enrich their minds with alternative points of view about a particular problem, as well as different intellectual ways to tackle that problem as some students are more creative thinkers, others are more critical thinkers, others are good at making pin-point deepening questions and others easily find contradictions in a speech while others are good at giving examples and finding counter-examples. To get all this richness from his dialogical group the teacher needs to make sure the participants have the opportunity to make good use of their cognitive toolbox and to do that he needs to structure the dialogue in an adequate and though- provoking way. Conforming to the rules of the Dialogue the group should be able to make the best use possible of all this different individual expertise's and in doing so everybody will learn something with everybody else. Even those students who normally remain silent in a Dialogue can have something to teach to the class as it normally happens that they are the best listeners and they can show with their own example an alternative way to be present and conscious in a Philosophical Dialogue.

It should be noted that in every Dialogue there are some usual difficulties and obstacles to reasoning and communication and students will have to deal with them. All students will show different attitudes and dispositions and sooner or latter the group will have to deal with them. Some will be eager to talk and less eager to think, some will be stubborn and won't recognize when they are wrong, others will only listen to themselves, etc.

These obstacles to reason and communication can also serve as stimulus for future reflections and dialogues. In the Philosophical Practice

tradition this dialogues about the process of dialogue and reasoning are called "meta-dialogues" and they allows the elements of the group to become more aware of their own though processes as well as their cognitive difficulties and capabilities. These meta-dialogues are essential for a group to improve the quality of its Dialogues and, at the same time, the thinking competences of all its elements.

3 - S.O.S. DIALOGUE

In this section we share some teaching strategies that can help philosophy teachers maintain dynamic and interesting Philosophical Dialogues with their students.

In order to understand the role of the teacher during a Philosophical Dialogue we use an analogy with a game. Like any game the Philosophical Dialogue has some rules that must be met, some obstacles that prevent us from playing the game and some techniques that we use to overcome these obstacles and play a better game. We begin by presenting some of the **rules** that should guide the teacher and his students during the "game of philosophy" as well as some of the **obstacles** that usually arise in a Philosophical Dialogue. After that we suggest some dialogue moderation **tools**, that can help our students overcome each of these obstacles by themselves..

Here it´s important to underline that the role of the philosophy teacher during the Philosophical Dialogue, "the game of philosophy", is that of a *referee*/*coach* rather than of a player. The teacher should work has a *referee* preventing the students from breaking the rules of the game and at the same time should work has a *coach* calling their attention for the obstacles that arise in the way of reason and dialogue and his students the necessary thinking tools to overcome these obstacles by themselves.

By now it should be clear that the only purpose of the teacher is to get his students to think as a group and resolve among themselves the various problems which arise within the Philosophical Dialogue. To achieve this the teacher should avoid to summarize, clarify and explain what the students should be able to summarize, clarify and explain. In other words the teacher should resist the temptation to do the work the students are supposed to do by themselves and that is to think.

Any comments from the teacher relating to the philosophical contents should be left to the end of the dialogue or, better yet, to a future dialogue. All the errors, doubts, misunderstandings or conceptual confusions that students are making should, as far as possible, be left to the

students themselves to think and live their consequences. Like this they will learn that philosophy is a "hard game" where one has to face frustration, misunderstanding, confusion, joy and even aggression from others to our own ideas. In a Philosophical Dialogue our students will see that philosophy is something real and they will feel that it is something close to their problems, values and ideas, something with real implications to their immediate lives. It is precisely in getting our students to live the obstacles to philosophical thought and forcing them to persevere to overcome them that we find the true meaning of the term "pedagogy of error".

In a Dialogue with our peers philosophy is no longer a distance and vague subject but a close and intense one. And this happens because the contents of every Philosophical Dialogue are the ideas, problems and values of its participants and not the ones of someone outside the dialogical circle. Philosophy matters to them because their ideas matter to them.

From our experience in organizing dialogues in philosophical communities of all ages we believe that the ideas and arguments that come out from a diversified group of people are interesting and rich enough to spend some hours per week thinking about them. The philosophy teacher just has to trust the ability of his students to have deep and meaningful philosophical ideas and prepare himself to cope with them. This is, of course, the difficult part for a teacher that is used to prepare in advance all he has to present to his students. No class plan can save you when you engage in a Philosophical Dialogue with your students. You have to jump into the river with them and "go with the flow", or torrent, of ideas that come out of their vivid and joyful minds. This is surely a dangerous and even terrifying idea to some traditional philosophy teachers who want to be sure of every step they are taking and do not allow any room for error in their classrooms. But if you are that kind of teacher perhaps you have chosen the wrong subject to engage with. Risk taking and uncertainty are at the heart of philosophical thinking. So take the risk or go teach history or mathematics but leave philosophy to the awake and brave.

3.1 - EIGHT RULES OF PHILOSOPHICAL DIALOGUE

Any person, regardless of his or her age or degree of schooling, can play the "game of philosophy ". In other words, anyone with a minimum of maturity and rationality is able to take part in a Philosophical Dialogue, provided that he is willing to comply with some very simple rules that structure the dialogue separating the "philosophical" from the "non-philosophical".

To give an example, in a game of tennis if you talk loud or insult your

opponent while he is serving the ball, or if you play without respecting the lines that limit the "tennis court" you are clearly breaking some fundamental rules of tennis. In a similar way if you insult or scream to your interlocutors in a Philosophical Dialogue, or if you insist in talking over them without listening to what they have to say or if you distort their words to something that fit your intentions and agenda, then you are also in a clear violation of the fundamental rules of the "game of philosophy". In a nutshell, the attitude you should bring to a Philosophical Dialogue, is the attitude of fair-play and friendship that you should also take to any sport or game you play with your friends. This two dimensions of "game" and "friendship" are essential to the process of dialogue. You must take them with you to a dialogue so that it does not become a "battle of egos" or a fruitless debate of opinions where every participant is just trying to defend is background ideas and values and is the least interested in understanding the ideas and values of others. If you see the other participants in a dialogue as your friends, and not as your opponents in a debate, you'll be prone to listen to them and even disagree with what they say in a respectful and rational way, without reacting to their words as if they were attacks on your person but, instead, you'll think and act about them in a rational and civilized way. Rationality and civilization are the backbone of Philosophical Dialogue, they are necessary conditions that have to be present if you want to do philosophy with your friends and not just talk about it with some strangers.

Maybe this is the reason why most of philosophy teacher hardly have real dialogical and philosophical interactions with their students. They are not being their friends but their teachers and they are not presenting philosophy in the spirit of a game but in the form of a lecture.

Here are what we consider to be the Eight Basic Rules of the "game of philosophy".

From our point of view these are the rules that prevent that a Philosophical Dialogue stumbles into a coffee-talk kind of conversation where you hardly get meaningful and connected propositions, arguments and questions but instead you get a multiplication of speeches that (with some luck) are closer to a series of monologues than to a real Philosophical Dialogue.

1 - Present your thesis

To initiate the "game" the students must present their thesis committing themselves to a certain position. To do this they must abandon the comfort of a neutral position that does not make a mistake because

nothing relevant is said. The fear of making a mistake in front of the teacher and colleagues also makes some students adopt a kind of a relativistic position that accepts every possible position without, however, committing to no positions at all. "It can be either way" or "It depends", are some of the ways students try to get rid of the responsibility to think by themselves. Without this real commitment to a given position the thought process does not advance and the Dialogue does not have any fuel to go proceed.

2 - Give reasons

After committing to a certain thesis a student must defend it with reasons and it is at this point that the "game of philosophy" truly begins. From here on students are asked to analyze the reasons presented to the group, to check the relevance of their connection to the issue at hand and to the thesis they claim to adhere. With more experienced and mature groups we can also ask our students to find the assumptions on which these reasons are based..

3 – Cultivate a dialectic spirit

This can only be achieved with the continued practice of Philosophical Dialogue with your students. The reason for this is that the more your students play the "game of philosophy" the better they´ll be at actually playing it, and more naturally all it´s moves and rules will feel to them.

To be able to do this in a natural and almost effortless way the philosophy teacher must continuously submerge his students in a dialectical environment that cultivates this active and critical attitude that questions, problematizes, analyses and discusses every reason and assumption presented in a dialogue.

4 - Give examples

Whenever you find necessary you should invite your students to put forward actual examples that illustrate their ideas. This is a way for helping them to clarify their reasoning by distinguishing an explanation from an example, but also by bringing the abstract concepts typical of any philosophical investigation closer to examples from their daily life, episodes that they can easily relate to.

5 - Look for counterexamples

One of the most common and effective intellectual movement that one can do in a Philosophical Dialogue is the counterexample. When we are teaching our students to play the "game of philosophy" we must make sure that they understand the implications of a good counterexample both to their own ideas and to the ideas of others. We should also teach them how to deal with the, some times, fatal implications of a counter-example.

When faced with a good counterexample to his generalization, definition or condition the author is forced to rethink his original idea, making it weaker or simply withdrawing it completely.

6 - Advance hypothesis

The exercise of philosophizing is largely an exercise of honest and accurate speculation, an effort we make to "see further" (the original etymological sense of the Latin verb speculari). In a Philosophical Dialogue getting our students to "see further" is to get them to see beyond their protective walls, their mental watchtower inviting them to "think the unthinkable". In order to do this we must ask them to advance hypothesis, i.e. to take step into the unknown, to venture to take a risk into thinking in a different way and from a different perspective from the one they are accustomed, often in a direction that points contrary to their fundamental beliefs.

Asking our students to advance a hypothesis is a way to get them out of their comfort zone where thought can hardly get any food and where mental mechanisms are prone to repeat thought processes that solidify beliefs and dogmas.

7 – See other perspectives

Sometimes a way to solve a particular philosophical problem (or, for that matter, other kinds of problems) is trying to see other perspectives on that issue.

Every time our dogmas and prejudices don't allow us to see beyond our certainties and subjective points of views our thought process becomes blocked.

The Philosophical Dialogue with their colleagues and friends can help our students unblock that "dead end" situation allowing them to have access to many other points of view that, all together, can enable them to see a problem with greater objectivity and clarity and, at the same time,

find different solutions and perspectives to that problem or, finally, to realize that there was no problem after all.

8 - Accept to change your opinion

Learning to play the "game of philosophy" is also a process of learning to reconcile with our fundamental limitations and fallibility.

Accepting to reformulate or change our initial thesis in the light of the suggestions and criticisms received during the dialogue or, even, to give it up completely if the reasons given against our position force us to do it, is a learning process that most of our students and many adults, including many teachers, have yet to accomplish.

The continuous exercise of Philosophical Dialogue confront us with the limitations of our own opinions, perspectives and values and doing so it works as a sort of spiritual exercise that helps us mature and become more attentive and conscious human beings.

In the following part of the paper we will be presenting some of the most common obstacle that a teacher will find in a Philosophical Dialogue with his students. For each obstacle presented we also suggest a specific tool to try to overcome it.

3.2 - SOME OBSTACLES AND SOME TOOLS

Ambiguity

A term is ambiguous when we can have more than one possible meaning to it. For example in the propositions "There is no knowledge without reason" the term "reason" in can either mean cause, reason, reasoning, or even consciousness. The meaning of the proposition depends on the specific sense we give to it.

Tool: *New Concept*

To remove the ambiguity of a term we should ask the author of the initial proposition in what sense is he using the term in question. To do that he should give us some *new concept* that allows us to distinguish the intended meaning of the proposition from other possible meanings. Thus, in the example given above, the *new concept* "reasoning" indicates the true meaning of the proposition. All too often our students won't be able to come up with a new concept to clarify their initial proposition and the teacher can then ask the group for hypothesis. The group suggests several possible concepts and from that list the student can choose the one he considers most appropriate.

Dogmatic Certainties

This is a very common anti-philosophical attitude that consists in defending an idea without having any reasons or rational basis for it.

Normally this is a blockage attitude for any kind of thought and dialogue and is accompanied by some other dogmatic attitudes such as the refusal to consider alternative ideas or explanations, to analyze and understand its assumptions and to see beyond the surface of ones own opinions or points of view.

Tool: *Imaginary Critic*

Sometimes a dialogue between students reaches a point where they all agree, more or less artificially with some point of view or gets stuck against the unanalyzed certainties and stubbornness of some participants. Both this artificial consensus and rigidity of thought are blockages to Philosophical Dialogue. To help students continue to think and try to relaunch the dialogue we can ask the group "what reasons would present someone who wouldn't agree with that position". This technique is a type of Devil's Advocate and encourages students to develop an "inner dialogue" that searches for alternative reasons and explanations when there is no real interlocutor to challenge their ideas. Sometimes students agree with the "imaginary critic" and end up changing their positions.

Irrelevance

This obstacle consists in changing the direction of the discussion by including elements that do not relate directly to it. This change of direction may be inadvertent, as when the student doesn't get the essence of the subject under discussion, or it may be intentional, as when the student wanders and refuses to answer directly to the questions that are asked and tries to evade them, or when there are appeals to the authority of any author or known personality or even when he chooses to attack someone personally in a discussion and not his arguments.

Tool: *Anchorage*

When the teacher feels that a students intervention is irrelevant to the subject in hand, or when this relevance is not clear, he should try to clarify things to everyone by bringing the discussion back to the initial question (or task) and simply asking the student "how is your intervention connected to the initial question?" In Asking for this "lost connection" the teacher will be working to improve the pertinence and accuracy of students interventions and also their argumentative consciousness.

The *anchorage* is also a good strategy to find subtleties in students reasonings that may have escaped us in the hit of the dialogue. Frequently a student speech that looks to us completely aside and absolutely irrelevant

reveals itself full of meaning and adequacy when it´s connections are explained to us. Here, as in many other places, we should also cultivate some sense of humility towards our student's capabilities. As a rule of thumb the teacher is well advised if he gives his students the benefit of the doubt and trusts that they can think for themselves. His students will feel that they are trusted upon and will most certainly try to live up to that vote of confidence. In the pedagogical jargon this is known as the "pygmalion effect".

Fear [of making a mistake]

The expressions "I don't know" or "I´m not sure" are two of the most common responses we hear from students when they are face-to-face with some big philosophical problems such as "Does God exist?", or "Does the Universe have a reason?" or "What is art?"

Our students response come to quickly and sudden "I don't know" and after this waits in expectation for an answer from the teacher or from some of his "smarter colleagues". The teacher all too often accepts this refusal to think and in this sad way the students thinking process comes to an end even before it was offered a chance to get started.

Tool: *Dare to think*

In a Philosophical Dialogue we are more interested in "how" the students thinks than in "what" he thinks. With this basic assumption at hand we can challenge our students to dare to think by themselves about what is being asked and a good strategy to achieve this is to ask our students for an hypothesis rather than for a definite answer. "What is your hypothesis?" rather than "What is your answer?" can help students loose themselves a bit more and forget about their fear of getting it wrong. "A dialogue is not an exam, your mistakes are welcomed", this is what a participant in a Dialogue as to know.

Precipitation

In a Philosophical Dialogue there is precipitation when someone shortcuts the reasoning process and rushes to reject in an intuitive and immediate way any criticism or argument that is advanced against his position. This is normally done in a reactive end emotive way without devoting any time to assimilate and truly understand the reasons for this criticism, as well as its main assumptions and implications.

Precipitation is probably a genetic defense mechanism that might have worked well with our ancestors in the *savana* but nowadays it is surely not a good strategy to bring to a Philosophical Dialogue, where what you want is not to defend yourself with hot blooded and instinctive decisions

and reactions but cool and reflected actions towards some ideas and forms of behavior. Remember what we wrote earlier that a dialogue is only possible between friends, and friends don't attack each other. When you are with friends you can leave your defensive weapons at home.

Tool: *Suspension of judgment*

When we ask our students to suspend judgment on a particular issue we are asking them to temporarily assign the same weight to the different positions presented in a dialogue. This is done in order to really analyze and understand the various possibilities that lay ahead of us in a Dialogue and prevent the "all too human" emotive and instinctive reactions to opposite claims that we tend to engage in.

To suspend his students judgments a teacher can ask them to "Look at the reasons for and against this idea", or, "Do not commit yourself to soon to that position, take a look these different opinions from yours."

The aim of putting a students particular statement "in brackets ´" is to invite him to further examine different reasons and arguments in favor or against that position, to give him time to find out more information on the subject or to listen to different perspectives on the same issue that will help him take a more well-founded decision.

4 - SHUT UP AND LISTEN

We are perfectly aware that for most people asking a teacher to be silent in a classroom is almost the same as asking a singer not to sing in a concert. Isn't teaching all about speaking and explaining stuff to students? By now you should already know our answer to that question and if you were convinced by our arguments and want to engage in a Philosophical Dialogue with your students you should skip the usual beginning of the year introductions and explanations about "What Is Philosophy" or "What Philosophers Do" and take the risk to engage directly in a philosophical confrontation with them. Ask them right at the first class a "hard-core" philosophical question you believe they will be interested in, or ask them what question they find to be "the most fundamental question of all"? Then go on from here to make some *juicy philosophical dialogues* with them, letting yourself "go with the flow" of dialogue and see were and how it is going.

Trust us, most of the times you will be surprised with how much "philosophy" your students have inside them. You just have to listen to them and that is the best philosophical teaching you can do.

So, if you want to teach your students how to think you must learn how to shut up and listen.

BIBLIOGRAPHY

Brenifier, Óscar, *El Diálogo en Clase*, Ediciones Idea, 2005, tradução Gabriel Arnáiz.

Kant, Immanuel, *An answer to the question: What is Enlightement?*, Penguin Books, 1991, tradução H.B. Nisbet.

POWER AND BEAUTY IN PSYCHOPATHOLOGY EUGEN BLEULER'S CONCEPT OF SCHIZOPHRENIA

João Machado Vaz

Early twentieth century psychopathology is indelibly rooted in the works of philosophers such as Schopenhauer, Nietzsche and Bergson. If these and other authors' work was successful in depicting some of the fundamental principles of human psychology, one could legitimately expect them to account for at least part of psychopathological phenomena as well.

Focusing on concepts intimately related to Schopenhauer's *will* and Nietzsche's *will to power*, this paper hypothesizes that the expression of some psychopathological states and processes can be envisaged as the result of a disruption of the relation of power between the individual and his world.

Self-made descriptions of patients with schizophrenia will be used as a starting point for a brief and exploratory research into the possible relations between power and beauty in the human psyche. The question of whether the experience of aesthetical contemplation is accompanied by an abnormal feeling of power will be dealt with. An example of time-limited psychotherapy will serve the purpose of trying to understand to what extent such a relation holds within the scope of non-pathological psychology.

Subsequently, Eugen Bleuler's concept of schizophrenia will be introduced. Bleuler considered that much of the symptomatology observed in these patients was the expression of a reaction of the individual's psyche to an unbearable situation. It is argued that Bleuler's theory provides us a

paradigmatic example of how philosophical concepts evolving around the idea of *will to power* can help explaining the behavior and mental processes through which patients with schizophrenia try to regain power over their dramatic existence. Bleuler's theories of catatonic symptoms, such as catalepsy and negativism, will be revisited as they represent the peculiar transformation that seems to take place in the relation of the patient with a world that he experiences as being hostile to him.

POWER

It is overwhelming to realize how power, in its immense variety of forms, is present in our psychology, from the simplest act or thought to the biggest accomplishment of a lifetime, from our self-conscious existence to the coded language of the body described by Nietzsche.

Amongst the infinite implications of man's self-reflexive capacities, consciousness of the limits of our own power over the variables of one's existence seems to be one of the strongest, if not the strongest, inputs in our psychological functioning. Man's industry and intellectual abilities alone could never explain the invention of religion, art or science – the subtle action of that limited power being impossible to be taken out of this equation. The dissimulation of those limits lies, therefore, at the epicenter of our psychology to such an extent that we believe legitimate to ask: what remainder of humanity would there be in an all-powerful man, free of any restriction, unaware of any resistance?

For Schopenhauer, human nature could be seen as the highest possible objectification of a *will* that is, otherwise, expressed in everything that exists. Unlike other animals and matter, human consciousness would be the artifact through which *will* would find and recognize itself. Thus, this unifying and genetic principle for all that exists and is created is, therefore, the substratum of the world (*the world as will*): *it appears in every blind force of nature and also in the pre-considered action of man ; and the great difference between these two is merely in the degree of the manifestation, not in the nature of what manifests itself* [1].

Nietzsche, influenced by Schopenhauer's concept of *will*, seems, nevertheless, to have given its possession back to the individual: he ceases to be the simple vehicle and temporary holder of the *will* to become its legitimate owner. Moreover, it is no longer the blind and directionless *will* that Schopenhauer[2] described, but it is the *will to power*.

[1] Schopenhauer, A. (1818). *The World as Will and Idea*, *p.*143;

[2] Nietzsche, F. (1900). *A vontade de poder*, Vol. I, *p.*175: *is that 'will' what Schopenhauer believes to be the thing-in-itself? My principle is that this 'will' is an unjustified generalization. That 'will' does not exist [...] what Schopenhauer calls 'will' is a mere empty word.*

Espinoza, in his *Ethics*, had already associated power with human psychology, while writing that *the mind, as far as it can, endeavors to conceive those things, which increase or help the power of activity in the body* [3] and that *pain is an activity whereby a man's power of action is lessened or constrained* [4]. The pertinence of this observation by Espinoza is quite remarkable. If we look at descriptions of pathological states such as mania and depression, it becomes clear its association with subjective experiences of increase and decrease of power over one's world. For example, in states of abnormally elevated mood – such as in mania episodes in bipolar affective disorder – an individual may evaluate his mental abilities as being extraordinary, believe that he will turn into a millionaire, or present an unbreakable conviction that he or she descends from historical personalities such as Napoleon or Alexander the Great. Nietzsche synthesizes this relation between power and psychology when he defines *pleasure as the feeling of power*[5].

Diametrically opposed to this feeling is that of depression. The background of depression is powerlessness: anything that in the past may have interested the depressed individual and the things he used to do with little effort to his own pleasure, would now force him to an unbearable use of energy that he believes to be in no possession of. He may find some comfort in a dark room or in endless hours of sleep, each and every small difficulty of everyday life being faced as an insurmountable obstacle. In cases of psychotic depressions, the patient could even get to the point of believing that he is, in fact, already dead, a walking dead unburied body.

Isn't it paradigmatic the fact that clinical guidelines for the prevention of suicide in severely depressed patients alert to a sudden recovery of mood as a possible sign that the individual has made the decision of taking his own life[6]? Would it be legitimate, in such cases, to interpret this decision as an ultimate and despaired act aimed at increasing one's power over his existence? Could the power to decide one's moment of death be the only feeling of power (and pleasure) some depressive patients can experience?

Let us consider a description made by a patient of German psychiatrist Hans Gruhle (1880-1958), as quoted by Karl Jaspers in his *General Psychopathology* [7]:

> *"I woke up one morning with the most blissful feeling that I had risen from the dead or was newly born. I felt supernatural delight,*

[3] Espinosa, B. (1677). *Ethics*, Part III, Prop. XII;
[4] Espinosa, B. (1677). *Ethics*, Part III, Definitions of the emotions, *p.332*;
[5] Nietzsche, F. (1900). *A vontade de poder*, Vol. II, *p.*27;
[6] For example, Canadian Mental Health Association (2006). *Suicide awareness*. [www.cmha.ca]
[7] Jaspers, K. (1959). *General Psychopathology*, Vol. I, *p.*114;

> *an overflowing feeling of freedom from everything earthly… brilliant feelings of happiness made me ask "am I the sun? who am I? I must be a shining child of God… Uncle A., changed into God, will fetch me… we shall fly straight into the sun, the home of all those risen from the dead"… in my blissful state I sang and shouted; I refused to eat and no longer needed to eat; I was waiting for paradise and to feast on its fruits."*

In this self-made description of a schizophrenic patient, the euphoric emotional tonality seems to emerge out of a background of omnipotence and grandiosity. The way the patient experiences his immense power seems inseparable from an elevation of his mood and would – on the footsteps of Nietzsche's philosophy – legitimate the question of knowing to what extent one differs from the other. In fact, it is otherwise unlikely that we would find a description in which a patient is simultaneously invaded with feelings of unlimited power and anguish. Mood seems, therefore, to comprise both how the individual appraises his reality, as well as the extent to which he experiences, consciously or not, the responsibility for the occurrence of that reality. If that judgment is of a negative valence, one may still not be able of exempting himself from the responsibility over that reality, the feeling of power being, in this case, replaced by that of guilt. In the case of Gruhle's patient, responsibility appears as the expression of the individual's power, in the latter, it comes as the result of the lack of it.

We would be much inclined to say that envisaging human psychology as being deeply rooted in the feeling of power is both pertinent and elucidative of many aspects of psychopathological phenomena. So at this point, the direction we set out to our research is to understand to what degree we can conceptualize these phenomena as the result of a disruption of the relation of power between the individual and his reality.

BEAUTY

It is self evident that any action implies imperfection. The idea of perfection that we find within ourselves results in that every action and every expression of our power is endowed with an aesthetical element. A paradigmatic example is that of some obsessive personalities, whose action is often impaired by a maladapted aspiration to perfection. In this case, the psychological functioning of the individual does not allow him a peaceful cohabitation with imperfection.

Going back to Gruhle's patient and the appraisal he makes of his reality, in what way could that appraisal be thought of as a judgment of an aesthetical nature? Another of Gruhle's schizophrenic patients says:

> *"All the people I speak to believe in me wholly and do what I tell them. No one tries to lie to me; most of them have ceased to believe in their own words. I have an indescribable influence on my surroundings. I think my look beautifies other people and I try this magic out on my nurses; the whole world depends on me for all its weal and woe. I will improve and rescue it."* [8]

In this patient's delusion, the different aesthetical elements are all too clear, alongside with the same grandiosity and feeling of power. The individual sees himself as the origin of a beauty that spills over to an ugly and mean world, making it beautiful and unthreatening. The delusion, filled with schizophrenic ambivalence, works as a psychic process which provides the individual with means of denying that ugliness and discomfort. So this *beautifying* of the world is or is it not by definition an act of power?

The centrifugal conception of psychopathological phenomena as an exaggeration of normal psychic functions is again quite useful. Let us take the example of a time-limited psychotherapy to illustrate what we mean[9].

Patient C. was 25 years-old when we first met. His mother had convinced him he could benefit from psychotherapy and he conceded coming to see us. She was most worried about her son's lack of academic and professional goals and his rather unhappy mood. He agreed to that analysis. He felt his mother knew him much too well and he saw her as his unconditional affective bond. His relation with his father was quite the opposite. Our patient's self-contained and shy personality was in clear contrast with his father's impulsiveness and lack of affective and communication competences. As a teenager, C. had been a promising and proficient musician and left his hometown in his early teenage years to pursue his musical studies in a bigger city. However, a serious injury and a great deal of physical pain made him leave his dream before the age of eighteen. He felt he had been "normal" until then. After that he remembers a persistent feeling of inadequacy, a lot of guilt, "thinking too much" and being too hard on himself. He'd been in several relationships with foreign women for some reason he was unable to understand. He'd also been in and out of different colleges and found no motivation in his work as a student albeit his family's business and their expectations of him continuing its activity. When asked about his expectations concerning psychotherapy he refers his need to understand his behavior and to give some "structure" to his

[8] Jaspers, K. (1959). *z*, Vol. I, *p.*121;

[9] This case's names, places and other factual information were altered in order to preserve the identity of the patient. The modifications introduced, nonetheless, are of little importance as regards the purpose of this exposition although the detail of the information is significantly reduced for the sake of this exposition's length.

ideas. "I feel like I'm drifting", he said.

Our patient found some relieve in travelling. He'd been in various and faraway countries by himself. He said it had given him the opportunity to start all over, to think less and feel good. He could open himself differently and enjoy other people he met. This young man was a very enjoyable and good looking person. By talking to him, however, one realized how he was very far from realizing his qualities and how he experienced great difficulty in appreciating himself.

One of the first aspects that we both tried to figure out was the fact that his mood seemed to vary depending upon his location. How come he felt much better when outside the country? What was it that made him feel so distressed when in his hometown? It was difficult for C. to give this question an answer that would satisfy him. He mentioned, nevertheless, how he seemed to expect less from himself when abroad. Self-expectation being many times a disguised form of expectation unconsciously attributed to others (either fairly or not), we tried to lead the conversation to his life's failures or unaccomplished goals. Clearly, the interruption of his musical studies stood out as his first great disappointment in life. Being a time-limited psychotherapy, we risked the early interpretation and asked him if, when on stage, he felt that he played for someone other than himself. He says no. But he immediately starts talking about how his relation with his father was profoundly marked by poor communication and that he'd been against him moving to a different city to study music when he was 12. He goes on to say that he felt he had to compensate his parents for the investment done which turned out to be unfruitful. When asked if that feeling of having to compensate his parents, specially his father, could be seen as a way of compensating his progenitor for his own disappointment as well, he prolonged the most agreeable silence one psychotherapist can witness – there was insight.

The case of C. is clearly one in which the obsessive symptoms seem to elapse from the unease of the patient to accept the inevitability of imperfection. His perfectionism may be seen as a maladaptive effort to revert the growing distance that separates him from his father. This motivation of affective proximity constitutes an obsessive nucleus present in his psychic functioning. Giving up on his studies and other life projects is the way through which C. is able to preserve the possibility of perfection, that is, the idealization of his existence.

When C. is abroad, the frustration that derives from his father's affective unavailability is suspended or attenuated. Being geographically close, on the contrary, makes the emotional distance to his father impossible to surmount. The impossibility of reaching out for his father's attention and affectivity is lived with anguish and as a lack of power over his reality. He

lacks the power to change a reality that reinforces the negative judgment he makes of himself. This experience of his reality constitutes a depressive nucleus, working together with the obsessive one, as two sides of the same coin. What the former tries to look for in the outer world, the latter fails to recognize in the inner one. Bergson's words seem appropriate when he says that *there are things only intelligence can look for but will never find and things only instinct would find but will never look for*[10].

Ultimately, C. seems unable to appreciate himself for as long as he is not perfect. As far as his obsessive thinking and drive for perfectionism are concerned, it unlikely he will ever find validation of his *self,* or experience it as good or beautiful. Perhaps he can unexpectedly find a satisfying compromise or ease his endless quest through the aesthetical contemplation of the exquisite locations he has been to, such as in Schopenhauer's pure knowing subject. In such a case, despite of his intents to control every aspect of his existence and reach self-perfection, he finds beauty independently of his own efforts and will.

Travelling, art, love, drug addiction, they all seem to provide the aesthetical dimension that underlies human experience and action. What is the relation between power and beauty? Is beauty an antidote of lack of power? Should we conceive it as the supreme form of dissimulation of human impotence? Or are we merely reducing aesthetical contemplation to a biologism filled with utilitarian values? We nonetheless argue that power and beauty can be thought of as inseparable.

This so sought beauty seems to be looked for externally, in the healthy psyche, and internally, fabricated in the pathological one. Could this be a fundamental difference between neurotic and psychotic psyches? It is the adequacy of the relation with the outer world that seems to change from one case to the other, the purpose of achieving beauty remaining unaltered. In the former, the relation of the individual with his aesthetical dimension is mediated by action; in the latter, the delusion compensates the individual's inability to grasp the outer world. Man is not distinguishable from other beings on the basis of the *will to power*. But his unique feature of self-reflexivity seems to imply an aesthetical dimension that would otherwise be inexistent.

The question to be asked would then be: in what way is the feeling of power present or absent in experiencing beauty? Is it, again with Schopenhauer, that contemplation enables the individual to subtract himself from the *will* of which he is an expression of? In what way are

[10] Bergson, H. (1907). *A evolução criadora*, *p.*140;

power and beauty related in the human psyche? Is there any room for a causal relationship between the two? Is it even possible to distinguish one from the other?

SCHIZOPHRENIA

The interest behind psychopathological phenomena exceeds the study of mental disorders alone and goes beyond clinical purposes. Much because of authors like Eugen Bleuler (1857-1939), the study of mental illnesses as the exaggeration of normal psychological processes was rendered possible. Psychiatric patients stopped being seen as individuals whose mental functioning was qualitatively different from what was expected. Their symptomatology became a kind of crack in the window of human psychology, enabling the development of new psychological and philosophical conceptions of human existence. Pio Abreu, reflecting on the legacy of Karl Jaspers, wrote that *psychopathology itself, through the analysis of extreme human phenomena [is] a source of anthropology*[11]. Bleuler refers to the psychopathology of schizophrenia as *one of the most interesting and intriguing, since it permits a many sided insight into the workings of the diseased as well as the healthy psyche*[12].

Henri Ey (1900-1977), French psychiatrist and translator of Bleuler's work, tries a description of schizophrenia for the layperson in the following terms: *the patients that we currently range in this group of diseases are «lunatics» who first struck us by their strangeness, their quirks and the gradual evolution of their disorder to a state of stupor, numbness and inconsistency*[13]. Here's a more technical and widespread description of symptoms, as used in clinical onsets: [*schizophrenia*] *includes delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior and negative symptoms such as affective flattening, alogia and avolition*. Subtypes of schizophrenias are also described depending on the relative prominence of the active symptoms: paranoid, disorganized, catatonic, undifferentiated and residual[14],[15].

[11] Pio-Abreu, J.L. (2009). *Introdução à Psicopatologia Compreensiva*, *p.*29;

[12] Bleuler, E. (1911). *Dementia praecox or the group of schizophrenias*, *p.*348;

[13] Ey, H., Bernard, P., Brisset, C. (1974). *Manuel de psychiatrie, p.*528;

[14] *Ibidem*;

[15] American Psychiatric Association (2002). *DSM-IV-TR*, *pp.*312-316: *Paranoid schizophrenia: preoccupation with one or more delusions or frequent auditory hallucinations [frequently of persecutory nature]; Disorganized type: a type of schizophrenia in which [...] all of the following are prominent: disorganized speech, disorganized behavior, flat or inappropriate affect; Catatonic type: a type of schizophrenia in which the clinical picture is dominated by at least two of the following: motoric immobility as evidenced by catalepsy (including waxy flexibility) or stupor, excessive motor activity (that is apparently purposeless and not influenced by external stimuli), extreme negativism (an apparently motiveless resistance to all instructions or maintenance of a rigid posture against attempts to be moved) or mutism, peculiarities of voluntary movement as evidenced by posturing (voluntary assumption of*

The following example is that of a young patient with catatonic schizophrenia:

> *"A young, unmarried woman, aged 20, was admitted to a psychiatric hospital because she had become violent toward her parents, had been observed gazing into space with a rapt expression, and had been talking to invisible persons. She had been seen to strike odd postures. Her speech had become incoherent.[…]*
> *The patient was agitated, noisy and uncooperative in the hospital for several weeks after she arrived, and required sedation.[…]*
> *Despite all those therapeutic efforts, her condition throughout her many years of stay in a mental hospital has remained one of chronic catatonic stupor. She is mute and practically devoid of spontaneity, but she responds to simple requests. She stays in the same position for hours or sits curled up in a chair. Her facial expression is fixed and stony."* [16]

This polymorphism that characterizes the symptomatology of schizophrenia explains the innumerous definitions and classifications to which it has been subjected over the two hundred years that have passed since the first steps of modern psychiatry. Today, Bleuler's conviction that schizophrenia comprises not a single disease but a group of diseases is still accepted by many – hence the use of the expression *group of schizophrenias* in his historical 1911 monograph *Dementia praecox or the group of schizophrenias*.

In the nineteenth century, French psychiatrists Philippe Pinel (1745-1826) and his disciple Jean-Étienne Esquirol (1772-1840) remarked a sort of «stupidity» that seemed to appear in youths, whose development had been normal up to then. They've named it accidental or acquired idiocy, as opposed to the concept of congenital idiocy (what is now called oligophrenia) [17]. Benedict Morel (1809-1873) described these young man and women as «déments précoces», upon which description Emil Kraepelin (1856-1926) will regroup these syndromes into a single nosographic entity under the name «dementia praecox». For Kraepelin, this disease implied a process of cognitive deterioration (*dementia*) of early onset (*praecox*), with

inappropriate or bizarre postures), stereotyped movements, prominent mannerisms or prominent grimacing, echolalia or echopraxia; Undifferentiated type: a type of schizophrenia in which [...]criteria are not met for the Paranoid, Disorganized, or Catatonic Type;

[16] Sadock, B., Sadock, V. (2004). *Kaplan & Sadock's concise textbook of clinical psychiatry*, *p.*144;

[17] Fonseca, A. Fernandes da (1986). *Psiquiatria e Psicopatologia*, Volume II, *p.*11;

delusions and hallucinations being its most common features[18]. According to him, other major groups of psychiatric disorders were, first of all, manic-depressive psychosis, which consisted in the intermittence of illness with periods of normal affective behavior, and paranoia, a concept which by then had a broader sense than that of persecutory delusional thinking and that comprised all other psychosis with neither signs of cognitive deterioration nor the intermittence of manic-depressive symptoms.

Bleuler subscribed Kraepelin's description of these patients' *peculiar destruction of the inner coherence of the psychic personality with dominant damage of the emotional life*[19]. However, he disagreed with Kraepelin's conviction that cognitive deterioration always took place. Thus, for Bleuler, the diagnosis of *dementia praecox* did and should not depend on the course of the illness, as Kraepelin did for purposes of differential diagnosis with manic-depressive psychosis, but on the presence of a *scission* of psychical functions (*spaltung*), that could be observed synchronically. In 1908, in a conference held by the German Psychiatric Association in Berlin, he introduces and coins the term *schizophrenia,* derived from the Greek etymons *esquizos* (division) and *frenos* (mind). Let us enter, then, Bleuler's concept of schizophrenia.

DIAGNOSIS, ETIOLOGY AND PATHOGENESIS OF SCHIZOPHRENIA

Bleuler realized the fact that prior efforts to delimitate and classify the phenomena implied in *dementia praecox* had all been based on symptoms that were, although apparent and exuberant, rather contingent and, therefore, unsuited for effective diagnosis. Bleuler named *accessory* the symptoms that, regardless of their striking effects, he considered to be contingent – *e.g.* hallucinations – and *fundamental* those that seemed to be always present[20]. The fundamental symptoms Bleuler pointed out were the *loosening of associations*, *inappropriate affect*, *ambivalence* and *autism*. Before we go into the description of these symptoms, let us first consider Bleuler's perspective on the etiology and pathogenesis of schizophrenia, for

[18] Sadock, B., Sadock, V. (2004). *Kaplan & Sadock's concise textbook of clinical psychiatry*, *p.*134;

[19] Kraepelin, E. quoted by Scharfetter, C. (2001). *Eugen Bleuler's schizophrenias – synthesis of various concepts*, *p.*35;

[20] Bleuler, E. (1911). *Dementia praecox or the group of schizophrenias*, *p.*13: *certain symptoms of schizophrenia are present in every case and every period of the illness [...] Besides these specific permanent or fundamental symptoms, we can find a host of other, more accessory manifestations such as delusions, hallucinations or catatonic symptoms. These may be completely lacking during certain periods, or even throughout the entire course of the disease; at other times, they alone may permanently determine the clinical picture.*

they are of utmost importance for the philosophical framing of the disease.

Alongside with this division of symptoms for diagnostic purposes, Bleuler proposed a different division according to etiological criteria, thus developing a theory of the symptomatology. He first defined primary symptoms as those resulting directly from the morbid, yet unknown, organic process that he believed to take place. Although he reckoned that other symptoms could be considered primary as well, the loosening of associations was the one he was certain about. Likewise, this symptom accounted for the disruption of the Ego's unity, thus bringing forward the necessity of the restitution of the mind's integrity. The historical importance and scope of his concept of schizophrenia lies precisely in the pathogenesis he proposed: *above all, we must endeavor to distinguish between the primary symptoms, which are part of the disease process, and the secondary symptoms, which develop only as a reaction of the afflicted psyche to the influences of its surroundings and to its own efforts*[21]. Thus, secondary symptoms were eligible for psychological comprehension and even explanation. For this purpose, Bleuler relied much on the work of his own disciple, Carl Jung, and on the entire psychoanalytic revolution that sprung from Vienna by that time. Let us, at last, briefly consider Bleuler's fundamental symptoms.

a. The loosening of associations

A decrease of coherence in the association of ideas takes place, resulting in the impoverishment of ideation, the absence of finality in speech and thought, loss of logical sequence of ideas, stupor, confusion, *etc*. Representations that have little or no relations whatsoever with the main idea, and that should therefore be excluded from the course of thought, may, nevertheless, produce effects in the outcome of speech and thinking processes, resulting in a dissociated, bizarre, inexact and abrupt speech[22]. It seems as if Thomas Hobbes' description of a *folly* was finally given the name he couldn't find himself: […] *but without steadiness, and direction to some end, great fancy is one kind of madness; such as they have that, entering into any discourse, are snatched from their purpose by everything that comes in their thought, into so many and so long digressions and parentheses, that they utterly lose themselves: which kind of folly I know no particular name for[…]*[23].

[21] Bleuler, E. (1908). *Die Prognose der Dementia praecox (Schizophreniegruppe).* Allgemeine Zeitschrift für Psychiatrie und psychischgerichtliche Medizin; 65:436-464 quoted by Kolle, K. (1968) *«Bleuler, Eugen.»* International Encyclopedia of the Social Sciences. [Encyclopedia.com];

[22] Bleuler, E. (1911). *Dementia praecox or the group of schizophrenias*, *p.*22;

[23] Thomas Hobbes, *The Leviathan*, *p.*43;

b. Inappropriate affect

According to Bleuler, the mind's ability to produce affects does not disappear, but can be seriously impacted or inhibited. In severe cases, an emotional flattening occurs leaving the patient in a state of indifference, with the conservation instinct kept to a minimum level, and little or no reaction to situations of abuse or imminent danger. In moderate cases, such indifference may be masked in the form of a superficial affectivity that Bleuler refers to as being easier to feel than to describe[24].

c. Ambivalence

This group of symptoms expresses the patient's tendency of simultaneously endowing psychic elements with positive and negative valences. Bleuler gives an example of this kind of dissociated thought when he refers a patient who found it difficult to like roses albeit its spines; the patient liked and disliked roses at the same time[25]. The inability of the individual with schizophrenia to perform a synthesis of conflicting psychic elements resulted, according to Bleuler, from the loosening of associations.

d. Autism

Bleuler coined the term *autism* to emphasize the restricted contact with reality of the schizophrenic patient, alongside with the relative or absolute prominence of his inner world. The reciprocity between inner and outer world assumes a very particular tonality: *the most severe schizophrenics, who have no more contact with the outside world, live in a world of their own. They have encased themselves with their desires and wishes (which they consider fulfilled) or occupy themselves with the trials and tribulations of their persecutory ideas; they have cut themselves off as much as possible from any contact with the external world*[26].

As far as etiology is concerned, Bleuler explains autism as a direct result of the schizophrenic split of the mind (*spaltung*). Due to the loosening of associations, autistic thinking becomes subordinated to the affective needs of the patient, that gain prominence over his logical needs and requirements. If the external world provides the patient with elements consistent with his affects, he will integrate them in his mental life. But, if necessary, he will reject or alter those elements in conformity with his affective needs, *displacing or falsifying reality*[27]. The patient's need

[24] Bleuler, E. (1911). *Dementia praecox or the group of schizophrenias*, *p.*42;
[25] *Idem*, *p.*374;
[26] *Idem*, *p.*63;
[27] *Idem*, *p.*373;

to find a substitute for an unsatisfactory reality within his imagination can be corresponded with little or no resistance at all. Moreover, this is a phenomenon which is not absent in normal psychology. Therefore, for Bleuler, autism is the exaggeration of a normal psychological process.

Furthermore, this pathogenesis excluded the existence of primary disturbances of perception, orientation, memory, motricity, or of more complex psychic functions such as attention or volition, in a way that any deficits detected at these levels ought to be explained as the combined effect of other symptoms.

Besides the improvement of this disease's nosography, the double division of schizophrenic symptomatology allowed, on one hand, a psychopathology-based synchronic diagnosis and, on the other, the birth of a conceptual framework of schizophrenia that brought Psychiatry and Philosophy closer as never before. How is it, then, that we can glimpse the peculiar relation of the individual with schizophrenia to his reality? And in what way is that relation instructive of man's *being-in-the-world*?

> *The paradigmatic example of catatonic symptomatology*
> *«The weakening of the logical functions results in relative predominance of the affects. Unpleasantly-toned associations are repressed at their very inception (blocking); whatever conflicts with the affects is split off. This mechanism leads to logical blunders which determine (among other things) the delusions; but the most significant effect is the splitting of the psyche in accordance with the emotionally charged complexes. Any unpleasant reality is split off by the operation of autism or transformed in the various delusional states. The turning away from the outer world can assume the form of negativism. The association-splitting can also lead to pathological ambivalence in which contradictory feelings or thoughts exist side by side without influencing each other»*[28]

The large number of symptoms and its combinations is well documented by Bleuler. Besides the primary symptoms, of which the most relevant of all is the disturbance of associations, he describes a broad number of secondary symptoms that include delusions, hallucinations, disorganized speech and writing, somatic and catatonic symptoms. From a philosophical standpoint, there is much interest in Bleuler's theory on the genesis of secondary symptomatology, to which he referred as *a more or*

[28] *Idem*, *p*.354;

less unsuccessful attempt to find a way out of an intolerable situation[29]. His perspective implied that these symptoms could be seen as the expression of what remains intact in the patient's psyche in face of the dramatic change he is subjected to. The study of psychopathological phenomena could therefore shed a light on the analysis of how the individual *is* in the world. For the sake of the brevity of this paper, we will briefly focus on two catatonic symptoms – *catalepsy* and *negativism* – as examples of such phenomena.

German Psychiatrist Ludwig Kahlbaum (1828-1899) first described *catatonia,* in his 1874 monograph *Catatonia or tension insanity,* as a disturbance characterized by unusual motor symptoms in which the voluntary motor activity of the patient is impaired[30]. Bleuler not only agrees with this clustering of symptoms – negativism, peculiar forms of motility, stupor, mutism, stereotypy, mannerism, and others – but he also contributes to the psychological analysis of its genesis.

Such is the case of *catalepsy*, a symptom which consists of a muscular tonus increase, that leads patients to maintain purposeless and uncomfortable positions for long periods of time (*Cf.* Figure 1). At first, these symptoms were thought to have an organic origin and were therefore seen as motor abnormalities. Bleuler refutes this conception in favor of a psychogenetic one, though admitting that there could be an organic predisposition for this symptom's expression: *according to our present state of knowledge, all motor symptoms are dependent upon psychic factors for their origin as well as for their disappearance*[31] *[…] certain patients become cataleptic only under definite conditions or circumstances. Whenever a patient thinks she is alone, she begins to sing merrily, laughs contentedly, or curses obscenely, only to become cataleptic immediately when she knows she is being observed*[32].

[29] *Idem*, *p.*460;

[30] Garrabé, J. (2003). *História da esquizofrenia*, *p.*32;

[31] Bleuler, E. (1911). *Dementia praecox or the group of schizophrenias*, *p.*445;

[32] *Idem*, *p.*183. In *The theory of schizophrenic negativism (1910)*, *p.*13, Bleuler says: *in spite of all my effort I have been unable to see a true motor disturbance in dementia praecox either at the root of negativism or elsewhere.*

Fig. 1 - Schizophrenics patients in cataleptic positions which they can maintain for hours[33].

Throughout his famous monograph, Bleuler is rather cautious in trying to explain catalepsy and other motor phenomena. Nonetheless, he advocates that disturbances in the process of thinking, such as blockings and the interference of split-off complexes (in the Jungian sense), can influence the patients' motility. Likewise, symptoms such as hallucinations, delusions and autistic thinking could precipitate the appearance of catatonic symptoms, as they modify the individual's perceived reality. Although Bleuler is unable to fully explain the process of schizophrenic catalepsy, again we see how part of the symptomatology could derive from the relation of the individual with an adverse reality, and could therefore be seen as a defensive psychic expression. In the case of *negativism*, Bleuler is somewhat more conclusive in defining its operating process.

Negativism refers to the patient's tendency to remain indifferent or do exactly the opposite of what is expected from him.

«When the patients should be getting up, they want to stay in bed. When they are supposed to be in bed, they want to get up. They will neither dress nor undress in accordance with the rules of the hospital. Neither will they go for their meals nor leave the table once they are there […] To 'good day' they say 'good-bye'. They do their work all wrong; sew buttons on the wrong side of the clothes. They eat their soup with a fork and their desert with a soup spoon […] In short, they oppose everyone and everything»[34]

[33] Sources: (Left picture) Catatonic Schizophrenic, 1894, Dr. H. Cruschmann, in *A Morning's Work: Medical Photographs from the Burns Archive & Collection*, 1843-1939; (Right picture) Sadock, B., Sadock, V. (2004). *Kaplan & Sadock's concise textbook of clinical psychiatry*, 2nd edition, *p.*144.

[34] Bleuler, E. (1911). *Dementia praecox or the group of schizophrenias*, *p.*192;

Bleuler finds different predisposing and interacting causes for negativistic phenomena[35]: *ambitendency* – the simultaneous existence of a counter tendency for every tendency present in the schizophrenic psyche; *ambivalence* – the reduced capacity of the patient to perform a synthesis out of opposed psychic elements; the *splitting of the psyche,* that renders the effective balance and integration of mental processes and actions less likely; the *lack of clearness and imperfect logic* of the schizophrenic thoughts that lead to an inadequate adaptation to the outer world.

It is the interpenetration of psychic elements of opposite signs – as made possible by ambitendency and ambivalence – and the diminished logical needs, derived from the split of the psyche – that concur to the emergence of negativistic phenomena. One could thus expect negativistic phenomena to occur randomly alongside with normal reactions. Bleuler says it is not so and, again, justifies his position with the change of the relation of the individual with the world: *the negativistic reaction does not appear merely as accidental, but as actually preferred to the correct reaction*[36]. He goes on to say that

«*the autistic withdrawing of the patient into his fantasies, which makes every influence acting from without comparatively an intolerable interruption […] appears to be the most important factor [and] in severe cases it alone is sufficient to produce negativism*[37]*[…] The autistic and negativistic patients are therefore mostly inactive; they have actively as well as passively narrowed relations with the outer world*»[38]

The position of *contradiction with reality*[39] of the schizophrenic thus seems to account for a significant part of secondary symptomatology, as is the case of catatonic symptoms or even autism.

CONCLUSION

The above exposition of Bleuler's concept of schizophrenia does not aim at characterizing this affection in scientific, clinical or historical terms. Elements of this nature, however, were included to the extent to which they proved useful to the standpoint of Philosophy and the reiteration of the existential dimensions that mental disorders seem to elicit.

Although Bleuler appears to have been very prudent as to the use of

[35] Bleuler, E. (1910) *The Theory of Schizophrenic Negativism*, *p.*1;
[36] *Idem*, *p.*2;
[37] *Idem*, *p.*2;
[38] *Idem*, *p.*20;
[39] *Idem*, *p.*33;

philosophical concepts in the theorization of schizophrenia, he nevertheless contributed to the opening of an era of intense psychiatric research on the grounds of Philosophy (much because of his theory of schizophrenic symptomatology that implied the possibility of psychological analysis). Unsurprisingly, some of Bleuler's disciples will review and extend his work from a philosophical perspective, the most notorious being Eugène Minkowski's use of the Bergsonian concept of *duration* and Ludwig Binswanger's *daseinanalysis* based on the philosophy of Martin Heidegger. One therefore has the impression that Bleuler could have gone further in the philosophical framing of schizophrenia, but perhaps chose not to, or considered himself lessened to undertake such a task comparatively to others[40]. Indeed, a sense of reverence towards Kraepelin, Freud or Jung can be felt throughout his writings, despite the fact that very often he's in complete contradiction with these authors' ideas (especially Kraepelin's). On the contrary, he's rather more persuasive when referring to other authors' works, particularly when arguing over rather clinical and less conceptual aspects of schizophrenia. Overall, albeit Bleuler's reluctance to explore and grasp the philosophical implications of his own theory, his clinical and scientific achievements remain invaluable and stand out as a cornerstone of subsequent psychiatric research, whether within the scope of schizophrenia or not.

One of the purposes of a Philosophy-driven Psychiatry is that of reducing the number of concepts needed for the framing of mental disorders and their symptomatic expression. How many concepts of phenomenology and existentialism (or other) are required to satisfactorily account for the majority of abnormal psychological processes? And to what extent are those concepts eligible to the understanding of non-pathological psychology as well? It is our perspective that a philosophical approach to mental disorders cannot do without the notions of *power* and *beauty*, beyond that of *identity*, of course.

Anxiety disorders, drug addiction, hysteria, they all seem to present the common feature of pointing out the relation of *power* between the individual and his reality; the *aesthetical judgment* of oneself and the outer world goes side by side with mood disorders; the *integrity of the Ego* and its functions plays a major role in schizophrenia and other psychosis. And so on. Could *being, acting* and *judging* – Plato's truth, good, and beauty? – be the three fundamental dimensions of human psychology, ill or sane?

[40] For example, Minkowski, E. (1927), *p.266: as to phenomenology, we believe we can say with a good degree of certitude that Bleuler ignored completely the work of Husserl when he wrote his monograph on the theme of schizophrenia.*

The amount of time people like Bleuler or Freud spent with the same patient is not to expect in current practice; at the same time, the progress in pharmacology makes it more difficult to observe pathological phenomena such as those aforementioned – to the benefit of the quality of life of patients and their families. Nonetheless, because Psychiatry as practiced one hundred years ago is no longer possible, it seems legitimate to argue that it's up to Philosophy to contribute, beyond the rediscovering of the phenomenological and existential dimension of mental disorders, to the limitation of the excesses of mechanistic reductionisms that insidiously underlie research and clinical practice, both in Psychiatry and Psychology.

The contribution of Eugen Bleuler lies precisely in this line of thought. His empirical and conceptual work seems to have gathered evidence in favor of the relevance of the philosophies of authors such as Schopenhauer or Nietzsche, even though these were not specifically aimed at the studying of pathological psychic phenomena. Bleuler dared to say that the occurrence of a delusion proved the existence of a morbid process but that it was not, by itself, an incomprehensible phenomena, as Karl Jaspers would put it. One could argue that some delusions are, therefore, to be considered somewhat short of the threshold of pathological phenomena, that is, within the scope of normal and expected behavior and thought processes. Could Nietzsche have been less right when he said *the fictitious world of subject, substance, reason, etc., is needed: there is in us a power to order, simplify, falsify, artificially distinguish […] "Truth" is the will to be master over the multiplicity of sensations…*[41]. Devoid of a vision of man, the epistemological foundation of psychological intervention is utterly unachievable.

BIBLIOGRAPHY

American Psychiatric Association (2002). *DSM-IV-TR: Diagnostic and Statistical Manual of mental Disorders, Text Revision*. Fourth Edition. American Psychiatric Association: Washington, DC.

Bergson, H. (2001) [1907]. *A evolução criadora.* Edições 70: Lisboa.

Bleuler, E. (1959) [1911]. *Dementia praecox or the group of schizophrenias.* Translated by Joseph Zinkin, M.D. International Universities Press: New York.

[41] Nietzsche, F. (1900). *A vontade de poder*, Vol. I, *p 71;*

Bleuler, E. (1908). *Die Prognose der Dementia praecox (Schizophreniegruppe)*. Allgemeine Zeitschrift für Psychiatrie und psychischgerichtliche Medizin 1908; 65:436-464 quoted by Kolle, K. (1968) [*see below*].

Bleuler, E. (1912) [1910] *The Theory of Schizophrenic Negativism*. The Journal of Nervous and mental Disease, No. 11. Translated by William A. White (New York) from the original German edition *Zur Theorie des schizophrenen Negativismus*, Psychiatrisch-Neurologische Wochenschrift, Vol. 12, 1910/11, Nr, 18, 19, 20, 21.

Canadian Mental Health Association (2006). *Suicide awareness*. [www.cmha.ca]

Fonseca, A. Fernandes da (1986). *Psiquiatria e Psicopatologia*, Volume II. Fundação Calouste Gulbenkian: Lisboa.

Garrabé, J. (2003). *História da esquizofrenia.* Climepsi Editores.

Ey, H., Bernard, P., Brisset, Ch. (1974). *Manuel de psychiatrie*, 4éme édition. Paris: Masson.

Espinoza, B. (1677) [1883] *Ethics*. Translated from the Latin by R. H. M. Elwes, The Project Gutenberg Literary Archive Foundation.

Hobbes, T. [1651]. *Leviathan or the Matter, Forme, & Power of a Common-wealth Ecclesiasticall and Civill*. Prepared fo McMaster University Archive of the History of Economic Thought, by Rod Hay.

Jaspers, K. (1959). *General Psychopathology*. Translated by J. Hoening and Marian W. Hamilton. The John Hopkins University Press, 1997.

Kolle, K. (1968) *«Bleuler, Eugen.» International Encyclopedia of the Social Sciences*. [Encyclopedia.com];

Minkowski, E. (2002) [1927]. *La schizoprhénie. Psychopathologie des schizoïds et des schizoprhènes.* Éditions Payot: Paris.

Minkowski, E. (1988) [1933]. *Le temps vécu. Études phénoménologiques et psychopathologiques.* Gérard Monfort: Brionne.

Nietzsche, F. (2004) [1900]. *A vontade de poder, Vol. I – O niilismo europeu.* Tradução de Isabel Henninger Ferreira. Rés-Editora Lda.: Porto.

Nietzsche, F. (2004) [1900]. *A vontade de poder, Vol. II – Crítica dos valores superiores.* Tradução de Isabel Henninger Ferreira. Rés-Editora Lda.: Porto.

Pio Abreu, J. L. (2009). *Introdução à Psicopatologia Compreensiva*. 5ª Edição. Fundação Calouste Gulbenkian.

Sadock, B., Sadock, V. (2004). *Kaplan & Sadock's concise textbook of clinical psychiatry*, 2nd edition.

Scharfetter, C. (2001). *Eugen Bleuler's schizophrenias – synthesis of various concepts*. Schweizer Archiv für Neurologie und Psychiatrie, 152, 1/2001, p.35;

Schopenhauer, A. (1818) [1883-1886]. *The World As Will and Idea*, transl. R. B. Haldane and J. Kemp. London: Routledge & Kegan Paul.

Note: the author takes full responsibility for the translation of the bibliography appearing in other language than English. If applicable, the year of the original publication is shown in square brackets.

EXPERIENCING THE WORLD JOHN MCDOWELL AND THE ROLE OF SENSIBILITY

João Santos[1]

> "I suggest that we can understand some of the central preoccupations of modern philosophy by making sense of a wish to ask 'How is empirical content so much as possible?' That would give expression to an anxiety about how our intellectual activity can make us answerable to reality for whether we are thinking correctly or not – something that is surely required if the activity is to be recognizable as thinking at all. The question whether some of our thinking puts us in possession of knowledge cannot even arise unless this prior condition, that our thinking can have empirical content at all, is met. I use the word 'transcendental', in what I hope is sufficiently close to a Kantian way, to characterize this sort of concern with the very possibility of thought's being directed at the objective world. And it is in this context of transcendental anxiety that I am primarily concerned with the question how we should conceive experience."[2]

In *Science and Metaphysics* Wilfrid Sellars claims that a correct interpretation of Kant's thesis about intuitions has to consider two readings of the Transcendental Aesthetics; according to the first one, intuitions

[1] FCT Scholarship - Ref: SFRH/BD/76166/2011; MLAG (Mind Language and Action Group) Researcher; Institute of Philosophy – University of Porto.

[2] McDowell, John, (2009a), *"Experiencing the World"*, in McDowell, *The Engaged Intellect*, Cambridge, Mass.: Harvard University Press, p. 243.

involve the co-operation of sensibility and understanding (and this will be also McDowell's position); according to the second, the manifolds of intuition are *prior* to any operation of understanding and therefore play a transcendental role in guiding the flow of conceptual representations in perception – what Sellars calls *sheer receptivity*. McDowell considers this second reading an incorrect interpretation of Kant and a form of the Myth of the Given, which Sellars himself had tried to avoid since *Empiricism and the Philosophy of Mind*. McDowell wants to overcome this commitment to sheer receptivity. He finds a clue in Kant, in the so-called Metaphysical Deduction – The clue to the Discovery of All Pure Concepts of Understanding. For him, this will also be the clue to a realist position regarding the metaphysical problem of mind and world. McDowell pays particular attention to the following passage in Kant: "the same function which gives unity to the various representations in a judgment also gives unity to the mere synthesis of various representations in an intuition" (A79/B104-5). The main purpose of this paper is to assess the dispute between McDowell and Sellars and in ultimately to discuss the role of sensibility.

"THINGS ARE THUS AND SO" – ON THE ABSENCE OF A METAPHYSICAL GAP

In order to understand what McDowell wants to say about the specific role of sensibility two main aspects must be considered: on the one hand, it is important to understand how sensibility appears as a central concept in McDowell's position regarding the metaphysical problem of Mind and World. As far as I can see, McDowell tries to argue about it in the second lecture of Mind and World and we can put in a nutshell (putting in McDowell terms): "*things are thus and so*". So the first step is to understand what McDowell means by it. On the other hand, it is crucial to understanding the role of sensibility to understand how the reading of Kant that McDowell developed can be seen not only as a contribution to the Mind and World problem but also structures McDowell's whole work. This last topic is going to be discussed on the next section – "Transcendental Idealism – a pathway to reach the world?"

"That things are thus and so" is a central expression in McDowell's thought. This is how McDowell presents its importance: "*That things are thus and so is the content of the experience, and it can also be the content of a judgment: it becomes the content of a judgment if the subject decides to take the experience at face value. So it is conceptual content. But that things are thus and so is also, if one is not misled, an aspect of the layout of the world:*

it is how things are. Thus the idea of conceptually structured operations of receptivity puts us in a position to speak of experience as openness to the layout of the world. Experience enables the layout of reality itself to exert a rational influence on what a subject thinks".[3] There are several observations to be drawn from this quotation: the first one is that the expression *that things are thus and so* has a multifaceted character for the fact that things are such and such – it regards both the content of experience and the content of judgments.

To justify this, McDowell, following Kant, defends a cooperation between sensibility and understanding. Yet, in order to overcome a gap one may believe exists between these two pieces of Kant's machinery, he assumes that the operating regime of spontaneity (its concepts, its rules) is already present in sensibility, in fact, it is actualized in sensibility, therefore eliminating the need of a process, ontologically grounded, that guarantees the transformation of brute data, given through sensibility, into knowledge, by means of understanding. This view of the given as requiring a process, a mediator, is what Sellars calls the *framework of Givenness*, characteristic of traditional empiricism. Thus, it becomes possible to argue that the "*conceptual capacities are recruited from experience itself, do not apply latter to what would be given in experience (...) intuition is (...) experiential reception, already with conceptual content, that things already are thus and so. The experience is thus passive and at the same time endowed with content and the world is neither external to the realm of concepts nor outside the exercise of spontaneity*."[4] McDowell believes this cooperation between receptivity and spontaneity is present in Kant's first Critique under the heading "*Thoughts without content are empty, intuitions without concepts are blind*." Or, as he says: "*Operations of our sensibility exert a rational influence on our formation of belief, impinging on our capacities for judgment from within the conceptual sphere*."[5]

McDowell's thesis of the *unboundeness of the conceptual* lurks: if we are not mistaken, the fact that *things are thus and so* also refers

[3] McDowell, John (1994), *Mind and World*, Cambridge, Mass.: Harvard University Press, p. 26.

[4] Miguens, Sofia (2008), *Será que a minha mente está dentro da minha cabeça – Da ciência cognitiva à Filosofia*, Porto, Campos das Letras Editores, p. 177. McDowell aims to recover the notions of intuition and concepts presented in Kant work by devoting the first lesson of *Mind and World* to the analysis of these two pieces by arguing that the way these two pieces were treated through time is part and origin of the anxiety that philosophy suffers from. McDowell argues that "we should make sense of the objective import of intuitions and the objective content of judgments together. Each is supposed to cast light on the other" Westphall (2006). In this paper, Westphal presents harsh criticism to the interpretation that McDowell makes of the Kantian thesis, namely, the failure to consider adequately the role of transcendental imagination and the resulting relationship between sensitivity and understanding.

[5] McDowell, John (2009a) *"Gadamer and Davidson on Understanding and Relativism"* in McDowell, *The Engaged Intelect,* Cambridge, Mass.: Harvard University Press, p. 134.

to things themselves, in an unmediated way. It follows from this that the unboundeness of the conceptual brings with it the collapse of the distinction between *phenomenon* and *things in themselves* and thus the whole framework of Kant's transcendental idealism. I shall come back to this on the next section. Together with the thesis of the unboundeness of the conceptual comes a conception of experience as openness to the world, where the unboundeness of the conceptual does not mean that reality is reduced or exhausted by the thinkable, but only that "there are no objects that it cannot embody"[6]. The fact that there are barriers or gaps that we can consider ontological, between what is and what can be thought, does not, therefore, imply that reality depends on what can be thought. That would lead us into idealism, which McDowell simply rejects and which according to him stimulates anxiety. According to McDowell, the world is independent of what might be thought. To think otherwise would imply a choice between "*a coherentist denial that thinking and judging are subject to rational constraint from outside, on one hand, and an appeal to the Given as what imposes the constraint, on the other*."[7]

So here's McDowell's resulting position, in his own words: "*Thought can mean the act of thinking, but is can also mean the content of a piece of thinking: what someone thinks. Now if we are to give due acknowledgement to the independence of reality, what we need is a constraint from outside thinking and judging, our exercises of spontaneity. The constraint does not need to be from outside thinkable contents (…) The fact that experience is passive, a matter of receptivity in operation, should assure us that we have all the external constraint we can reasonably want. The constraint comes from outside thinking, but not from outside what is thinkable.*"[8] There is an interesting aspect here that should be considered. On the one hand, one is totally unable to make present to one's thought reality as a whole; on the other hand, the need for a constraint is required and even necessary. But what is the source, the root, of this constraint? As McDowell tells us, this constraint appears to be outside thought, but not outside thinkable content. So, what is thinkable must be seen not as a phenomenon, subject to an ontologically dubious operation, a mysterious or mystical transformation of empirical data into cognitive content, but rather as belonging to the world, although never fully presented to one. That is, the last thing one ever will find in the realm of justification will always be a thinkable content. McDowell proceeds by saying the following: "*But these final thinkable*

[6] Miguens, 2008, p. 179.
[7] McDowell, 1994, p. 26.
[8] McDowell, 1994, p.28.

contents are put into place in operation of receptivity, and that means that when we appeal to them we register the required constraint on thinking from a reality external to it. The thinkable contents that are ultimate in the order of justification are contents of experiences, and in enjoying an experience one is open to manifest facts, facts that obtain anyway and impress themselves on one's sensibility. To paraphrase Wittgenstein, when we see that such-and-such is the case, we, and our seeing, do not stop anywhere short of the facts. What we see is that such-and-such is the case."[9] Now, what is important here is that, what McDowell finds mysterious in Kant is the transcendental *operation* – the operation and not the transcendental machinery. Once he does away with the operation, receptivity emerges as already conceptual, and not as brute data that appear to us through experience. Therefore, the access to things is no longer mediated, but appears to us immediately – things themselves are directly presented to one without any interface, as they are, although our cognitive capacities can only access such content perspectively. Thus comes the externalism in Mind and World – the idea that one's mind is not inside one's heads. Another way to explain this issue is the following: "*the world is essentially graspable in conceptual thought. The world, McDowell insists, is everything that is the case, where something's being the case is something thinkable – a possible content of thought*."[10]

However, it is possible to isolate two aspects that ensure that the constraint warrants access to the things themselves. The first one is that the application of concepts in experience requires that these same concepts are integrated in a worldview which extends beyond any experience, actual or possible; the other aspect concerns the need for continuous revising, scrutiny, of our conceptual systems in the light of new experiences.[11] Willascheck identifies these arguments in *Mind and World*. What I intend to take from these arguments is that although we have direct access to things themselves, the experience that we have of them is always partial and the need to revise our system of concepts is a process that ensures the independence of reality from our thought. Yet, no system of concepts will ever exhaust reality and as a consequence, this process will always be an

[9] McDowell, 1994, p.29.

[10] Willascheck, Marcus (1999), *"On "The Unboundedness of the Conceptual" in* Willascheck (ed.), 'John McDowell - *Reason and Nature - Lecture and Colloquium in Münster*', Münster: LIT press; http://web.uni-frankfurt.de/fb08/PHIL/willaschek/mcdowellkolloq.pdf.

[11] These aspects are mentioned in the comment Marcus Willascheck made after a conference held by McDowell, called "On "The Unboundeness of the Conceptual"" which can be found at the site mentioned in the previous reference. It is not our goal to comment Willascheck argument - although what he wants is to counter the thesis of conceptual. The argument serves only to illustrate the impossibility of making the whole world presented to us, and that this impossibility not only can be used as an argument against idealism but also allows the defense of the independence of reality from thought.

ongoing revising process, which will not come to an end. This inability to finish the process is due to the fact that new experiences necessarily prompt revision of our systems of concepts.

My goal in this section was to see McDowell's expression *things are thus and so* as key to understanding the role of sensibility. It is now time to develop another aspect that, as far as I can see, can make a powerful contribute to further understanding of this topic, and that is McDowell's interpretation of the transcendental deduction of Kant's Critique of Pure Reason. By doing this I want to show how McDowell tries to argue that one can objectively have reality in view and, if so, that (and how) one's sensibility can be about world.

TRANSCENDENTAL IDEALISM – A PATHWAY TO REACH THE WORLD?

In order to understand McDowell's solution, two main aspects of his interpretation of Kant's Critique of Pure Reason must be considered: one is to be found in the Metaphysical Deduction and the other in the Transcendental Deduction. In the third section of the Analytic of Concepts of Kant's Critique of Pure Reason - *The clue to the Discovery of All Pure Concepts of Understanding* (the so-called Metaphysical Deduction) – Kant tells us the following: "*the same function which gives unity to the various representations in a judgment also gives unity to the mere synthesis of various representations in an intuition; and this unity, in its most general expression, we entitle the pure concept of understanding*".[12] Endorsing this, McDowell wants to argue that it is because of the cooperation between sensibility and understanding, receptivity and spontaneity, that we can vindicate objective purport not only to intuitions - and experience - but also to judgment.[13]

To sustain this assumption, McDowell introduces a third element that can also be found in Kantian philosophy: the *I think* or Transcendental Unity of Apperception. In paragraph 18 Kant says: "*The transcendental unity of apperception is that unity through which all the manifold given in an intuition is united into a concept of the object*".[14] In *Mind and World*, this same idea is used when McDowell introduces one of his characteristic expression "*That things are thus and so*" – the multifaceted character of

[12] Kant, Immanuel (1781/1789), *Critique of Pure Reason* (from now on CPRu), trans. Norman Kemp Smith. London, Macmillan (1929), A79/B104-5.

[13] John McDowell (2009b), *"Hegel's Idealism as Radicalization of Kant"* in McDowell, *Having the World in View*, Cambridge, Mass.: Harvard University Press. (Selected Papers), pp. 70-73.

[14] CPRu, B139

this expression states that because *things are so and so*, not only are they presented as the content of experience but also as the content of a judgment, and, if one is not misled, as a layout of the world. According to McDowell's reading of Kant, it is possible to sustain a realist and thoroughly conceptual thesis about mind-world relations only and only if an interaction between the following three factors (intuition, judgment and the transcendental unity of apperception) occurs. He argues that "*by invoking the unity of apperception we enable ourselves to make sense of the objective purport of intuitions and the objective purport of judgment together. The Deduction elaborates the idea of a subjectivity that is both intuitionally in touch with objective reality and able to make judgments about it*".[15] Therefore, the logical structure of intuitions will be the same as that of judgment, which are linked by the *I think*, the unity of apperception.

To understand the strength of this, a closer look into the transcendental deduction is essential. The following question can then be made: why does the transcendental deduction play such an important role in Kant's argument and in McDowell's solution? To sustain McDowell's *bold* thesis – that co-operation between understanding and sensibility results in an objective content of thought directed to the world – one needs to overcome one objection. The objection is less concerned with the realm of understanding, where the pure concepts of understanding will carry out their unifying role of synthesis, and more with the fact that it is not possible to ensure the objectivity of pure forms of sensibility, i.e. space and time, in order to reject subjective idealism. McDowell believes that the B Deduction tries to argue for the objectivity of experience focusing only on pure concepts of understanding. Now the problem is that the deduction only ensures thought, or thinkability, as McDowell says[16]. However, when the question regards the conditions for an object to be given to our senses, difficulties begin to emerge. These difficulties are materialized in the thesis of transcendental idealism, in which space and time are presented as the pure forms of sensibility that organize, give shape, to intuition. According to McDowell, "*the transcendental Aesthetic has already supplied an independent condition for objects to be able to be given to our senses: they must be spatially and temporally ordered. For all Kant can show, objects could satisfy that condition for being present to our senses without conforming to the requirements of the understanding*"[17]. Thus,

[15] McDowell, 2009b, p. 71.

[16] To an understanding of McDowell's argument see "*Hegel's Idealism as Radicalization of Kant*" in McDowell, *Having the World in View*, pp 69-89.

[17] McDowell, 2009b, p. 73.

the problem that the deduction needs to deal with, according to McDowell, is the subjective imposition to which the Transcendental Ideality leads us by assuming an ontological difference, a gap, between things that are presented to our senses and things themselves.

McDowell considers that this objection renders vulnerable the objectivity that Kant intends to defend with his argument. McDowell argues that the way Kant overcomes this awkwardness is by denying that the Transcendental Aesthetic offers independent conditions for objects to be given to our senses. To McDowell, the solution for the problem should be that the "*capacities that belong to apperceptive spontaneity are actualized in intuitions. That goes in particular for the pure intuitions of space and time. So the formedness of our sensibility, the topic of Aesthetic, cannot after all be fully in view independently of apperceptive spontaneity. The unity constituted by conformity to the requirements of our sensibility, which is the unity of the formal intuitions of space and time, is not a separate unity, independent of the unity that consist in being informed by the categories*"[18]

However, the shadow of Transcendental Idealism remains and subjective idealism turns out to be the greatest threat to the Kantian notion of objectivity. Kant argues against this position by saying that the requirements of understanding are not only subjective but are requirements on the objects themselves. As Kant tells us, the conditions for the possibility of experience are the same conditions for the possibility of the objects of experience[19]. Combining this idea with the thesis that the synthetic unity of consciousness is the objective condition of all knowledge, it becomes possible for Kant to defend the objectivity of empirical knowledge, even with the brand of transcendental idealism still present, even if dissimulated.

Thus, experience can only reach the limits defined by the pure forms of intuition, its spatiality and temporality, suggesting that there is something outside the unifying power of apperceptive spontaneity, something like brute facts that are beyond space and time and therefore, beyond any cognitive range. This implies a particular form of the Myth of the Given. There is, then, the question of how, in this framework, it is possible to reach knowledge that is both a priori and genuinely objective. "*The harshest way to put this criticism is to say that though the Aesthetic purports to ground* a priori *knowledge that is objective, in the only sense we can make intelligible to ourselves, what it puts in place is indistinguishable from subjectivistic psychologism. Whatever is the case with the requirements that reflect the discursiveness of our intellect, the requirements that reflects*

[18] McDowell, 2009b, p. 74.
[19] CPRu, A157/B198.

how our sensibility is formed – the requirements of spatial and temporal ordering – look like subjective imposition. Transcendental Idealism, which is just this insistence that the apparent spatiality and temporality of our world derive from the way our sensibility is formed, stands revealed as subjective idealism".[20] Thus, the way sensibility is presented by Kant, according to McDowell's reading, affects the objective validity that is the cornerstone of the demonstration of pure concepts of understanding. In the B Deduction (B148-9), Kant tells us that categories are just ways of thinking, forms of thinkability, devoid of any content at all, without any objective validity, unless empirical intuition is presented by our forms of sensibility along with the synthetic unity of apperception in order to provide the content. This is then categorized by the understanding, leading to cognition, consequently "establish*(ing)* that experience has its objective purport"[21]. Nevertheless, a dependence of the pure forms of sensibility that constitute the limits of what can be given through our senses is always present.

Thus, if the aim of the Transcendental Deduction is to ensure that the requirements of both understanding and sensibility are objective conditions on objects themselves, the project is always dependent on Kant's Transcendental Idealism. It is this that ensures that the same objective conditions, which are also subjective conditions because our ability to know things can be achieved only as spatially and temporally, are mandatory by transcendental aesthetics. Thus, there is always something outside the conceptual, something that can never be presented to our cognitive activity, and so as a consequence, can never be touched by our minds. In other words "*if we cannot know whether things themselves are really spatially and temporally ordered, that undermines the possibility of recognizing as knowledge the supposed knowledge we are supposed to be able to achieve within the boundary. That in turn ensures that the deduction cannot succeed in vindicating a genuine objectivity for the requirements of the understanding*"[22]

What McDowell argues is that Kant has almost managed to achieve his purpose. The reason why this aim was not accomplished is due to the conditions imposed by transcendental aesthetic – the *a priori* forms of space and time and the doctrine of transcendental idealism. To overcome this problem, McDowell defends a destranscendentalization of Kant's

[20] McDowell, 2009b, p. 76.

[21] McDowell, 2009b, p. 73.

[22] McDowell, 2009b, p. 79 The position presented in the first section tries to show that the constraint between the mind and the world is indeed necessary; but he argues that this constraint arises from outside the thought and not from outside the thinkable content.

Transcendental Aesthetic[23]. But what does McDowell mean by this? Well, what McDowell wants to say is that if we remove space and time, the ideal forms, from the process of knowledge and knowing the world by a thinker, nothing will be left outside the unity of apperception, of the *I think,* and one's mind's relation with objects will be a direct one, with no intermediaries mediating that encounter. As such, the co-operation between sensibility and understanding can be conceived as genuinely objective, about the world. By arguing along these lines, McDowell can then say that our experience of the world by means of sensibility can, in fact, reach the world and we can make judgments about it, which, if not misleading, can in fact be a layout of the world. Through sensibility one has, thus, the world in view.

By considering these two aspects of McDowell philosophy, I will now make an attempt to make clear the role of sensibility through McDowell's work.

EXPERIENCING THE WORLD – THE ROLE OF SENSIBILITY

1. In his seminal work, Mind and World, McDowell explores the co-operation that he believes exists between sensibility and understanding, receptivity and spontaneity. Clearly influenced by Kant, McDowell's diagnosis led him to assert that philosophy suffers from an anxiety that arises from a clash between two ways of thinking about the nature of experience, i.e., the idea of a normative tribunal (the tribunal of reason) and the idea of experience as stimulation of the senses. Thus, McDowell, influenced by Sellars - in particular by his distinction between "the logical space of reasons" and "logical space of nature," the space of causality –, wants to show that the metaphysical question about mind-world is poorly worded and generates two different solutions that leave the world untouched by the mind: either subscribing to the Myth of the Given, the solution inspired by Quine, or a coherentist framework, as that developed by Davidson. The proposal presented in *Mind and World* tries to overcome this dilemma. McDowell's claim can then be synthesized under the above discussed motto "*That things are thus and so*".

It is in the second lecture of his book that McDowell presents his thesis of the co-operation between sensibility and understanding. He does this by questioning whether Kant, in the Critique of the Pure Reason, considers any role for sensibility besides the mentioned co-operation

[23] McDowell, 2009b, pp. 17-18. See also note 26.

with the understanding in order to justify the objectivity of experience and knowledge. McDowell gives a double answer: he answers "yes" and "no". "No," because from the standpoint of experience, in providing an independent role to receptivity "*one commits oneself to something Given in experience that could constitute the ultimate extra-conceptual grounding for everything conceptual.*"[24] Now, what it is being questioned in the diagnosis is whether it is reasonable to suppose that, by the mere impact of the senses outside the sphere of the conceptual, experience can have objective validity or, as mentioned above, that experience should be defined as openness and responsiveness to the world. According to McDowell, "*in experience we take in, through impacts on the senses, elements in a reality that is precisely not outside the sphere of thinkable content*"[25]. Subsequently, special emphasis is given to this notion of "thinkable content", since it is central to McDowell's metaphysical solution.

However, McDowell believes that it is possible to give a positive answer to the question. To this end, the philosopher tells again the story of transcendental philosophy in which receptivity appears to be susceptible to the impact of a supersensible reality, a reality that is characterized by its independence from any conceptual activity[26]. This supersensible reality, viewed from the outside, or from nowhere, presents a difficulty that seems obvious: by endorsing a supersensible reality, beyond any conceptual content, sensibility, with its *a priori* forms of time and space – the thesis of transcendental idealism – will necessarily entail subjective idealism which, according to McDowell, is a manifestation of the Myth of the Given. As a corollary, receptivity, in addition with its co-operation with spontaneity, also has a space of its own: non-conceptual, sheer receptivity, on which intuition flows. I will come back to this.

Nevertheless, having receptivity in co-operation with spontaneity ensures access to reality without any trace of idealism. So, according to McDowell, "*if we restrict to the standpoint of experience itself, what we find in Kant is precisely the picture I have been recommending: a picture in which reality is not located outside a boundary that encloses the conceptual sphere (…) the fact that experience involves receptivity ensures the required constraint from outside thinking and judging. But since the deliverances of receptivity already draw on capacities that belong to spontaneity, we can coherently suppose that the constraint is rational.*"[27]

[24] McDowell, 1994, p. 41.
[25] McDowell, 1994, p. 41.
[26] McDowell, 1994, p. 41.
[27] McDowell, 1994, p. 41.

This is the main thesis that McDowell advocates about the cooperation between sensibility and understanding. My aim here is to assess the role of sensibility and how the notion evolves in McDowell's work.

2. In *Mind and World*, McDowell does not explicitly address the consequences of the idea of cooperation between sensibility and understanding. It will be through the philosophical work of Sellars and Kant that this vindication will be made.

Although "Empiricism and the Philosophy of Mind"[28] was the seminal work of Sellars, and the root of some of the main theses of McDowell that developed and presented in *Mind and World*, it will be in "*Science and Metaphysics: Variations of Kantian Themes*", Sellars main collection of essays published ten years later, that McDowell will find the strengthening of his philosophical assumptions. Nevertheless, the core of McDowell's argument can in fact be found in Sellars' major work; in fact we can find in the beginning of that paper a Hegelian orientation that will be the core of Sellars' attack to the "Framework of the Givenness".

According to McDowell, the greatest revolution that Sellars' brought to philosophy was the following: "*There is a special category of characterizations of states or episodes that occur in people's corresponding characterizations of the people in whose lives the states or episodes occur, for instance, characterizations of people as* knowers."[29] Sellars calls this type of categorizations the "logical space of reasons", as opposed to the "logical space of nature." Thus, while the first is the ability to justify, *to give reasons for*, the second takes us to the field of science, into the realm of causality, scientific law. One can then say that for Sellars it is possible and feasible to divide and isolate the logical space of reasons and the logical space of nature, into specific and ontologically different realms.[30] According to McDowell, this artificial division is of special importance for Sellars because it will prevent intrusions from both spaces, one into the other, thus avoiding category mistakes, which inevitably end in the adoption of the myth of the given, whereby justification and causation will be mysteriously intertwined. So the main purpose of Sellars is to reveal the existence of a difference between epistemic facts and natural facts.[31] Following Sellars

[28] Wilfrid Sellars (1956). *Empiricism and the Philosophy of Mind*, Edited by Robert Brandom (1997), Cambridge, Harvard University Press.

[29] John McDowell (2009b), "*Sellars on Perceptual Experience*" in McDowell, *Having the world in a view*, Cambridge, Mass.: Harvard University Press. (Selected Papers), p. 4.

[30] McDowell, 2009b, p. 5

[31] McDowell, 2009b, p. 5. By natural capacities, Sellars means the "capacities that their subjects have at birth, or acquire in the course of merely animal maturation".

thesis, this ontological difference between these two logical spaces lead us, among other things, to the problem of perceptual experience. This particular problem is intrinsically connected with the role of sensibility and its integration into the experience of the world.

Therefore, Sellars' intention, in McDowell's interpretation, is the following: "*to arrive at an acceptable picture of how the sensory and the conceptual – sensibility and understanding – combine so as to provide for the intentionality of perceptual experience, and to provide for how perceptual experience figures in the acquisition of a knowledgeable view of the world*"[32]. In order to understand how these two structures interrelate, it becomes necessary to use an image of an imaginary line that divides them: *above the line*, we find, in perceptual experience, (Sellars uses the example of visual perceptual experience) a conceptual contribution that will allow us to have beliefs about the world, but beliefs of a special kind - conceptual beliefs, beliefs that allow us access to a world as it is, the world in itself. This reveals a very specific nature of the conceptual: these conceptual experiences already show the kind of co-operation between sensibility and understanding that McDowell tries to defend. According to McDowell, "*the above-the-line episodes that figure in Sellars's picture of visual experience are, as conceptual episodes of their special kind, already conceived as conceptual shapings of sensory, and in particular visual, consciousness*".[33]

Sellars distinguishes those beliefs (conceptual beliefs) from ostensible beliefs, beliefs without any conceptual content – this distinction is an epistemological one, and it is important because the *below-the-line* space of our imaginary line is full with a multiplicity of perceptive sensations, sensations that are devoid of any conceptual content.[34] The question that arises is why Sellars needs a space below the line, a dimension of pure receptivity when the elements above the line already have a manifest, explicit, co-operation between understanding and sensibility, giving the latter a status of openness and responsiveness to the world. This question is relevant inasmuch as Sellars tries to defend the existence of this dimension of pure sensibility, while McDowell objects to it.

To vindicate a role to sheer receptivity, Sellars appeals to a transcendental argument in order to sustain the need for the *above-the-line* space of the imaginary line. Here is how McDowell presents his argument: "*the idea is that we are entitled to talk of conceptual episodes in which*

[32] McDowell, 2009b, p. 9.

[33] McDowell 2009b, "*The logical from of an intuition*" in McDowell, *Having the world in a view*, Cambridge, Mass.: Harvard University Press. (Selected Papers), p. 23.

[34] McDowell, 2009b, p. 23.

claims are ostensibly visually impressed on subjects – the above-the-line element in the picture – only because we can see the flow of such conceptual representations as guided by manifolds of sensations: non-concept-involving episodes or states in sensory and specifically visual, consciousness".[35] Based on this idea, Sellars argues that it is only because there is a sheer receptivity, without any kind of conceptual content, where the multiplicity of intuitions flows, that our perception can be about the world.

However, for McDowell, this reading of Sellars – a reading which he (Sellars) argues is what Kant should have said but did not say –, is not defensible, since it entails a mysterious and inexplicable passage of raw impact on our sensibility to a cognitive state of understanding. Thus, the *below-the-line* space constitutes a barrier, an obstacle to how perception and thought are directed towards objects and, consequently, to the world. It is now time for a reflection about the role of intuition, as a central element of sensibility.

3. Kant refers to intuition as follows: "*In whatsoever mode, or by whatsoever means, our knowledge may relate to objects, it is at least quite clear that the only manner in which it immediately relates to them is by means of an intuition. To this as the indispensable groundwork, all thought points. But an intuition can take place only in so far as the objects are given to us. This, again, is only possible, to man at least, on condition that the objects affect the mind in a certain manner*"[36]. By considering this operational definition of Kant, what can be first said is that this is a clear example of the Myth of the Given, mainly because of Kant's thesis that the objects are given to us. Yet it can be shown that this observation is only accurate in a surface level.

Sellars argues that what Kant means by intuition is something like "*representations of individuals that already involve the understanding, the faculty associated with concepts (…) be taken to represent an individual as this such (…) we might describe intuitions on this interpretation as shapings of sensory consciousness by the understanding*"[37]. The cooperation between sensitivity and understanding is evident here and it is clearly located on the *above-the-line* image (note that this thesis is central to my argument). This argument is not only compatible with McDowell's solution but is also an alternative to the Myth of the Given.

[35] McDowell, 2009b, pp. 23-24.
[36] CPRu, A17/B 33.
[37] McDowell, 2009b, p. 24.

Nonetheless, Sellars isolates a specific mode of intuition, which he calls "*Sheer Receptivity*". McDowell tells us that "*Sellars is convinced that Kant also needs to speak about sensibility in a way that belongs below his line, as the talk of sensory consciousness with which we can gloss this first notion of intuition does not, because in intuitions, on this first interpretation, sensory consciousness is already shaped by the faculty of concepts. Sellars think the transcendental role that Kant needs sensibility to play consists in its supplying manifolds of sensory items that are not shaped by the understanding, to guide the flow of conceptual representations in perception*".[38] Thus considered, intuition will play a dual role in the logic of experience as openness to the world. The latter – sheer receptivity – manifests itself as a mediator, an element capable of synchronizing the multiplicity of representations whose main aim will be the unification of an intuition; in the course of the workings of transcendental machinery this will generate an empirical, cognitive knowledge. There is, however, the question of how this is possible.

These two interpretations of intuition are, according to Sellars, both possible, although Kant does not mention the second one, because intuition is integrated with the understanding by means of transcendental imagination and transcendental schematism. This is something Sellars will try to integrate in his own interpretation. It seems, however, that Sellars' error was to consider Kant's transcendental aesthetic isolated from the remaining transcendental machinery and, therefore, consider that it may play an essential role, even mysterious, on acquisition of cognitive content. Sellars believes that Kant did not see the exact role that, in general, intuition and sensibility need to play in the whole project of the Critique and, more specifically, considering our aim, the objective experience of the world. McDowell, however, tells us that when considering this specific aspect of intuition, Sellars omits something crucial, when he puts aside the interactive vector that goes along with Kant's reflection in the Critique of Pure Reason. According to McDowell, "*we are supposed to account for the outwardness of outer sense by invoking space as an autonomous form of sensibility, intelligible independently of any involvement on the part of the understanding. When Kant then brings the understanding into play, in the Transcendental Analytic, the outwardness that, on this reading, the Aesthetic has already provided for takes on a new from, as directedness towards determinate* objects. *On this reading, space as the Aesthetic considers it would after all do its informing work below something corresponding to Sellars's line, with operations of the understanding above the line*".[39] And

[38] McDowell, 2009b, p. 25.
[39] McDowell, 2009b, p. 27.

quoting Sellars in *Science and Metaphysics* "*the characteristics of the representations of receptivity as such, which is what should properly be meant by the forms of sensibility, are never adequately discussed, and the so-called forms of sensibility become ever more clearly, as the argument of the Critique proceeds, forms of conceptual representations*".[40] Sellars is thus convinced that a proper interpretation of Kant's thesis requires pure receptivity. The reason Sellars assigns such an essential role to pure sensibility is due to the fact that he argues that all empirical cognition has to be constrained by something external to the cognitive activity in order to prevent the fall into some kind of idealism. However, the constraint must indeed exist but should be located outside thinking and judging and not outside thinkable contents.[41]

Sellars gives a transcendental role to sensibility because he believes this is an essential requirement to achieve the Critique's main aim. However, McDowell thinks that by defending such a position – i.e. giving a specific role to sensibility, beyond its cooperation with the understanding –, one falls into a subjective psychologism where the transcendental forms of space and time will produce a gap, an ontological gulf, between the *object-to-me* and the *object in itself*. According to Sellars, "*for thought to be intelligibly of objective reality, the conceptual representations involved in perceptual experience must be guided from without*".[42] Sellars argues that there is a non-conceptual space that, somehow, will make not only the conceptual possible, but also the directedness towards the world, the openness to the world. It seems to me that this interaction, a mysterious, magic one, that Sellars intends to defend when he assigns more than one role, the decisive transcendental role, to sensibility, begins to collapse, though in a different way than what was outlined in *Empiricism and the Philosophy of Mind*, which focuses on the critique of the multiple forms of sense-data. "*In characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing*

[40] Sellars, Wilfrid (1967), *Science and Metaphysics: Variations of Kantian Themes*. London: Routledge and Kegan Paul (reissued, Atascadero, California.: Ridgeview, 1992), p. 30.

[41] The reason why Sellars argues his position against McDowell is: "Sellars thinks the ordinary objects that seem to be present to consciousness in perceptual intuition are strictly unreal. He thinks "scientific realism" requires this deniel of reality to those constituents of, as he puts it, the manifest image. It cannot be those merely apparent ordinary objects that are the source of the required constraint from an external reality. What do really exist are the constituents of the scientific image that correspond to those merely apparent ordinary objects: swarms of elementary particles or something of the sort. Empirical cognition can be subject to genuinely external constraint only by way of impacts on our senses from those genuinely real items. Sellars puts his forward as an interpretation for Kant's distinction between phenomenal objects, constituents of the manifest image, and things in themselves, which Sellars identifies with constituents of the scientific image" John McDowell (2009d), *"Self-determining subjectivity and external constraint."* in McDowell, *Having the world in a view*, Cambridge, Mass.: Harvard University Press. (Selected Papers), pp. 98-99.

[42] McDowell, 2009b, p. 39.

it in the logical space of reasons, of justifying and being able to justify what one says"[43].

Until now I have been discussing the topic of sensibility by evoking a double role that Sellars gives to it, and by presenting McDowell's main objections. Now I think is important to discuss why McDowell argues in favour of a cooperation of sensibility and understanding. To do that, it's important to remember an aspect that was mentioned before about the topic of transcendental idealism. By arguing around the *Clue*, McDowell says that the logical structure of intuitions is the same as that of judgment, and that both of them are linked by the *I think*, the unity of apperception. As such, sensibility cannot have any particular role independently of understanding, not if one wants to argue that one can have direct access to the world through perception and judgments. By linking this argument with the argument of the unboundeness of the conceptual, McDowell tries to argue that one's experience of the world, if we are not mistaken, if everything goes well, has the following result: "the idea of conceptually structured operations of receptivity puts us in a position to speak of experience as openness to the layout of reality. Experience enables the layout of reality itself to exert a rational influence on what a subject thinks".[44]

4. In his essay "*Avoiding the Myth of the Given*", McDowell makes an attempt to clarify his notion of experience and sensibility, in a discussion with Charles Travis and the possibility or not to sustain the thesis that we can find conceptual content in our perceptual experience. I will dedicate this next section to understanding how these conceptions evolved in McDowell's thought, thus adding some more pieces to this transcendental puzzle.

According to McDowell, in this last mentioned article, the role of sensibility is directly related with his particular way of defining the Myth of the Given. According to him, "the Myth is the idea that sensibility by itself could make things available for the sort of cognition that draws on the subject's rational powers".[45] Until now, McDowell did not seem to move away from his initial remarks, mainly his Kantian argument and his objections to Sellars' thesis about the specific role of sensibility in our

[43] McDowell, John, (2009a) *"Naturalism in the Philosophy of Mind"*, in McDowell, *The Engaged Intelect,* Cambridge, Mass.: Harvard University Press, p. 257. McDowell quotes directly from Sellars - Wilfrid Sellars (1956). *Empiricism and the Philosophy of Mind*, edited by Robert Brandom (1997), Cambridge, Harvard University Press. See also Richard Rorty (1979). *Philosophy and the Mirror of Nature*. Princeton University Press, Princeton.

[44] McDowell, 1994, p. 26.

[45] McDowell, John (2009b), *"Avoiding the Myth of the Given"* in McDowell, *Having the world in a view*, Cambridge, Mass.: Harvard University Press. (Selected Papers), p. 257.

experience of the world. Thus, considering the requirement that states that in order for one to have knowledge of the world one must consider some kind of cooperation between sensibility and understanding, it is essential that our capacities that belong to reason be present not only in judgment but also in experience itself. By claiming this, McDowell wants to reject a possible misinterpretation that states that experiencing implies that "rational capacities are operative only in responses to experience, not to experiences themselves."[46] This is how McDowell's thinks that the myth should be overcome.

However, in this article, McDowell felt the need to make a revision of some of the claims that he once thought to be insightful, mainly the claim that experience has propositional content. Since my aim is to discuss the plausibility of his conception of the role of sensibility, I think that it is important to understand this last move of McDowell's. In order to do this, let's remember Kant's slogan that McDowell seems to take so seriously: "*the same function which gives unity to the various representations in a judgment also gives unity to the mere synthesis of various representations in an intuition; and this unity, in its most general expression, we entitle the pure concept of understanding*". According to this statement, there should be some kind of correspondence between the propositional content of judgment (discursive content) and an intuition. Following McDowell, a judgment like "This is a cube", must have correspondence with the intuition "This cube"[47]. So, because McDowell claims that receptivity and spontaneity cooperate and that sensibility alone cannot have an independent role in the process, then experience should only have a propositional content. But this was not McDowell's final word.

So, what has changed in McDowell's claims? As I said, it is due to Charles Travis that the change happens. The change is presented by the expression *Having the world in view*.[48] What this implies is a new look on the notion of intuition. What Travis[49] brings to McDowell's thought is the idea that intuitional content cannot be conceptual and one cannot have access to things themselves. So "*When Travis says experiences do not represent things as so, he does not mean that experiences are intuitions in the sense I have been explaining. He says that experience is not a case*

[46] McDowell, 2009b, p. 258.

[47] McDowell, 2009b p. 260.

[48] Interesting fact: this is the name of McDowell's book. I think that through the book, McDowell tries to establish the evolution of his thought concerning the problem of experience and makes an attempt to explain how the world becomes open to us. The question is: Can it?

[49] The remarks about Charles Travis thought that one is using are from McDowell's interpretation. It is not my aim to assess Travis work. My aim is to show how McDowell's reading of Travis lead him change is way of seeing the problem of experience.

of intentionality, and I think it is fair to understand him as denying that conceptual capacities are in play in experience at all (…) In Travis's picture conceptual capacities are in play only in our making what we can of what visual experiences anyway bring into view for us, independently of any operation of our conceptual capacities. In Travis's picture, having things in view does not draw on conceptual capacities, having things in view must be provided for by sensibility alone".[50] According to this argument, *that things are thus and so* (McDowell's thesis) would not make any sense because sensibility would have a specific role that goes beyond the co-operation with the understanding.

The discussion with Travis forces McDowell to revise on the notion of intuition. So, the question that McDowell now poses is: what kind of content can one find in intuition? If not a propositional one, a discursive one – following Kant's terminology – what kind of content should one find in intuition that would not put at risk the main thesis of McDowell and the co-operation between sensibility and understanding? This is an important question because, according to McDowell, although Travis is right about the problem of intuition, his solution is a manifestation of the Myth of the Given.[51]

McDowell agrees with Travis in thinking that experience brings our surroundings into view – *having the world in view*. But he disagrees that experience does not have any kind of conceptual content. So the change in McDowell's way of seeing intuition is that he now thinks that its content is not propositional, although it still remains conceptual. But how can this be? This is how McDowell solves the problem: "*though they are not discursive, intuitions have content of sort that embodies an immediate potential for exploiting the same content in knowledgeable judgments. Intuitions immediately reveal things to be the way they would be judged to be in those judgments*"[52]. By arguing like this, McDowell not only states that sensibility does not have any special role independent of understanding, but also that intuitions – experience – "*entitle us to judgments that would exploit some of the content of an intuition, and it figures in our entitlement to judgments that would go beyond that content in ways that reflect capacities to recognize things made present to one in an intuition. But as I have insisted, in intuiting itself we do not deal discursively with content*"[53]. Therefore the

[50] McDowell, 2009b, p. 267.

[51] "Travis thinks the idea that experiences have content conflicts with the idea that experience directly brings our surroundings into view (...) wanting, as is reasonable, to keep the idea that experience directly brings our surroundings in to view, he is led to deny that experiences have content". (*Idem*, p. 268)

[52] *Idem*, p. 267.

[53] McDowell, 2009b, p. 270.

main aim of McDowell's thesis remains untouched and the Myth of the Given is avoided.[54]

One last aspect must be considered: the potential for a discursive activity. This aspect is important because the idea of having the world in view does not mean that everything that passes through sensibility – through the form of an intuitional content – becomes judgment – a discursive content. Although both contents should be conceptual, the former is an unconscious process that has the potentiality to become a discursive claim[55]. By arguing this, it seems that McDowell falls in the pitfall that he is trying to avoid. But what he is arguing is that, following Kant, sensibility is not alone: if "the same function which gives unity to the various representations in a judgment also gives unity to the mere synthesis of various representations in an intuition" then both sensibility and understanding are present in the process. What I think is important to notice is that "*it is in the intuition in a form in which one could make it, that very content, figure in discursive activity. That will be to exploit a potential for discursive activity that is already there in the capacities actualized in having an intuition with that content*"[56]. So, the fact that every intuitional content may not became discursive content, is not an argument for the Myth of the Given, since both contents are conceptual and can be, according to McDowell, both incorporated in the slogan that is the cornerstone of McDowell's main thesis: *that things are thus and so*.

The problem is that the marks of Kantian transcendental idealism remain. Thus, it seems that the world is still locked to us[57].

[54] *"(...) the conceptual content that allows us to avoid the Myth is intuitional, not propositional, so experiencing is not taking things to be so. In bringing our surroundings into view, experiences entitle us to take things to be so: whether we do is a further question"* (*Idem*, p. 269).

[55] To a more detailed interpretation of McDowell's argument see "Avoing the Myth of the Given", p. 264. About the notion of unconsciousness see also John McDowell (2009c), *"Hegel's Idealism as Radicalization of Kant"* in McDowell, *Having the World in View*, Cambridge, Mass.: Harvard University Press. (Selected Papers), pp. 71-72.

[56] McDowell, 2009b, p. 265.

[57] Travis, C. (forthcoming). *Unlocking the Outer World*. http://mlag.up.pt/.

Kant, Immanuel (1781/1789), *Critique of Pure Reason,* trans Norman Kemp Smith. London, Macmillan (1929)

McDowell, John (1994), *Mind and World*, Cambridge, Mass.: Harvard University Press

McDowell, John, (2009a), "*Experiencing the World*", in McDowell, *The Engaged Intellect*, Cambridge, Mass.: Harvard University Press, p. 243

— (2009a) "*Gadamer and Davidson on Understanding and Relativism*" in in McDowell, *The Engaged Intelect,* Cambridge, Mass.: Harvard University Press

— (2009a) "*Naturalism in the Philosophy of Mind*", in McDowell, *The Engaged Intelect,* Cambridge, Mass.: Harvard University Press

McDowell, John (2009b), "*Avoiding the Myth of the Given*" in McDowell, *Having the world in a view*, Cambridge, Mass.: Harvard University Press. (Selected Papers)

— (2009b), "*Hegel's Idealism as Radicalization of Kant*" in McDowell, *Having the World in View*, Cambridge, Mass.: Harvard University Press. (Selected Papers)

— (2009b), "*Sellars on Perceptual Experience*" in McDowell, *Having the world in a view*, Cambridge, Mass.: Harvard University Press. (Selected Papers)

— (2009b), "*The logical from of an intuition*" in McDowell, *Having the world in a view*, Cambridge, Mass.: Harvard University Press. (Selected Papers),

Miguens, Sofia (2008), *Será que a minha mente está dentro da minha cabeça – Da ciência cognitiva à Filosofia*, Porto, Campos das Letras Editores,

Sellars, Wilfrid (1956). *Empiricism and the Philosophy of Mind*, Edited by Robert Brandom (1997), Cambridge, Harvard University Press.

Sellars, Wilfrid (1967), *Science and Metaphysics*: *Variations of Kantian Themes*. London: Routledge and Kegan Paul (reissued, Atascadero, California.: Ridgeview, 1992), p. 30

Travis, C. (forthcoming). *Unlocking the Outer World*. http://mlag.up.pt/

Willascheck, Marcus (1999), "*On "The Unboundedness of the Conceptual" in* Willascheck (ed.), 'John McDowell - *Reason and Nature - Lecture and Colloquium in Münster*', Münster: LIT press; http://web.uni-frankfurt.de/fb08/PHIL/willaschek/mcdowellkolloq.pdf.

MEETING OTHER MINDS

Tero Vaaja

Descartes argued that human beings have rational souls, while animals are automata, functioning in a wholly mechanistic fashion. How do we know that God has joined a soul with each human body? In *Discourse on the Method*, part V, Descartes mentions two ”tests of a real man”:

> “[I]f any […] machines had the organs and outward shape of a monkey or of some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals; whereas if any such machines bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could never use words, or put together other signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words which correspond to bodily actions causing a change in its organs (e.g. if you touch it in one spot it asks what you want of it, if you touch it in another it cries out that you are hurting it, and so on). But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do. Secondly, even though such machines might do some things as well as we do them, or perhaps even better, they would inevitably fail in others, which would reveal that they were acting not through understanding but only from the disposition of their organs. For whereas reason is a universal instrument which can be used in all kinds

> of situations, these organs need some particular disposition for each particular action; hence it is for all practical purposes impossible for a machine to have enough different organs to make it act in all the contingencies of life in the way in which our reason makes us act." (CSM I, 139-140; AT VI, 56-57)

This is sometimes (e.g. Plantinga 1967) taken to suggest that Descartes endorsed what is now called an analogical argument for the existence of other minds. However, as Avramides (1996, 2001) notes, Descartes seems to think that only a single judgment is enough to assure us that a fellow human has a mind, because it is a fundamental assumption for him that all and only beings that have the human shape and form are endowed with a mind. This fundamental assumption prevents Descartes from seeing a need for an analogical argument of the form "this body is a human body (i.e. it moves and talks like human bodies do); therefore, it very probably has a mind".

With his separation of material things from thinking things, Descartes anyway opened up the logical possibility of a mechanical, material body without a mind joined together with it – that is, the logical possibility of a human-shaped automaton. When his aforementioned fundamental assumption is dropped, it seems that the best we can do to justify our everyday belief in the existence of minds in others is to entertain some kind of argument from analogy. This is the basic (and perhaps the most commonsensical) way of explaining how we know about other minds.

A usual point of reference for this argument is John Stuart Mill's exposition of it:

> "I conclude that other human beings have feelings like me, because, first, they have bodies like me, which I know, in my own case, to be the antecedent condition of feelings; and because, secondly, they exhibit the acts, and other outward signs, which in my own case I know by experience to be caused by feelings. I am conscious in myself of a series of facts connected by a uniform sequence, of which the beginning is modifications of my body, the middle is feelings, the end is outward demeanor. In the case of other human beings I have the evidence of my senses for the first and last links of the series, but not for the intermediate link. I find, however, that the sequence between the first and last is as regular and constant in those other

> cases as it is in mine. In my own case I know that the first link produces the last through the intermediate link, and could not produce it without. Experience, therefore, obliges me to conclude that there must be an intermediate link; which must either be the same in others as in myself, or a different one: I must either believe them to be alive, or to be automatons: and by believing them to be alive, that is, by supposing the link to be of the same nature as in the case of which I have experience, and which is in all other respects similar, I bring other human beings, as phenomena, under the same generalizations which I know by experience to be the true theory of my own existence." (*An Examination of Sir William Hamilton's Philosophy* (1865), quoted through Malcolm 1958, 969)

For many thinkers, at least those who hold that minds have an essentially subjective element that cannot be reduced into any publicly observable physical facts, the argument from analogy is still the best available account of our knowledge of other minds (e.g. Chalmers 1996, 246). Anyway, as has been long pointed out, it is problematic in a number of ways. The standard criticisms against the analogical argument point out that it is an inductive generalization based on a single case, as well as an inference to an uncheckable conclusion (see Hyslop 1995, chapter 4, for a critical examination of these issues). But in addition to these, the analogical argument also gives rise to a deeper problem having to do with solipsism and the possibility of intersubjective understanding. The argument from analogy assumes that I first come across the notion of experience (mental states) first-personally, as referring to what I feel and experience, and then extend that notion to cover others. But it seems questionable whether, assuming that I only have experiences that are *mine* to start with, I will be able to form as much as an idea of experiences that are *not mine*. So we encounter a problem put forward by Wittgenstein:

> "'If you pity someone for having pains, surely you must at least *believe* that he has pains'. But how can I even *believe* this? How can these words make sense to me? How could I even have come by the idea of another's experience if there is no possibility of any evidence for it?" (Wittgenstein 1958, 46)

This aspect of the problem can be referred to as the *conceptual* problem of other minds, distinguishable from the pure epistemological problem. And finally, insofar as the problem of other minds is felt to be a

problem with real philosophical weight, it is arguable whether the argument by analogy can in any case be a satisfactory solution to it. The analogical argument concludes that the data provided by my own case gives me a fairly good reason to *assume* that others have mental states too. This ends up as an uncomfortably *detached* view of our relation to others: minds of others lie somewhere beyond what is manifestly in view for us, and we have to infer ourselves into knowledge about them. If our best justification for the existence of other minds only allows us to *inductively infer* or *postulate a theory* that others *probably* have minds, one might be left with a strong feeling that the core of the problem has not yet been touched upon at all. Being left with an answer that gives only probability, albeit a very high probability, is almost as bad as being faced with the original problem. The argument from analogy seems to keep us from meeting on the outermost surfaces of our bodies.

In a way, this final point against analogical argument has an existential flavor to it. It draws attention to the actual phenomenology of reacting to others as conscious beings. In clear cases of attributing a phenomenal experience – sensation, feeling or such – to another human being, we are not aware of first making a judgment about their physical movement, then an inferential step to a mental state. Of course, it is possible to say that the analogical argument is not necessarily the way we *in practice* proceed when attributing experiences to others; it is merely the way to epistemologically justify our beliefs about those experiences, should we feel the need of such a justification. But this distinction seems an important one to make, since it underlines the fact that in practice, we respond to certain kinds of bodily movements and sounds directly, taking them as manifestations of mentality in their own right, rather than something that allows us to *infer* the presence of mentality.

There is a way of approaching this problem, finding its place of origin in the later works of Wittgenstein, that declares the analogical argument to be wrong-headed and attempts to draw a more adequate picture of our epistemic relation to others. This approach states that some behavioral features serve as *criteria* for mental states. This is taken to mean that these behavioral features are somehow logically linked to mental states, so that they provide a special kind of evidence for such states: they serve as a non-inductive and immediate way of telling that the other is minded.

Wittgenstein uses the notion of criterion in contexts where he is rethinking the role of private objects of experience, stating that "an 'inner process' stands in need of outward criteria" (Wittgenstein 1953, §580). This

kindled an idea that the relation between "inner processes" (experiences) and patterns of behavior typical to them is an intimate one, so that access to the latter could be seen also as providing an access to the former. Bruce Aune characterized in 1963 this Wittgenstein-inspired idea that had been embraced in the then-recent discussion about other minds:

> "The traditional assumption that we must make a weakly-justified ontological leap when, on the basis of a person's observed behavior, we conclude that he is having a certain feeling or sensation is completely erroneous. The truth of the matter is rather this: the things we call 'psychological states' are so intimately connected with certain patterns of observable behavior that the occurrence of the latter provide us with criteria that are logically adequate for determining the presence of the former." (Aune 1963, 187)

What is it for something to be a logically adequate criterion for another thing? As John McDowell (1982) notes, Wittgenstein's own remarks do not give the impression that he is giving "criterion" or "Kriterium" a special technical meaning, except in a singular passage in the *Blue Book*, where he introduces the term to make a sharp distinction between criteria and symptoms:

> "Let us introduce two antithetical terms in order to avoid certain elementary confusions: To the question 'How do you know so-and-so is the case?', we sometimes answer by giving 'criteria' and sometimes by giving 'symptoms'. If medical science calls angina an inflammation caused by a particular bacillus, and we ask in a particular case 'why do you say this man has got angina?' then the answer 'I have found the bacillus so-and-so in his blood' gives us the criterion, or what we may call the defining criterion of angina. If on the other hand the answer was, 'His throat is inflamed', this might give us a symptom of angina. I call 'symptom' a phenomenon of which experience has taught us that it coincided, in some way or other, with the phenomenon which is our defining criterion. Then to say 'A man has angina if this bacillus is found in him' is a tautology or it is a loose way of stating the definition of 'angina'. But to say, 'A man has angina whenever he has an inflamed throat' is to make a hypothesis." (Wittgenstein 1958, 24-25)

The thing that Wittgenstein is calling the criterion for angina in this scenario is its defining criterion. We are able to tell that the patient has angina by observing the bacillus just because it has been defined in language that having the bacillus *is what it means* to have angina. Here the presence of the bacillus is surely a "logically adequate" criterion for angina, because "all people with angina have the bacillus so-and-so in their blood" is an analytic statement. Having the bacillus entails having angina. But this use of "criterion" in the sense of defining criterion is not representative of Wittgenstein's use of the term elsewhere; in other contexts, he seems to be mean by "criteria" conventionally stipulated, adequate ways of telling that something is the case, but such that it is always at least in principle possible that further evidence puts it into question whether that something is indeed the case. According to this latter reading, the relation between the criterion and what it is the criterion of is not an entailment. This latter reading is also the one that is interesting in the context of other minds. Saying that certain patterns of observable behavior are logically adequate criteria for mental states in virtue of being their *defining* criteria amounts to logical behaviorism; such position defines minds so that they are fully observable, so that any special problems about our access to the minds of others find no place to start with.

When the idea of Wittgensteinian criteria is invoked to explain how our access to the behavior of others could also serve as an access to their mentality, the relevant criteria are accordingly taken to be *defeasible* ones. That is, the satisfaction of criteria for X's being the case is compatible with X's not being the case. It is always possible that someone displaying typical e.g. pain-behavior is actually not in pain (he may be pretending or play-acting, or lacking pain-experiences for some other reason); the hypothesis of his pretending or otherwise lacking pain-experiences can be, depending on the circumstances, more or less implausible, but in no case is there anything *contradictory* in it. However, the relation between the criterion and its corresponding state of affairs is taken to involve more than just *symptoms* or *signs* of that state of affairs. The behavioral criteria for pain, for example, are taken to be such criteria, in some sense, as a matter of "logical" fact.

In Wittgenstein's terminology, that certain behavioral features serve as criteria for someone's being in pain is part of the "grammar" of our pain-language. This makes it seem like a matter of linguistic convention that people displaying pain-behavior are said to be in pain; as if when we are talking about sensations and experiences, our talk is equivalent to

talk about bodily and verbal behavior. If the idea of the criterial relation being "logical" or "grammatical" is understood in this straightforward way, in the context of other minds it produces a behavioristic account of our knowledge of minds. As already said, such an account is not very interesting in itself, and not easily attributable to Wittgenstein; it is actually a position he is continually guarding himself against (e.g. Wittgenstein 1953, §§304-308). The Wittgensteinian idea of behavioral features serving as criteria for mental states is interesting insofar as the "logical" relation of behavioral criteria to mental states is a more subtle one: not entailing them, but being more than mere signs of them.

The outcome of this is an idea of behavioral criteria not as defining criteria, but still as a special kind of "logically adequate" evidence, and furthermore, as (at least sometimes) conclusive evidence:

1) Behavioral criteria are defeasible; but
2) Behavioral criteria provide necessarily good evidence for mental states, and
3) Behavioral criteria are able to settle beyond doubt the question of whether an other is in a mental state.

The third point involves the idea that it is possible to come up with a new kind of solution to the problem of other minds, distinct from an analogical argument, by the help of a Wittgensteinian notion of criteria. The interpretation that criterial evidence is evidence that establishes the existence of something with certainty is what Cavell (1999, 6-7) calls the "Malcolm-Albritton reading" of Wittgenstein's relevant remarks. It is also what Wright (1984, 383-384) calls the "principal point" of criteria in the eyes of most advocates of that notion: recognition of the satisfaction of criteria for P can confer skeptic-proof knowledge that P.

The problem with this is that two mutually exclusive features seem to be expected from criteria: defeasibility and anti-skeptical power. We attribute e.g. pains to others appealing to the typical pain-behavior we observe in them; the pain-behavior serves here as our criterion, as our way of telling that the other is in pain. But the satisfaction of the criterion is not to be taken to be *constitutive* of the fact that he is in pain. That approach would treat the behavioral criterion as a defining criterion, resulting in a behavioristic account. Rather, what we attribute to others appealing to the criterial evidence is a circumstance distinct from the satisfaction of the criteria – an "inner state" or experience. Thus, it is always in principle

possible that the other is not having pain-experience, even though the criteria for his being in pain are satisfied. This is what the defeasibility of criteria is meant to ensure. But this simple admission seems to be in an irresolvable tension with the idea that criteria can settle beyond doubt whether the thing they are criteria of is present or not. Skeptic-proof knowledge about the mentality of another person requires me to be acquainted with such circumstances that are incompatible with the other lacking mentality. But insofar as criteria are defeasible, claiming them to have anti-skeptical power is to say that "knowing that someone else is in some 'inner' state can be constituted by being in a position in which, for all one knows, the person may not be in that 'inner' state. And that seems straightforwardly incoherent (McDowell 1982, 371)".

Probably a direct anti-skeptical solution to the problem of other minds is too much to ask from the notion of criterion. But we can still assess the claim that behavioral criteria constitute a special kind of evidence, an especially adequate way of telling whose connection to persons' experiences is not a contingent matter. What it is that sets up such an assumed connection between certain ways of behaving and "inner" experiences? Wittgenstein has a passage in *Philosophical Investigations* that hints at a possible answer; this answer seems to be based on the insight that our use of language involving pain terminology is itself a sophisticated kind of pain-behavior:

> "How do words refer to sensations? – There doesn't seem to be any problem here; don't we talk about sensations every day, and give them names? But how is the connexion between the name and the thing named set up? The question is the same as: how does a human being learn the meaning of the names of sensations? – of the word 'pain' for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behavior.
>
> 'So you are saying that the word 'pain' really means crying?' – On the contrary: the verbal expression of pain replaces crying and does not describe it." (Wittgenstein 1953, §244)

Here Wittgenstein is giving a possible account of how the connection between a sensation – an "inner" experience – and its name in language is set up. The conjecture – arguably not at all implausible – is that verbal

expressions of pain, and subsequently the language involving pain-sensations in general, get introduced into language by replacing the natural, involuntary expressions of pain. This serves to set up a connection between pain-language and certain ways of behaving that is not a matter of mere contingency: those behavioral patterns that resemble the natural expressions of pain (aversive behavior, crying out, etc.) are the paradigm cases for the correct application of the concept of pain. The special connection between criterial pain-behavior and pain resides in the idea that, as Malcolm (1954, 544) writes, the satisfaction of the criterion "repeats the kind of case in which we were taught to say ['pain']".

This goes at least some way towards explaining how Wittgenstein could see it as a matter of "grammar" that some behavioral features serve as criteria for experiences in others. In a way, the connection is a matter of "convention", in the sense that it is set up in language and upheld in language, but it is not a matter of convention in the sense that we had just decided to call certain ways of trembling or crying "pains". Rather, the link between having painful experiences and displaying pain-behavior is a naturally fixed one, and it is via the publicly observable pain-behavior that we introduce a linguistic expression for the phenomenon of pain.

This way, Wittgenstein can be seen to offer an account of our epistemic relation to others such that directly observable bodily states of others – their pain-behaviors – can provide occasions which we *immediately* see as proper for attributing pain to them. Analogical argument to other minds assumes that pain-behavior is something I note to be constantly *associated* with pain in my own case, allowing the inductive inference from pain-behavior to pain in the case of others. According to the Wittgensteinian account, pain-behavior is not just associated with pain; rather, it is the paradigmatic case for correctly applying the concept of pain. Thus, a person mastering this aspect of pain-language is not required to make an *inference* from observing pain-behavior in another person to the conclusion that the other must be in pain. He will be making the attribution of pain to the other in a single judgment. But the fact that some occasions are the paradigmatic cases for attributing pains to others is not, on any occasion, a guarantee against being mistaken in one's attribution. So the assumption that criteria would be able to settle beyond question the presence of the phenomenon they are criteria of is left controversial; the Wittgensteinian account provides an alternative to the framework of the analogical argument, but it hardly can refute other minds skepticism.

Wittgenstein's point relieves us from one problematic assumption of the analogical argument's framework: if we accept the point, we don't need to construe our judgments about the mentality of others as an inference, starting with a judgment about their bodily behavior and concluding with a judgment about their experiences. The judgment about the mentality of others will be seen as a single judgment, an immediate response to the bodily movements of the other. But it will still be an open question whether any such judgment can be justified without recourse to something like the argument from analogy. I recognize some cases to be paradigmatic occasions for correctly saying that another creature is in pain, but no matter how paradigmatic the case is, for all I know it is still possible that the creature is not experiencing pain; *that* fact still seems to lie outside the limits of my knowledge. And is not the reason I am anyway not too concerned about the possibility of others being automatons simply this: I know that in my own case, this human behavior and this human physiology consistently go together with experiences, so I have reason to be quite reassured that the same is the case with others too? We still seem to have a coherent skeptical question about the existence of other minds that demands an analogical argument as an answer.

It is often pointed out that as soon as skeptical problems are allowed to rise, they seem insoluble. Thus, the most effective anti-skeptical strategy may be thought to be showing that skeptical problems are incoherent or illegitimate to start with. Wittgenstein's own attitude to skepticism, in *On Certainty* (1969) and elsewhere, is of this type: not attempting to meet skepticism in its own terms, but to "silence" the skeptic. This kind of idea seems to be operative in McDowell's "Criteria, Defeasibility and Knowledge" (1982).

The point of McDowell's approach is that skepticism takes the relationship between "good cases" and "deceptive cases" in the wrong way. We will make a distinction between these types of case like this: in a good case, I perceive some state of affairs such that it looks in all respects like P is the case, and it is actually true that P is the case. In a deceptive case, I (again) perceive some state of affairs such that it looks in all respects like P is the case, but P is not actually the case. Now, following McDowell, we should resist the idea that our gaining knowledge of the world through experience is concerned only with "looking-like-somethings" that may or may not be veridical. The root of skepticism lies in being concerned only with the common factor that the good and deceptive cases share – their perceptual indistinguishability – and assuming that the common factor is

all we can appeal to when assessing whether we have knowledge of this or that thing. The antidote for this is, then, to focus instead on the difference between the good and the bad cases: any instance of perception can be either a situation where X perceives P to be the case or a situation where it merely looks to X as if he were perceiving P to be the case. McDowell presents his account as an alternative to a "highest common factor" view about perception, the latter being the view that only the appearances that are shared by good and illusory cases have epistemic relevance.

This is to insist that perception fundamentally involves a kind of openness to the world; opposing the view that experience forms a "veil of representations" between the subject and the world he lives in. (McDowell 1982, 408n19) The latter is seen as the ultimate motivating thought of skepticism.

McDowell's stance is based on the externalist insight that instances of knowing are, so to speak, world-dependent: it is a philosophical mistake, based on the tradition of skeptical arguments in epistemology, to suppose that we should be able to build knowledge just from the materials that are available in our singular subjective experience.

Perceiving a person displaying typical pain-behavior can constitute perceptual knowledge about the person's experience of pain, provided that he indeed *is* experiencing pain – that is, provided the case is a good case and not a deceptive case. The good case calls for a different kind of characterization altogether from that suggested by the skeptic. The skeptic says that in the good case, I perceive something indistinguishable from what I perceive in the deceptive case. The correct, skeptic-silencing, account is however this: in the good case, I *perceive* another person to be in a mental state, while in the deceptive case it merely *seems to me as if I were perceiving* him to be in a mental state.

If one tries to read McDowell's suggestion as an attempt to meet and refute skepticism head-on, it seems hardly successful. The reason, pointed out e.g. by Glendinning & de Gaynesford (1998), is that the skeptic can just reformulate his question as a question about second-order knowledge. Maybe we are indeed able to know that the relevant state of affairs obtains in good cases; but how are we supposed to know *when* we are faced with a good case and not a deceptive case?

> "We are to suppose that this subject's best theory of his or her current perceptual standing (the appearance that such and

> such is the case) is that it is *either* a mere appearance *or* the fact that such and such is the case making itself perceptually manifest. *But no skeptic need deny this*. The skeptic's conclusion is only that, in every case, one must suspend judgment as to *which*." (Glendinning & de Gaynesford 1998, 29; emphases in the original)

But this just confirms the idea that McDowell's argument is better taken as an attempt to undermine skepticism rather than refuting it: it is a suggestion to correct our philosophical intuitions about knowledge so that skepticism can be seen as wrong-headed from the start. The leading idea is the idea of *openness*; we are not radically cut off from the world by a veil of representations, but our perceptions are able to "reach all the way" to the worldly facts themselves, not "falling short of them" in any significant way. It can be asked whether such a position can provide more than a dogmatic, and thus inadequate, response to skepticism (Glendinning & de Gaynesford 1998 argues that it doesn't). I choose here to take the idea of openness at face value, and present some cautious remarks about what the idea might amount to in the case of external world skepticism, on the one hand, and in the case of other minds skepticism, on the other hand.

Following McDowell's clue, we could say that external world skepticism is ultimately motivated by a false philosophical idea like this:

> *We never encounter worldly objects directly, but always as mediated by images or representations of them; and these representations may always be deceptive.*

The same false idea in the case of other minds skepticism would then be:

> *We never encounter the inner, conscious states of others directly, but always as mediated by observations of their behavior; and there may always be behavior without mentality.*

I assume that there is relatively little temptation for anyone with no philosophical aspirations to hold on to anything like the first thesis. In contrary, it seems very natural to think that upon, for example, seeing a tree, it is the tree itself that is affecting the observing subject, not any representational proxy of it. The phenomenology of it, one could say, is that of being in touch with the world directly, characterized by unfaltering

certainty. The distinction between the appearance of the tree and the actual tree only finds a place in contexts where we know the circumstances to be somehow abnormal; when we suspect that there is an optical illusion in play, for example, causing the tree to appear as swaying, while actually it is stationary. In that kind of situation, the tree and the abnormal surroundings can be said to cause an appearance of a swaying tree. But in normal conditions, we don't normally say that we see an appearance of a tree caused by a tree. Rather, we simply see a tree.

What about the second thesis? I think it is worth noting that here there seems to be much more prima facie plausibility in saying that *often*, even in normal, favorable conditions, what we see in the other is a behavioral state caused by a mental state. McDowell seems ready to admit this, at least to an extent. According to his idea of openness, in the good cases, where our perceptions of worldly states of affairs are the result of those states of affairs actually obtaining, what is disclosed to us in experience are those facts themselves, not intervening appearances or representations of them:

> "Suppose someone is presented with an appearance that it is raining. It seems unproblematic that if his experience is in a suitable way the upshot of the fact that it is raining, then the fact itself can make it the case that he knows it is raining." (McDowell 1982, 388)

And when one's experiences of some states of affairs is, as said, "in a suitable way the upshot of" those states of affairs, then "the fact itself is their object" (McDowell 1982, 389). The talk of one's experiences being the *upshots* of states of affairs make it seem like there might be a sense in which the states of affairs themselves remain detached from our experiences of them, and thus remain after all external to one's subjectivity (this point of criticism is pushed in Glendinning & de Gaynesford 1998). Regarding the case of other minds, McDowell particularly makes a crucial disclaimer: when we say that a worldly fact itself, and no intervening substitute, is disclosed to us in perception,

> "[i]n the most straightforward application of the idea, the thought would indeed be […] that the fact itself is directly presented to view […]. But a less straightforward application of the idea is possible also, and seems appropriate in at least some cases of knowledge that someone else is in an 'inner' state, on the basis of experience of what he says and does. Here we might

> think of what is directly available to experience in some such terms as 'his giving expression to his being in that "inner" state'; this is something that, while not itself actually being the 'inner' state of affairs in question, nevertheless does not fall short of it […]." (McDowell 1982, 387)

It is surely true that when some piece of behavior in another person is presented to us as *an expression of a mental state*, then it would be contradictory to hold the behavior to be such an expression *and* also remain skeptical about whether there is a mental state present. In this way, perceiving such an expression would not "fall short of" the inner state of affairs in the sense intended by McDowell. But it seems that some distinction between an expression and what it is an expression of is inevitable, and this introduces a divide which makes it questionable whether we can really enjoy comfortable openness to others. It provides a divide where skeptical doubts find a place to live: it makes it possible to doubt, in any given situation, whether what we take as expressions in another person actually *are* expressions of something.

McDowell talks against an "objectifying conception of the human" (1982, 393), which assumes that insofar as pieces of human behavior are expressive, the expressiveness resides not in the nature of those pieces of behavior, but in their being the observable effects of hidden, "internal" events. But it still seems inevitable that a piece of human behavior's being "expressive" requires its standing in the right kind of relation – maybe causal – to whatever it is expression of; and we seem to be always able to coherently ask whether we can be sure that this relation holds.

Maybe something more or less like this is inspiring Cavell, when he in *The Claim of Reason* (1999) holds that there is a difference in character between external world skepticism and skepticism about other minds. Marie McGinn paraphrases him:

> "[T]here is an asymmetry between scepticism about the external world and scepticism about other minds. While the former is strictly unliveable, confounded by our natural inability to own the sceptic's doubts or feel them as real for us, the latter, [Cavell] suggests, has its roots in our everyday experience of others. […] Sceptical doubt about the external world is 'lunatic'; even when we are caught in the sceptical net of philosophical argument, we never lose sight of our ability to

> put a stop to the fascination by 'joining again … the healthy, everyday world, outside the isolation of the [study]'. In the case of scepticism about other minds […] 'I can live my scepticism', for 'there is no comparable, general alternative to the radical doubt of the existence of others … such doubt does not bear the same relation to the idea of lunacy.'" (McGinn 1998, 45)

Cavell argues against the reading he attributes to Norman Malcolm and Rogers Albritton, which states that observing behavioral criteria for a mental state is sufficient to determine with certainty that another person is in the mental state. Rather, Cavell sees Wittgenstein's demand that we judge others to be in "inner" states based on behavioral criteria as meaning that eventually,

> "'criteria come to an end' […]. There is no final assurance that the other is not a machine; we have the power to grant humanity to the other, and something about the other 'elicits this grant from us', but in the end this granting depends upon an act of 'emphatic projection' whose appropriateness can never be a matter of certainty." (McGinn 1998, 46)

Cavell's insight suggests that it is a part of the human condition to live in some amount of fundamental uncertainty about the extent and nature of the inner lives of other beings. I think there is reason to say that this view was shared by Wittgenstein. And one could find from Wittgenstein also support for the idea that humanity is precisely something we *grant* to the other. Observable behavior, considered just by itself, seems inadequate to determine the "inner states" of others because seeing such behavior as meaningful is essentially a result of our own cognitive input: a result of our capacity to take human behavior as expressive.

"But can't I imagine that the people around me are automata, lack consciousness, even though they behave in the same way as usual? -- If I imagine it now – alone in my room – I see people with fixed looks (as in a trance) going about their business – the idea is perhaps a little uncanny. But just try to keep hold of this idea in the midst of your ordinary intercourse with others, in the street, say! Say to yourself, for example: 'The children over there are mere automata; all their liveliness is mere automatism.' And you will either find these words becoming quite meaningless; or you will produce in yourself some kind of uncanny feeling, or something of the sort.

Seeing a living human being as an automaton is analogous to seeing one figure as a limiting case or variant of another, the cross-pieces of a window as a swastika, for example." (Wittgenstein 1953, §420)

The suggestion here is that seeing others as automata is seeing them under a certain aspect. This aspect may suggest itself whenever I consider people as objects of natural-scientific study: as moving bodies whose future movements I am interested in explaining and predicting. Wittgenstein's calling it a "limiting case or variant" indicates that it appears as a weaker or secondary aspect; it is this because it is pragmatically useless, idle, in the normal course of life. I understand the words by which skeptical doubt about other minds is expressed, and I can even attach some kind of idea to that supposition, but I am unable to do anything with that supposition. This kind of point is echoed by P.F. Strawson in his *Skepticism and Naturalism* (1985), where Strawson suggests that it is a respectable naturalist answer to skeptical worries to point out that in some classes of belief, it is just not up to us to decide whether to hold those beliefs or not. Rather, it is an in-built feature of ours that we respond to human faces in a distinctive way that is not applied to inanimate things. It might be appropriate to say that the default case of our knowledge of other minds is not knowing-that, but knowing *how* to interpret happenings on the surface of the bodies of others as manifestations of mentality.

REFERENCES

Aune, Bruce (1963): "Feelings, Moods and Introspection". *Mind*, vol. 72, no. 286, 187-208.

Avramides, Anita (1996): "Descartes and Other Minds". *Teorema* vol. XVI/1, 27-46.

Avramides, Anita (2001): *Other Minds*.

Cavell, Stanley (1999): *The Claim of Reason*. New York: Oxford University Press.

Chalmers, David (1996): *The Conscious Mind*. Oxford: Oxford University Press.

Glendinning, Simon & de Gaynesford, Max (1998): "John McDowell on Experience: Open to the Skeptic?" *Metaphilosophy*, vol. 29, issue 1-2, 20-34.

Hyslop, Alec (1995): *Other Minds*. Dordrecht: Kluwer Academic Publishers.

Malcolm, Norman (1954): "Wittgenstein's Philosophical Investigations". *The Philosophical Review*, vol. 63, no. 4, 530-559.

Malcolm, Norman (1958): "I. Knowledge of Other Minds". *The Journal of Philosophy,* vol. 55, no. 23, 969-978.

McDowell, John (1982): "Criteria, Defeasibility and Knowledge". In McDowell: *Meaning, Knowledge, and Reality*, Cambridge: Harvard University Press, 1998.

McGinn, Marie (1998): "The Real Problem of Others: Cavell, Merleau-Ponty and Wittgenstein on Scepticism about Other Minds". *European Journal of Philosophy*, 6:1, 45-58.

Plantinga, Alvin (1967): *God and Other Minds*. Ithaca and London: Cornell University Press.

Strawson. P.F. (1985): *Skepticism and Naturalism. Some Varieties*. New York: Columbia University Press.

The Philosophical Writings of Descartes (1985). Translated by Cottingham, Stoothoff & Murdoch. Cambridge: Cambridge University Press. = CSM

Wittgenstein, Ludwig (1958): *The Blue and Brown Books*. Oxford: Blackwell.

Wittgenstein, Ludwig (1953): *Philosophical Investigations.* 3rd edition. Translated by G.E.M. Anscombe. Oxford: Blackwell.

Wittgenstein, Ludwig (1969): *On Certainty*. Ed. by Anscombe & von Wright. Translated by Anscombe & Paul. Oxford: Blackwell.

Wright, Crispin (1984): "Second Thoughts About Criteria". *Synthese* 58, 383-405.

PAUL CHURCHLAND'S CALL FOR A PARADIGM SHIFT IN COGNITIVE SCIENCE

Daniel Ramalho

1. INTRODUCTION

At the very beginning of the interdisciplinary enterprise of cognitive science a schism took place in its midst over the issue of how the nature of cognition should be construed and, consequently, what methodology for studying it should be preferred. The resulting two radically opposing views commonly fall under the headings of "*symbolic paradigm*" (according to which psychological phenomena are rule-governed symbol-based computational processes), and the "*connectionist paradigm*" (which holds that all biological cognition can and should be explained at the subsymbolic level of neuronal activity).

Over the course of the past three and a half decades, Paul Churchland (along with his spouse and colleague, Patricia Churchland)[1] has spearheaded a movement directed at overturning the long-standing dominance of the symbolic paradigm in favor of a reductionist, naturalistically-minded alternative grounded on computational neurobiology and connectionist artificial intelligence. My aim in this paper will be to present a synoptic overview of Churchland's philosophical work and of his main arguments in support of a paradigm shift in cognitive science. In doing so, I mean to join him in claiming that the connectionist paradigm opens far more promising research avenues than its contender and that it should as such become the standard theoretical and methodological model in the scientific study of the mind.

[1] Paul's work is interwoven with that of his wife's to the extent that he has claimed he often feels they 'have become left and right hemispheres of a single brain' (Churchland, 1996a: xii)

In section 2 I will offer a brief account of the origins, main tenets and rise of the symbolic paradigm in cognitive science. In section 3 the historical evolution of connectionist systems will be presented, along with a description of their main functional characteristics (which will be of crucial importance for grasping the technical aspects of Churchland's proposal). In section 4 the main aspects of Churchland's theory will be expounded. Section 5 will then segway into a description of the latter's main advantages over the symbolic paradigm. In section 6, I conclude with some remarks on the future of cognitive science.

2. THE SYMBOLIC PARADIGM IN COGNITIVE SCIENCE

The first formulation of the idea that symbol manipulation is essential to all intelligent cognitive activity – biological and artificial – is commonly identified with Newell's and Simon's *physical system hypothesis*: 'A physical symbol system has the necessary and sufficient means for general intelligent action. By "necessary" we mean that any system that exhibits general intelligence will prove upon analysis to be a physical system. By "sufficient" we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence'" (Newell & Simon 1976: 116). The general conception of mental processes as essentially *computational* processes, however, dates further back. As Newell (1980: 137) himself acknowledged, Whitehead (1927) had pointed in the direction of symbolic cognition half a century earlier. In linguistics, during the late 1950's, Noam Chomsky put forward the greatly influential thesis that the internal structure of all human language is based on a common finite set of recursive rules, and in the following decade Hilary Putnam's proposed the *computational theory of mind*, according to which the mind is to be understood as an algorithmic information processor of discrete internal representations. Similar trends became mainstream across the several disciplines gathered under the umbrella term "cognitive science" long before the latter began entering the scientific jargon in the last quarter of the 20th century.

The catalyst that precipitated this generalized turn to computationalism in the philosophy and sciences of the mind was undoubtedly the theoretical work done by the logician and mathematician Alan Turing, along with that of the mathematician John von Neumann who pioneered the field of computer science by expanding on Turing's legacy. To quote from Newell: '[…] the thread through computer science and artificial intelligence has made a distinctive contribution to discovering the nature of human

symbols. Indeed, in my view, the contribution has been decisive' (1980: 137).

Turing's most enduring contribution to cognitive science was his description in 1936 of what would later become known as the "Turing Machine".[2] A Turing Machine, in broad terms, is an abstract computational device designed for processing information through the sequential manipulation of symbols according to a fixed set of recursive rules. The pairing of this hypothetical computational model with mathematician and logician Alonzo Church's work on recursively calculable functions resulted in the *Church-Turing thesis*, which established that every effectively computable function that can be finitely specified by some recursive procedure is computable by a Turing Machine. This principle implies that a "Universal Turing Machine" can simulate any Turing machine whatsoever or, in other words, that it can be configured so as to deploy a special-purpose set of recursive rules targeted at executing any given computational task. Based on Turing's work, John von Neumann would in 1946 develop the stored-program digital computer design (the "von Neumann architecture").

This invention would become a landmark in the history of cognitive science. The computational prowess displayed by early artificial intelligence digital systems modeled after the von Neumann architecture could leave no doubt as to the latter's theoretical potential, which would eventually cause the theory that human cognition is a matter of rule-based manipulation of symbol-tokens to became widespread within cognitive science. For this reason, says Churchland, 'those many who hoped to account for cognition in broadly computational terms found, in functionalism, a natural philosophical home' (2008e: 18).

Functionalism was originally put forward in 1960 by philosopher of mind Hilary Putnam as a non-reductionist alternative to type-identity theories (i.e. mental states are identical to brain states) and behaviorism (i.e. human and animal behavior can be fully accounted for without resorting to any psychological terminology). According to its orthodox rationale, mental states are abstractly definable in terms of the *functional* role they play in the cognitive system and can, as such, be realized in any physical substrate whatsoever provided it possesses the necessary computational requirements to instantiate them.[3] This is called the argument for the *multiple realizability* of mental states, originally phrased by Putnam as follows: '[…] the functional organization (problem solving, thinking) of

[2] Turing originally used term "automatic machine" (or "a-machine") to designate the UTM.

[3] As Putnam eloquently put it, "we could be made of Swiss cheese and it wouldn't matter" (1975b: 291).

the human being or machine can be described in terms of the sequences of mental or logical states respectively (and the accompanying verbalizations), without reference to the nature of the "physical realization" of these states' (Putnam 1975a: 373).

The symbolic paradigm of cognition that resulted from the coupling of nonreductive philosophical functionalism and symbolic artificial intelligence necessarily entailed favoring a "top-down" methodology for pursuing cognitive science research – one that focused on the *algorithmic* level of cognitive processes rather than the neurological or behavioral. The goal of cognitive science so understood became that of isolating the "software" responsible for human cognition independently of the "hardware" in which it happens to be implemented. This statement's necessary corollary is that scientific inquiry aimed at understanding the brain's microstructure and computational architecture can be deemed useful in a sense but it is ultimately secondary, for the same reason that analyzing the wiring in the innards of a digital computer is not strictly necessary for the task of understanding the operative system it is running.

From the viewpoint of the symbolic paradigm, therefore, cognitive psychology and "classical" artificial intelligence should take the forefront as the leading disciplines in cognitive science and remain methodologically independent from the neurosciences, inasmuch as what is being sought is the "Platonic Function" (Churchland 2008f: 119) running in all human brains, regardless of their implementation-level idiosyncrasies.

3. THE CONNECTIONIST COMPUTATIONAL ARCHITECTURE

Connectionism consists of a fundamentally *subsymbolic* model of information processing. Unlike classical computation systems based on the serial manipulation of discrete symbols ("concepts") according to a fixed set of recursive rules ("syntax"), connectionist systems process information in an entirely nonlinguistic, nonserial and nonlocalized fashion.

The connectionist model was originally presented in a famous article by McCulloch and Pitts (1943) describing a computational architecture inspired in the physical structure of biological brains. Briefly put, they proposed a system in which information processing was entirely carried out at the subsymbolic level of artificial neurons divided into a variable number of interconnected layers.[4] Having established the principles underlying the functioning of artificial neural networks in the language of

[4] Hence the term "connectionism" or, alternatively, "artificial neural networks".

propositional logic, McCulloch and Pitts formally demonstrated that these systems could in principle perform any computational task that could be executed in a finite number of steps through the basic logical operations of conjunction, disjunction and negation. In doing so, they proved that connectionist systems possess, in abstract, the computational power of a Universal Turing Machine.

Based on this theoretical framework, the computer scientist Frank Rosenblatt (1958) developed the *perceptron* neural network design (shown in figure 1), which would become the standard model for future connectionist systems.

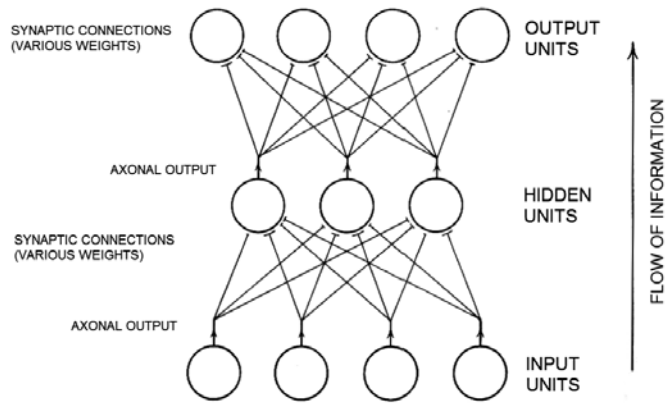

Fig. 1 - The perceptron's computational architecture (Churchland 1992: 162)

The perceptron architecture is composed or three layers of artificial neurons: an input (or sensory) layer, a middle (or hidden) layer, and an output (or response) layer. Communication between neurons in this model is strictly unidirectional ("feedforward"), as neuronal signals can only travel from the input layer to the output layer. Each neuron projects an output connection ("axon"), which then ramifies into several others, linking it to all the neurons of the next layer. When active, a neuron will send a signal that will be conveyed through the connections to all neurons in the following layer.

Neurons in the input layer become active depending on the stimulus they receive from the external environment. The value of the signal each

input neuron sends to the subsequent layer is modified depending on the synaptic *weight* of the connection through which it travels (excitatory or inhibitory). Each neuron in the next layer then calculates the values of all the incoming signals from the previous layer and, depending on the result, a signal will be fired to the next layer, or not. The same process will then repeat itself all the way up to the output layer, which will dictate the network's response to the initial stimulus.

McCulloch's and Pitts' model already embodied the same computational structure proposed by Rosenblatt. However, Rosenblatt introduced two important innovations: i) the weights of the neuronal connections in his model were continuous rather than binary (i.e. their value was not simply "excitatory" or "inhibitory" but could vary in degree within each polarity); and ii) he introduced mathematical procedures for adjusting the value of each individual synaptic weight, thus allowing for the improvement of the network's overall performance. This innovation introduced the possibility of *training* neural networks to execute complex tasks such as pattern recognition (by presenting the sensory layer with successive stimuli of a similar kind and adjusting the middle layers' synaptic weights at each turn until the network predictably produced the same output given the same type of input).

Although such early connectionist models displayed considerable power in executing a wide variety of cognitive tasks, in practice they remained computationally deficient compared to their digital counterparts. This was so for more than one reason, but mainly because there was no autonomous procedure available for adjusting weights in artificial neural networks. This entailed that connectionist systems could only learn with the help of a human "tutor". As a result of this limitation, research in artificial neural networks went through a long period of stagnation and was almost universally considered a research dead-end among cognitive scientists. It was not until the mid-1980's that interest in connectionism reemerged, at which time a new computational algorithm was developed that made it possible for artificial neural networks to operate efficiently without human supervision: the "generalized delta rule", commonly known as "the backpropagation algorithm".

The backpropagation algorithm allows for a computationally very powerful *recurrent* neural network design (as opposed to it's strictly *feedforward* ancestors). In such networks, the output layer can send signals *back* to the middle layers. By using the backpropagation algorithm to calculate the discrepancy between the network's actual output and the desired output, the network can autonomously recalibrate the *synaptic*

weights[5] in the middle layers until after a sufficient number of training trials they reach the optimal configuration of weights for executing the task at hand. In other words, the backpropagation algorithm endowed artificial neural networks with a versatile method for *learning* without human intervention and, in doing so, liberated these systems from many of the computational shortcomings that had previously humbled them in comparison to digital computers. The backpropagation algorithm was of paramount importance in the development of the computational architecture that characterizes most contemporary connectionist systems and which is at the heart of Churchland's theory of cognition: that of *parallel distributed processing* (PDP).[6]

Although Churchland does not argue that brains deploy the backpropagation algorithm in their computational activity,[7] he proposes that an account of biological cognitive activity based on the PDP computational model far surpasses the symbolic alternative in a wide variety of aspects. In the following section, the two core aspects of Churchland's proposal will be presented: i) the *Domain-Portrayal Semantics* model, which is Churchland's theory of how the brain acquires and stores information (i.e. how it embodies *knowledge*); ii) and the *Dynamic-Profile Approach*, which is his proposed answer to the question of how information processing in the brain becomes *conscious*.

4. PAUL CHURCHLAND'S CONNECTIONIST MODEL OF COGNITION

Domain-Portrayal Semantics

The basic PDP computational architecture is illustrated in figure 2. Its structure and information processing dynamics will be described in the following paragraphs in order to contextualize Churchland's connectionist model of cognition.

[5] Each neuronal connection has a synaptic weight, which determines the measure of the effect that a signal travelling through a given connection will have on the receiving neuron (i.e. the amount of influence it will have in causing that neuron to fire).

[6] This designation is owed to the fact that information storage in such systems is not localized in a memory database, but distributed along the synaptic weights of the connections linking parallel layers of neurons.

[7] Churchland (1992c: 246-250) believes the most likely candidate for explaining how synaptic weight adjustment takes place in brains is *Hebbian learning* (a process by which a synaptic connection between two neurons increases its weight as a function of the amount of times the presynaptic neuron contributes to the firing of the postsynaptic one). He acknowledges that 'the problem of what mechanisms actually produce synaptic change during learning is an unsolved problem', but adds that 'the functional success of the generalized delta rule assures us that the problem is solvable in principle, and other more plausible procedures are currently under active exploration' (Churchland 1992b: 187).

As was the case with the basic perceptron architecture, each individual neuron in PDP networks is connected to every other in the subsequent layer. However, as mentioned in the previous section, in contrast with perceptrons, the *activation levels* of individual neurons in PDP networks are not binary but graded along a numerical scale (in the example of the network depicted in figure 1, this scale includes ten degrees ranging from 0 to 1). A neuron's activation level will dictate the strength of the signal it will emit (which will be transformed in accordance with the weight values of the connections through which the signal will travel).

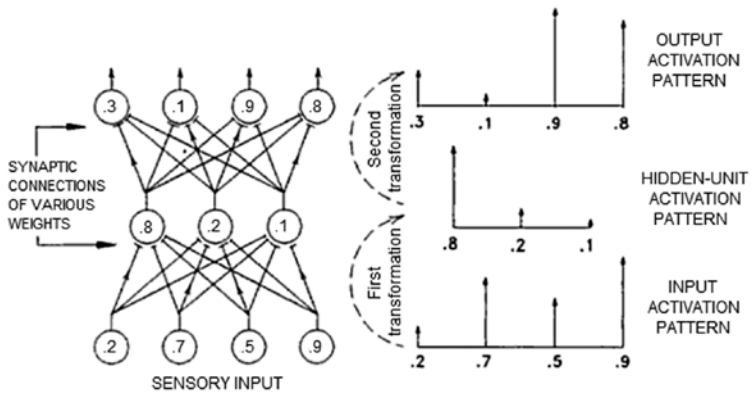

Fig. 2 - Schematic depiction of the parallel distributed processing computational architecture and respective activation vectors (Churchland & Churchland 1998a: 58)

The set of activation levels of all neurons in the input layer is that layer's *representation* of the input stimulus. Inasmuch as that representation is defined by an ordered set of numbers, it can be mathematically described as a *vector*. Each pattern produced by the simultaneous activation of neurons in a given layer is therefore that layer's *activation vector* (i.e. the activation vector of the input layer in fig.2 is [.2, .7, .5, .9]).

Once calculated, each layer's activation vector is transmitted to the following layer throughout the several weighted connections. That layer will then likewise acquire an activation vector which will in turn become *its* representation of the stimulus received. The same process will be repeated until the final activation vector is calculated by the output layer, dictating the networks final response. Thus, Churchland defines PDP networks as 'multistage device[s] for successively transforming an

initial sensory activation vector into a sequence of subsequent activation vectors embodied in a sequence of downstream neuronal populations' (Churchland 2008b: 98).

As depicted in figure 3, an activation vector locates a *coordinate* within a multidimensional *activation space* – or, to use Churchland's most common designation, *state space*. A state space is an abstract graph depicting the entire range of a neuronal layer's possible activation vectors. Each dimension (or axis) of a state space represents a single neuron of its corresponding layer. Considered as a coordinate, an activation vector can be represented as a *point* within the state space (see figure 3).[8] Therefore, an *activation point* denotes a neural network's current representational state.

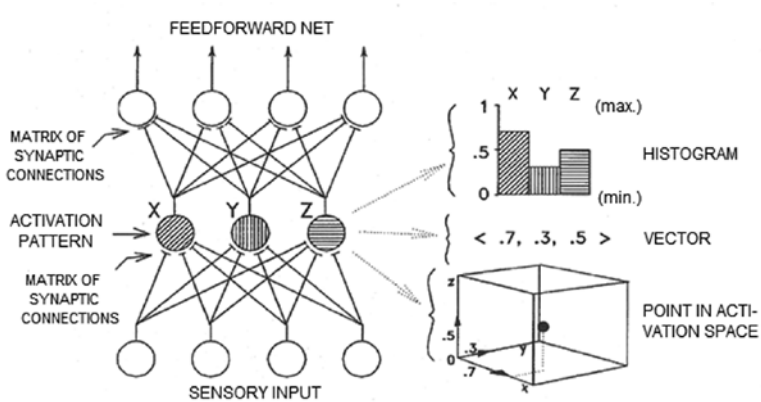

Fig. 3 - Coding in an elementary network (Churchland 2008b: 97)

Neural networks can spontaneously learn to partition their abstract internal spaces into subvolumes by gradually adjusting the hidden layers' synaptic weights during the training period, and thus create primed regions for categorizing certain types of stimuli. These regions are the network's *concepts*. Input that produces activation vectors located within the same subvolume will be equally categorized by the network (i.e. it will result in the same behavioral output). At the center of each subvolume is the *prototype-point* for that category which, to paraphrase Churchland, is

[8] The state space depicted in figure 3 corresponds to a neuronal layer of only three neurons and can therefore be three-dimensionally represented as a cube. However, state spaces can be mathematically described as *n*-dimensional objects.

something like the "Platonic form" for the class of stimuli it represents (for example, in figure 4, each prototype point corresponds to the activation vector of an "ideal" color). The measure of reliability with which a network will categorize a certain stimulus will be determined by the proximity of the activation vector it elicits to a given prototype point: 'the nontypical or marginal cases of the concept reside toward the periphery of that volume, and its center-of-gravity point represents a prototypical instance of that concept' (Churchland 2008a: 145).

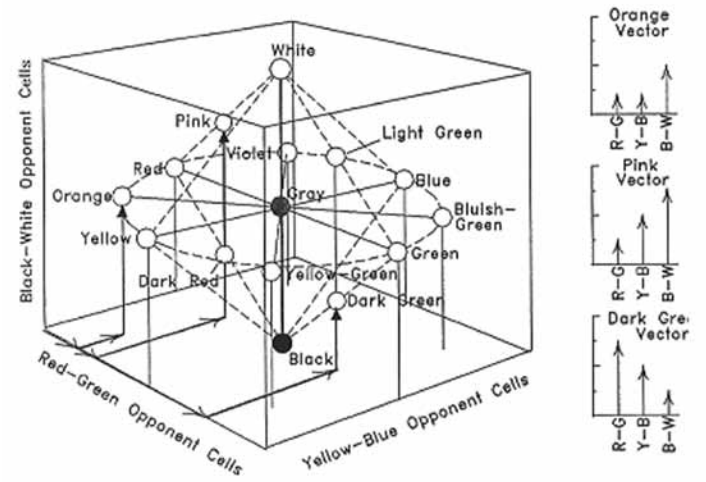

Fig. 4 The state space for human color coding vectors (Churchland 1998a: 23)

Inasmuch as this description of the learning process of PDP networks applies to both artificial and biological networks, Churchland formulates the following definition of *concept*: 'having-a-concept is having-the-*capacity* to represent each of a variety of relevantly related particular cases as lying within the same narrowly confined subvolume of an activation space, a subvolume that bears a relatively fixed set of distance relations to a great many others such preferred subvolumes. Crudely speaking, a concept is not an image, but an isolated and graded range of *possible* images. And to *have* a concept is to *command* that well informed range of possible representations' (ibid.). Concepts are therefore construed in Churchland's Domain-Portrayal Semantics theory as *subsymbolically defined regions within multidimensional abstract state spaces, physically embodied in a given*

neuronal population's web of weighted synaptic connections.

Conceptual frameworks in this perspective obtain their intentional content much in the same way as *maps* do (Churchland 2008a: 153-160). The same target domain of properties can be mapped with equal precision in many different representational media (e.g. photographs, blueprints, drawings, etc.). A typical road map, for example, is an accurate representation of a road system if the graphic elements determining its internal configuration portray the distance relations of the road's relevant geographical features in a correct scale (the relevant features being, in this case, those that are necessary for driving successfully along the road). Likewise, a neural network can be said to possess *knowledge* of a given domain of properties if the internal configuration of the family of the prototype points within its state space is *homomorphic* with the abstract structure of the salient elements of that target domain.

Brains can thus be regarded as natural-born "cartographers", continuously engaged in drawing and updating an enormous variety of "cognitive maps" representing all the dimensions in which they navigate, pinpointing their immediate position in each of them and adjusting their course along the way by constantly plotting new hypothetically favorable routes. This analogy is adequate insofar as it illustrates both the *internalist* and *holistic* aspects of Churchland's semantic theory. Respectively, neural maps (like any regular map) represent domains of properties without being *causally* linked to them, and derive their representational content from the *collective* configuration of their internal elements rather than the intrinsic semantic value of each prototype point.[9] However, it is only partially adequate given that typical maps are confined to representing *geographical* properties. The representational capacity of neural networks, on the other hand, extends to an indefinitely vast array of possible target domains. To list a few modest examples, the intricate web of a normal human brain's weighted synaptic connections typically embodies state spaces for discriminating sensory stimuli (e.g. colors, tastes, smells, etc.), performing complex motor tasks (e.g. swimming, driving, writing, etc.) and even successfully navigating the *social* world by identifying the appropriate position of morally relevant actions within a *moral* state space (e.g. stealing, nurturing, cheating, assisting, etc.).

There is therefore no difference from the neurological point of view between PDP networks trained for discriminating colors, faces, animals, bodily movements or English words, apart from their respective input-

[9] Taken by itself, a prototype point is as meaningless as a "you are here" sign in an otherwise empty map.

output systems. Conceptual frameworks of virtually any kind can be equally embodied in the medium of interconnected neurons.

The Dynamical-Profile Approach

Its explanatory power and broad reach notwithstanding, Churchland's proposal would remain incomplete if it failed to provide an account of phenomenal consciousness. Being a naturalist and a reductionist, Churchland argues for the complete identity of the neural and the phenomenological levels. In his view, qualia are not something over and above neural activity. Rather, 'qualia and […] vectors are not distinct things at all: they are identical; they are one and the same thing, although known to us by two different names. (Churchland 2008d: 195). Their co-occurrence, says Churchland, is no more mysterious than that of the substance *snow* and the substance *neige* (ibid.).

To expand on the reductionist character of Churchland's theory with an example, consider the experience of watching a clear-sky sunrise. When the first light beams begin peering from the horizon, the observer's visual cortex reacts by translating the incoming low-wavelength luminous input into an activation vector that will index a position near the prototype point for the color "red" within its abstract chromatic state space (see figure 4). Simultaneously, an activation vector in the subject's *emotional* state space will code for the appropriate feeling of wonder such a moment merits. As the Sun continues its apparent upward motion, the activation point in the awestruck observer's chromatic state space gradually moves away from the "red" prototype point, which phenomenologically translates into a growing difficulty in categorizing the light's color with certainty. The rising Sun continues to become increasingly "reddish" until the point's trajectory along the chromatic state space leads it outside the boundaries of the color red's subvolume. The activation point will then enter the outskirts of the adjacent subvolume, causing the light's color to again become categorized with less and less ambiguity, this time as progressively similar to *orange*. This process will continue until the activation point (mimicking the Sun's trajectory along the sky) reaches its apex within the state space, coding for "bright white".[10]

As the preceding example illustrates, Churchland does not mean to provide an account of how qualia *correlate* with brain activity. Inasmuch as the theory he proposes offers a method for *quantifying* all conscious

[10] This may strike one as a radically reductive view of human conscious experience in more than just the epistemological sense. Sensing the possibility of such an interpretation of his theory, Churchland writes: 'This should occasion neither horror nor despair. For while we now know these phenomenological roses by new and more illuminating names, they present as sweetly as ever. Perhaps even more sweetly, for we now appreciate why they behave as they do' (Churchland 2008d: 194).

experience (i.e. as fleeting patterns of activation-levels across neural populations, mathematically described as activation vectors in abstract state spaces), "qualia" simply cease to merit that designation and should therefore be discarded as useless philosophical artifacts.[11]

Although Churchland does not presume to have solved the mystery of consciousness, he believes that enough evidence from connectionist AI and neuroscience is currently available to warrant the claim that a fully workable theory of consciousness can be conceived strictly on the basis of the computational properties of *recurrent* neural networks (Churchland 1996a: 215-226; 2008c).[12] As previously explained, recurrent networks project descending axonal pathways leading back from the output layers to the middle layers. This feedback-loop, Churchland claims, endows these systems with a form of *short-term memory*: 'Some of the information present in the second layer's activation vector two or three cycles ago may still be implicit in the stimulation vector currently arriving there via the recurrent pathways. Such information decays over a number of cycles rather than disappearing after only one' (Churchland 1996a: 216). In other words, recurrent neural networks can possess *current* knowledge of their *past* activity.

The temporal dimension that this computational property of recurrent networks unfolds is the thread with which Churchland knits together an explanation of both conscious and unconscious brain processes into a single unifying blanket theory encompassing all cognitive activity. Churchland's Dynamical-Profile Approach to consciousness is strictly coherent with the previously outlined Domain-Portrayal Semantics theory in that it construes consciousness as being wholly independent from its *subject matter*. Consciousness is not about *what* is being processed in the brain, it's about *how*. To paraphrase Churchland (2008c: 12), it is the cognitive *activities* in which representations are involved and the *computational context* in which those activities take place that will determine if a given cognitive process will become conscious – not their content.

By taking this stance, Churchland sets himself apart from all theories of consciousness as *self*-consciousness (e.g. Dennett, 1991; Damásio, 2010) on the grounds that the brain's basic computational architecture is the same

[11] It could be argued that this is too strong a metaphysical claim given that this proposal is compatible with several varieties of non-reductive physicalism. However, to such objections Churchland need only reply that having provided a plausible *scientific* explanation of phenomenal consciousness, the "burden of proof" is no longer his to bear. Until an empirically testable non-reductive alternative with superior explanatory and predictive potential is provided, no strong ontological theory concerning the nature of qualia can claim the right to be the default philosophical position.

[12] There is no doubt that biological brains are recurrent neural networks. In Churchland's words, 'It is a rare neural population that sports no descending projections at all' (Churchland 1996a: 99).

for every one of its representational subsystems, regardless of their specific target domains being internal or external to the organism. Consequently, according to this position, consciousness is far from being an exclusively human trait: '[…] the contrast between human and animal consciousness has to go […], for nonhuman animals *share* with us the recurrent neuronal architecture at issue. Accordingly, conscious cognition has presumably been around on this planet for at least fifty million years […]' (Churchland 2008c: 17).

5. CHURCHLAND'S CASE FOR A PARADIGM SHIFT IN COGNITIVE SCIENCE

In this section I will present an overview of Churchland's main arguments supporting the claim that a paradigm shift should take place in cognitive science research. These arguments will be divided along three broad categories: *biological plausibility*, *eliminative materialism* and *reductionism*.

Biological Plausibility

The main reasons for crediting Churchland's connectionist model with greater biological plausibility than the symbol-based account of cognition stem from the evident discrepancies between digital computers and brains in terms of their information processing *performances* and *architectures*.

The first such discrepancy pertains to the conspicuous mismatch between the computational *speed* of biological and digital information processors. Despite the fact that electronic signals in digital computers travel in average a million times faster than neuronal signals in brains, the latter outperforms the former in sheer computational speed by several orders of magnitude. The reason for this apparent paradox is the *von Neumann bottleneck* problem, inherent to the von Neumann architecture. Inasmuch as the digital computer's CPU accesses the system's database *sequentially*, a certain amount of lag is inevitable given that regardless of how fast the information processing goes it will always be constrained by the fact that all information must pass in single-file through the narrow channel ("bottleneck") of the CPU's limited bandwidth. Connectionist systems are not affected by this problem on account of their entirely different computational architecture, which is unconstrained by a limited

throughput capacity[13] as digital computers are: 'Specifically, the biological brain is a massively *parallel* piece of computational machinery: it performs trillions of individual computational transformations – within the 10^{14} individual microscopic *synaptic connections* distributed throughout its volume – *simultaneously* and *all at once*' (Churchland 2008e: 21).

The second reason for distrusting the digital computer analogy concerns the matter of computational *accuracy*. If brains were indeed to process information sequentially, the performance of every recursive step would contain an inevitable measure of error due to the fact that individual neurons are functionally much more unreliable than electronic components in digital systems. In the course of thousands of successive steps, such individually negligible errors would collectively accrue to computational disaster. In connectionist systems, however, such errors are easily overcome, given that each computational task is assigned to thousands or millions of neurons *at the same time*, which causes the combined errors of even hundreds of them to become negligible in the context of the entire network's performance. 'In sum', says Churchland in regard to both the previous arguments, 'if the brain were indeed a general-purpose digital serial computer, it would be doomed to be both a computational tortoise and a computational dunce' (2008f: 120).

It could be objected that these limitations of digital computers cannot be construed as *a priori* arguments against the biological plausibility of the symbolic cognition thesis inasmuch as they merely reflect the shortcomings of present-day computational technology, which could in principle be overcome in the future. However, even if one is to accept this reply, a third more fundamental disanalogy between brains and digital computers remains. The claim that all human brains are running the same basic "software" entails the assumption that the latter is somehow embodied in the weight configuration of the brain's synaptic connections, like a program installed in a general-purpose digital computer.[14] The fatal flaw condemning this hypothesis to failure from the outset is that brains are simply not equipped to install generic "software". Whatever function a brain is running can only be embodied in its "hardware" inasmuch as its overall internal configuration *is* its memory database. Due to the highly idiosyncratic organization of each brain's neuronal layers and synaptic weights, it is impossible for two brains to stumble upon the configuration

[13] The rate at which information travels through a communication channel.

[14] Dennett, for instance, has argued at length that human brains are genetically adapted to post-natally "install" a "virtual machine" – a pattern of rules imposed upon the brain's microstructure which '[…] vastly enhance[s] the underlying powers of the organic hardware on which it runs […]' (Dennett 1993: 210). According to this perspective, the brain is both a neural network *and* a von Neumann-like machine.

that will embody the exact same set of "rules". In Churchland's words: '[…] no two of us normal humans are computing exactly the same abstract function. Its existence, as that which unifies us, is a myth' (2008a: 125). The necessary condition for the mutual understanding of two normal human beings to occur is the existence of *isometrically sculpted families of prototype points in their respective abstract state spaces* (i.e. their having different "cognitive maps" depicting the same objective reality), not a shared "cognitive software". For this reason, if ever there was a "Platonic function" running in a human brain, it died along with Plato.

Eliminativist Materialism

The second cluster of objections Churchland raises towards the symbolic paradigm pertains to the fundamental notion instilled in the classical cognitive science program that our commonsense psychological framework ("folk psychology") is a crude but generally accurate description of cognitive activity. According to this view, the framework of psychological concepts we typically deploy in our daily lives (e.g. propositional attitudes such as "believing, "remembering", "desiring", etc.) constitutes a reliable blueprint of the aforementioned "cognitive software", and therefore an explanation of intelligent cognition should take the form of a fine-grained version of that blueprint.

Churchland has adamantly criticized this view since the very beginning of his career, adhering to the philosophical position known as *eliminative materialism*. Its core tenet, as he defines it, is that '*our common-sense psychological framework is a false and radically misleading conception of the causes of human behavior and the nature of cognitive activity*.' (Churchland 1988: 43). In light of the fact that the familiar mentalistic terminology of folk psychology is incommensurable with that of neuroscience (at which level the most accurate description of cognitive activity is to be attained), the former should be *eliminated* and give way to the latter.

Four main objections are raised by Churchland towards the conception of folk psychology as a faithful mirror of human cognition. Firstly, Churchland points out that folk psychology is a millennia-old *theory* and should be looked upon with the same measure of suspicion as any other theory of such remote origins: 'The FP of the Greeks is essentially the FP we use today, and we are negligibly better at explaining human behavior in its terms than was Sophocles' (Churchland 1992a: 8). Eliminative materialism holds that such a long period of 'stagnation and infertility' makes it implausible that folk psychology is any more accurate than folk physics, alchemy or the Ptolemaic theory of the motion of

celestial bodies, wherefore it should join them in the pantheon of archaic pseudoscientific mythologies.

In addition to its poor explanatory *depth*, Churchland argues that folk psychology also suffers from an exceedingly narrow explanatory *breadth*, as it is utterly incapable of shedding light on the nature of such a wide array of cognitive phenomena as sleep, perceptual illusion, mental illness, learning, imagination and complex motor behavior, among many others.

A third objection is that the bulk of human cognitive activity is executed independently of propositional attitudes. Churchland offers a familiar example in support of this claim: 'I may awaken from a long and fierce reevaluation of Hume's argument for rejecting miracles, only to realize that I can remember nothing of the last twenty miles of perfectly successful highway navigation. If any propositional attitude was fixed during that period, their topic was Hume's skeptical philosophy, not the details of road and traffic' (Churchland 2008b: 91). Additionally, even in the extreme case of global aphasia[15] affected individuals remain capable of executing such complex tasks as cooking, driving, shopping and playing chess. An account of human cognition as *fundamentally* linguaformal is hard-pressed to account for such cases as these.

Churchland's final objection towards folk psychology is twofold, and it pertains to nonhuman animal cognition. Firstly, it is undeniable that despite being unable to command propositional attitudes, a great number of creatures remain nonetheless capable of complex behavior which must therefore be accounted for without recourse to symbolic cognitive processing. Secondly, the assumption that human cognition is distinct from that of any other animal because of its essentially symbolic/sentential nature implies postulating that our linguistic capabilities emerged as a result of an unprecedented evolutionary *qualitative leap*. The introduction of such an amendment to the history of evolution to account for the cognitive activity of a single species should not be warranted if a plausible, more parsimonious theory is available.

Churchland has shown that such an alternative exists. According to his proposal, linguistic concepts, unique as they undoubtedly are to our species, remain nevertheless only a *kind* of concept among a great variety of others. There is therefore no reason to believe that they are fundamental to human cognition in any sense, particularly in light of the fact that they are, evolutionarily speaking, the very *latest* to have appeared. This does not mean that Churchland intends to downplay the genuine originality

[15] A neuropathology characterized by the complete loss of the capacity for expressing and understanding language.

of human language in the history of evolution. His intention is rather to describe that originality in terms of a complexification of preexisting brain structures and cognitive capabilities, eschewing the need for postulating the emergence of entirely new ones. Language, says Churchland, 'is an acquired *skill*, both a motor and a perceptual skill. But do not think of it as the skill of producing and recognizing strings of words. Think of it instead as the acquired skill of perceiving (opaquely, to be sure) and manipulating (again, opaquely) the brain activities of your conspecifics, and of being perceptually competent, in turn, to be the subject of reciprocal brain manipulation' (Churchland 2008a: 159).

In construing language as an acquired *skill* and providing a neurocomputational explanation of it as such, Churchland discards any notion of exclusively human "software" as Fodor[16] and Dennett would have it, as well as the innate Chomskyan language module.[17] The basic human cognitive architecture as described by Churchland is evolutionarily very old and philogenetically widespread.

Reductionism

One of the criticisms Churchland most often raises against the classical functionalist picture of cognition is that in positing the irreducibility of mental phenomena to their underlying brain activity, it fails to provide an adequate account of how the psychological level "fits in" with the lower-level sciences. This failing, as Patricia Churchland remarks, constitutes an unacceptable philosophical anachronism: 'in a curious way, brain-averse functionalism is methodologically close to Cartesianism. In place of Descartes' *nonphysical mental substance*, functionalism substituted "*software*"' (Churchland 2002: 27).[18] The resulting marginalization of psychology from the remaining natural sciences can no more be warranted than suggesting the (also somewhat Cartesian) aforementioned "cognitive gap" segregating human minds from those of all other animal species on the basis of the purported uniquely symbolic nature of our mental lives. The potential for intertheoretic reduction of any scientific account of cognition must be a basic criterion for judging its validity. In that respect, Churchland's proposal is far more promising than the symbolic approach.

It is common to conceive of Paul Churchland as a staunch

[16] Jerry Fodor (1975) famously proposed the radical theory that human mental processes take place in the medium of an internal *language of thought*, and that *all* lexical concepts as well as syntactical rules are innate.

[17] The claim that a connectionist account of human language is possible is empirically supported by the fact that artificial neural networks (specifically, "Elman networks") have been successfully trained to determine the grammaticality of complex English sentences (Churchland 1996a: 137-143, Elman 1992).

[18] In its original context, Patricia Churchland's objection was directed at Jerry Fodor.

eliminativist towards all things psychological, both folk and scientific. There is, however, unambiguous evidence in his writings that this is a misconstrual of his philosophical program: '[…] it should not be assumed that the science of psychology will somehow disappear in the process [of reducing mental phenomena to neuroscience], nor that its role will be limited to that of a passive target of neural explanation. […] At this level of complexity, theoretical reduction does not appear as the sudden takeover of one discipline by another; it more closely resembles a long and slowly maturing marriage.' (Churchland 1998b: 79). Also, in a short but sharp text, the Churchlands write: 'Our iconoclastic reputations aside, we count ourselves among the most fervent of the Friends of Psychology' (1996b: 219). An accurate description of Churchland's reductionism, therefore, must emphasize that his criticism of scientific psychology is directed at its status of preferred level for explaining cognitive phenomena and not its status as a legitimate field of research.

There is another more overarching sense in which Churchland's proposal is reductionist. In his 1944 book "*What is Life?*", the physicist Erwin Schrödinger suggested that the evolution and behavior of complex organic systems can ultimately be accounted for in terms of the basic principles of thermodynamics. Living beings thus construed are *energy dissipaters* – open thermodynamic systems in a constant struggle to avoid reaching energetic equilibrium by exploiting low-entropy energy sources in their environment (e.g. the Sun) and using that energy to increment their physical structure (e.g. growing, reproducing, healing, etc.). Metaphorically speaking, from Schrödinger's viewpoint, life is what happens when the second law of thermodynamics begins cannibalizing itself so as to avoid death by starvation.

Churchland (1982; 2008e) claims that this reduction of biological phenomena to nonequilibrium thermodynamics can be extended to cognitive phenomena. "Cognitive metabolisms" feed off low-entropy *information* sources in much the same way that biological metabolisms feed off low-entropy *energy* sources. That information is thereafter embodied in the configuration of the brain's synaptic weights, after which it is dissipated back into the environment as high-entropy energy (e.g. heat).

To put it in a different way, cognitive beings are 'extrasomatic information multipliers' (Churchland 2008e: 30) – tireless epistemic foragers continuously perusing their "epistemic niches" for information (in the form of light, sound, smell, etc) so as to accrue to their limited genetically inherited supply and so enhance their chances for survival. Hence, Churchland's proposal provides not only a route for reducing cognitive activity to the neurobiological level but also for reducing it to the

fundamental level of physics, thus setting psychology squarely in line with the lower-level natural sciences all the way to the bottom.

6. CONCLUSION

We have seen that Paul Churchland offers a radically different perspective on cognitive activity than that of classical functionalism; an account that looks to real neurons rather than abstract symbols for answers. Computational neurobiology and artificial intelligence have come a long way in the past few decades and are now ready to offer us an unprecedented understanding of the hidden mechanisms of biological brains at the most fine-grained level. Time has come for cognitive science to mature as well and eschew its anthropocentric linguaformal prejudice.

Interestingly, Churchland suggests that this maturation should take the form of a return to *infancy*, in two distinct but interrelated senses. In order to advance, cognitive science should trace its steps back to its early years and heed the advice of Alan Turing, who in spite of being commonly regarded as 'the consensus patron saint of the classical research program in AI' (Churchland 2008f: 113) would more aptly be described as the 'unsung patron saint of the more recent and biologically inspired program of research into *artificial neural networks* (ibid.: 122). This is so because although Turing's work was undoubtedly crucial for the development of digital information processing systems and, therefore, to the rise of the cognitive paradigm, he never subscribed to the view that artificial intelligence should look for the "cognitive software" running in every human adult brain. On the contrary, he claimed it should strive to understand and possibly recreate the *child*'s brain: 'Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. Presumably the child brain is something like a notebook as one buys it from the stationer's. Rather little mechanism, and lots of blank sheets.' (Turing 1950: 456).

In light of this argument it is clear that Churchland's reductionist proposal is fully compatible with functionalism's multiple realizability argument. He agrees that mental states can be instantiated in a variety of different physical substrates, but only so long as they emulate the computational properties of *neural networks*, not *digital computers*. In short, any artificial intelligence system must like any cognitive creature *learn* how to think by painstakingly embodying knowledge from its external environment in its internal representational system. Simply attempting to upload the program for human cognition in a hard-drive just

will not do. The characteristic richness of our mental lives is not the glitter of our wealth of symbols but rather the humdrum of billions of neurons tirelessly trading on the universal currency of all biological cognitive economics: the *vectorial* attitude (Churchland 2008b: 98). Until cognitive science acknowledges this fact, our minds will remain shrouded in mystery to our own eyes.

REFERENCES

Churchland, P. M. (1982), "*Is* Thinker *a Natural Kind?*", Dialogue 21, 2: 223-238

Churchland, P. M. (1988), *Matter and Consciousness* (revised edition, first published in 1984), Cambridge, Massachusetts, MIT Press

Churchland, P. M. (1992), *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, Massachusetts: MIT Press

Churchland, P. M. [1981] (1992a), "*Eliminative Materialism and the Propositional Attitudes*", in Churchland, P. M. (1992): 1-22

Churchland, P. M. [1989] (1992b), "*On the Nature of Theories: A Neurocomputational Perspective*", in Churchland, P. M. (1992): 153-196

Churchland, P. M. (1992c), "*Learning and Conceptual Change*", in Churchland P. M. (1992): 231-253

Churchland, P. M. (1996a), *The Engine of Reason, the Seat of the Soul*, Cambridge, Massachusetts, MIT Press

Churchland, P. M. (2008), *Neurophilosophy at Work*, New York, Cambridge University Press

Churchland, P. M. [2001] (2008a), "*Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds*", in Churchland (2008): 126-160

Churchland, P. M. [2001] (2008b), "*What Happens to Reliabilism When It Is Liberated from the Propositional Attitudes?*" in Churchland (2008): 88-112

Churchland, P. M. [2002] (2008c), "*Catching Consciousness in a Recurrent Net*", in Churchland (2008): 1-17

Churchland, P. M. [2005] (2008d), "*Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience*", in Churchland (2008): 161-197

Churchland, P. M. [2005] (2008e), "*Functionalism at Forty: A Critical Retrospective*" in Churchland (2008): 18-36

Churchland, P. M. [2006] (2008f), "*On the Nature of Intelligence:*

Turing, Church, von Neumann, and the Brain" in Churchland (2008): 113-125

Churchland, P. M. & Churchland, P. S., (1996b), "*The Future of Psychology, Folk and Scientific*", in *The Churchlands and Their Critics*, (ed.) McCauley, R. N., Cambridge, Massachusetts, Blackwell Publishing: 219-221

Churchland, P. M. & Churchland, P. S. (1998), *On the Contrary: Critical Essays, 1987-1997*, Cambridge, Massachusetts, MIT Press

Churchland, P. M. & Churchland, P. S. [1990] (1998a), "*Could a Machine Think?*", in Churchland & Churchland (1998): 47-64

Churchland, P. M. & Churchland, P. S. [1990] (1998b), "*Intertheoretic Reduction: A Neuroscientist's Field Guide*", in Churchland & Churchland (1998): 65-79

Churchland, P. M. [1998] (1998c), "*Conceptual Similarity across Sensory and Neural Diversity: The Fodor-Lepore Challenge Answered*", in Churchland & Churchland (1998): 81-112

Churchland, P. S. (2002), *Brain-Wise*, Cambridge, Massachusetts, MIT Press

Damásio, A. (2010), *Self Comes to Mind: Constructing the Conscious Brain*, London, William Heinemann

Dennett, D. (1993), *Consciousness Explained*, Penguin Books, (first published in 1991)

Elman, J. L. (1992), "*Distributed representations, simple recurrent networks, and grammatical structure*", Machine Learning 7: 195-225

Fodor, J. (1975), *The Language of Thought*, New York, Crowell

Laakso, A., Cottrell, G. (2000), "*Content and cluster analysis: Assessing representational similarity in neural systems*", in *Philosophical psychology*, 13: 47-76

McCulloch, W.S., Pitts, W. (1943), "*A logical calculus of the ideas immanent in nervous activity*", Bulletin of Mathematical Biophysics, 5: 115-133

Newell, A. (1980), "*Physical Symbol Systems*", Cognitive Science 4: 135-183

Newell, A., Simon, H. A. (1976), "*Computer Science as empirical enquiry: Symbols and Search*", Communications of the ACM, 19: 113-126

Putnam, H. (1975), *Philosophical Papers – Mind, Language and Reality*, Cambridge University Press, 2

Putnam, H. [1960] (1975a), "*Minds and Machines*", in Putnam (1975): 362-385

Putnam, H. [1975] (1975b), "*Philosophy and Our Mental Life*", in Putnam (1975): 291-303

Rosenblatt, F. (1958), "*The Perceptron: A probabilistic model for information storage and organization in the brain*", Psychological Review, 65: 386-408

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986), "*Learning representations by back-propagating errors*", Nature, 323: 533-536

Turing, A. (1950), "Computing Machinery and Intelligence", Mind, New Series, 59, No. 236.: 433-460

Whitehead, A. N. (1927), *Symbolism: its meaning and effect*, New York, Macmillan

THE IDEA OF MORAL PERSONHOOD UNDER FIRE

Oscar Horta

1. INTRODUCTION

The concept of moral personhood plays a central role in a number of moral positions. It is used to distinguish those entities that have certain capacities that are morally relevant from any other entities there may be in the world. Some views claim that only persons are morally considerable, others claim that they deserve some special consideration other entities are not worthy of.

This is not the only way in which the term 'person' is used. In fact, we can find it in several fields apart from the moral one. It is commonly used in metaphysics to name the kind of beings that humans usually are. And it is also used in the legal realm to name those entities that have the capacity to sue. Finally, in common language its usage is widespread to name those who belong to the human species.[1]

It is usually assumed that persons in the metaphysical, the legal, the moral and the common sense coincide, that is, that they are the same entities — this being the reason why they are all called persons. The picture that results from this is one in which humans are entities of a certain kind, of which all other entities (including all other conscious beings) are excluded, and that only they must enjoy moral and legal significant protection. That is, all humans, and only humans, are believed to have certain capacities that turn them into metaphysical, legal and moral persons.

To be sure, there is a long tradition that stresses the difference

[1] In fact, there are other fields in which the term 'person' is also used, such as psychology or grammar. I will not consider them here since their relevance to the issue we are dealing with is not so central, but this shows how widespread the use of this term is.

between the concept of a person and that of a human being. This tradition tracks back to Locke, and is very much in use currently in the discussions concerning abortion.[2] Despite this, and setting aside the fact that even those who draw this distinction have a certain idea of what human beings usually are when they describe persons, we must note that this view is not widely accepted anyway. In fact, the view that all humans are moral persons is widespread and also has a long tradition behind it, with Kant being its most significant defender. Even if for Kant any rational being could be a person, he assumed all humans were persons, excluded all other animals from this category and as a matter of fact, often used the term 'human' as a synonym of 'person'. This reproduces a common practice among those who use the concept of personhood: although they present some strict criterion to qualify as a person, they also end up assuming all humans, and only humans, will satisfy it.[3]

Moreover, the importance that the language of persons has in metaphysics, ethics and the law has much to do precisely with the fact that it is humans that are believed to be persons. The assumption that humans are central in all these realms determines that the concept of personhood is also enormously relevant in them.

In this paper I argue that this perspective, prevalent as it is today, must be rejected. In it I defend two different claims. First, that personhood, as currently understood in ethics as in other fields, cannot be considered to be an attribute coextensive with humanity, and second, that, in fact, the language of moral personhood should be abandoned altogether.

To defend these claims, in section 2 I start by presenting the meaning that the term 'person' has in ethics, as well as in other fields such as metaphysics, law and common language. I also consider the claim that there are persons *simpliciter*. Then, in section 3 and 4 I examine the conditions that must be met to be a person in metaphysics and law, and whether all and only humans satisfy them. Next, in section 5 I consider this problem in the moral arena, for each of the different meanings the term 'moral person'

[2] See Locke, John, *An Essay concerning Human Understanding*, Dent, London, 1968 [1690]. See also, for instance, Tooley, Michael, "Abortion and Infanticide", *Philosophy and Public Affairs*, 2, 1972, 37-65; "Personhood", in Kuhse, Helga & Singer, Peter (eds.), *A Companion to Bioethics*, 2nd ed., Wiley-Blackwell, Chichester, 2009, 129-139; Warren, Mary Anne, "On the Moral and Legal Status of Abortion", *The Monist*, 57, 1, 1973, 43-61; or Engelhardt Jr, H. Tristram., *The Foundations of Bioethics*, Oxford University Press, Oxford, 1986, in particular p. 104. In fact, there has been a reaction against this precisely by those who oppose abortion. They have argued that it is the fact of being human that matters, and have worried about those humans who do not qualify as persons (see for instance Weiss, Roslyn, "The Perils of Personhood", *Ethics*, 89, 1978, 66-75).

[3] Kant's "humanity formula" of the categorical imperative actually expresses this quite clearly. See Kant, Immanuel, *Groundwork of the Metaphysic of Morals* , Harper and Row, New York, 1964 [1785], 4:429. See also, for instance, ibid., 6:442; *Critique of Practical Reason*, Cambridge University Press, Cambridge, 1997 [1788], 5:76.

may have. After this, in section 6 I conclude that the extension that the term 'person' has in common language and the moral, legal and the metaphysical realms differs significantly. In no other field apart from common language is personhood and membership to the human species coextensive. This means that the common assumption regarding the identification of the different dimensions of the term 'person' must be rejected.

After pointing this out, in section 7 I claim that the use of the term 'person' in all these different fields and with all these different meanings is not warranted and is due to an anthropocentric bias. Because such view is not justified, this gives us reasons to doubt whether the use of the concept of personhood is really adequate. Finally, in section 8 I argue against the view that the idea of moral personhood could be of some use regardless of all this. I contend that if the concept of moral personhood can be used to modify what an analysis of the features that can be morally relevant entail, then it cannot be a justified concept. Otherwise, if it is seen as a mere synonym of "moral consideration", it ends up being a superfluous and confusing one. I point out that similar conclusions can be reached for other meanings of moral personhood.

2. WHAT IS A PERSON?

In order to discuss personhood we need to start by distinguishing the different meanings we may find in each area. As we will see now, they need not be closely connected, even though they sometimes are.

2.1. Moral persons

There are several ways in which the term 'moral person' may be understood. It has been used to name, for instance:

(i) Individuals who are morally considerable.[4]

(ii) Individuals who have certain special interests that are particularly important in moral terms by means of the kind of beings they are, in particular the interest in living.[5]

(iii) Individuals who are moral agents.[6]

[4] See for instance Sztybel, David, "Animals as Persons", in Castricano, Jodey (ed.), *Animal Subjects: An Ethical Reader in a Posthuman World*, Wilfrid Laurier University Press, Waterloo, 2008, 241-257.

[5] See Singer, Peter, *Practical Ethics*, 3rd ed., Cambridge University Press, Cambridge, 2011 [1979]; Tooley, "Abortion and Infanticide"; Harris, John, "The Concept of the Person and the Value of Life", *Kennedy Institute of Ethics Journal*, 9, 1999, 293-308; McMahan, Jeff, *The Ethics of Killing: Problems at the Margins of Life*, Oxford University Press, Oxford, 2002.

[6] See for instance Frankfurt, Harry, "Freedom of the Will and the Concept of a Person", *Journal of Philosophy*, 68, 1971, 5-20; Dennett, Daniel, "Conditions of Personhood", in his *Brainstorms: Philosophical Essays on Mind and Psychology*, MIT Press, Cambridge, 1981, 267-285; Scott, G. E., *Moral Personhood: An Essay in the Philosophy of*

2.2. Metaphysical persons

Metaphysical persons can be defined as entities that have some kind of existence, but that are not necessarily also the entities that qualify as moral persons. There are two views regarding what the nature of metaphysical persons can be:

(i) According to an anti-reductionist approach, they are instances of a certain sortal that cannot be defined in terms of a different sortal.[7]

(ii) According to a reductionist approach, they are entities that have the features to qualify as members of a certain type, defined by the possession of certain mental capacities. For instance, some have argued that persons are those conscious entities whose unity is granted by common memories.[8] Others have argued that persons are continuous or connected contents of consciousness.[9] Others have claimed that persons are embodied minds.[10] Others, differently, have claimed that persons are agents.[11] Others have claimed that persons are intentional beings with certain complex cognitive capacities.[12] Others have argued that they are conscious beings.[13] We can easily see, then, that there is no consensus here.

2.3. Legal persons

Legal persons are entities that can start legal actions, that is, that can sue. According to naturalism, the reason why there are, or there must be, legal persons is that there are natural persons, which may be seen either as human beings, as persons *simpliciter*, as metaphysical persons, as moral persons or as both metaphysical and moral ones. These natural persons are, or must be, legal ones too. If instead we assume a positivist viewpoint, we may claim that natural persons do not exist, that it is irrelevant whether they do exist, or that there are natural persons, although the existence of legal persons is not determined by their existence.

However, there are also many legal persons who do not appear to satisfy any account of what natural persons are. This happens, for instance, in the case of companies, states, corporations, etc. In fact, the term 'legal person' is commonly used to name only those persons recognized by the

Moral Psychology, State University of New York Press, Albany, 1990; Farson, Timothy E., *Contemplating Moral Personhood*, Master Thesis, San Diego State University, San Diego, 2010.

7 See Strawson, Peter, *Individuals: An Essay in Descriptive Metaphysics*, Routledge & Kegan Paul, London, 1959.

8 Locke, op. cit.

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

10 McMahan, op. cit.

11 Korsgaard, Christine M., "Personal Identity and the Unity of Agency: A Kantian Response to Parfit", *Philosophy and Public Affairs*, 18, 1989, 103-131.

12 Dennett, op. cit.

13 Sztybel, op. cit.

law that are not assumed to be "natural persons". Some theories claim that these "non-natural" legal persons are actually real entities, and thus persons with an ontological status similar to that of "natural" ones. Others claim that they are mere aggregates of the real persons (natural ones). Yet others claim they are mere fictions.[14]

2.4. Persons in Common Language

Despite all we have seen before regarding different accounts of personhood in several realms, the fact is the term 'person' is mostly used to mean something different. In common language, a person is, simply, a human being.

2.5. Persons *simpliciter*

Finally, it is sometimes claimed that persons are persons *simpliciter*. These can be defined as entities of a certain kind that have such features that qualify them as persons in all the senses the term has (moral, metaphysical, legal, common language and any other). This definition is circular, but at least it allows us to see that those who believe in the existence of persons *simpliciter* assume that (i) whatever persons exist, in any sense, they must be persons in all senses; and (ii) that persons do exist. Some anti-reductionistic accounts of metaphysical personhood appear to assume that metaphysical persons are persons *simpliciter*.

2.6. The Common Assumption

The question here arises, of course, as to how these different conceptions of what a person is can be related. According to a common assumption, humans are persons in every sense of the term (from now on I will call this "the common assumption"). That is, humans are believed to be metaphysical, moral and legal persons. In some accounts, they are also persons *simpliciter*.

Now, despite the general acceptance of the common assumption, it is a wrong view. In fact, there are several reasons why this is so. In order to examine them we will now see what entities can be persons in each field.

3. WHAT ENTITIES ARE METAPHYSICAL PERSONS?

We have seen that there are many different accounts of what

[14] See on this Phillips, Michael J., "Corporate Moral Personhood and Three Conceptions of the Corporation", *Business Ethics Quarterly*, 2, 435-459.

metaphysical persons are. If we assume an anti-reductionist view, this question remains mysterious, since we are left with no criteria to determine who is or who is not a person. If we assume a reductionist account, things change, since that means we will be able to verify whether different entities can qualify or not as persons. The problem we face here, however, is that since there are so many views on what persons are, there will be beings who satisfy some criteria for metaphysical personhood but not others. (This may be the case, for instance, of a sentient being with a very simple mind who does not have complex cognitive capacities.)

Now, in light of this, we can ask, are all and only humans persons according to all these criteria? Are they persons only according to some of them? Or rather is it the case that there is no criterion of metaphysical personhood that all and only humans satisfy?

Despite the common assumption,[15] the only question we can answer affirmatively among these ones is the last one. There are many humans who satisfy all the mentioned criteria for personhood. But there are others who fail to satisfy some of them. For instance, there are humans with very serious brain injuries or with certain congenital conditions who have minds, but lack complex intellectual capacities. And there are also others who do not have minds, and who thus fail to meet any condition for metaphysical personhood.

Moreover, there are nonhuman animals who satisfy criteria for metaphysical personhood that humans, as we have seen, fail to satisfy. Some of them have complex intellectual abilities, and many of them are conscious beings.

4. WHAT ENTITIES ARE LEGAL PERSONS?

What about legal personhood? As we saw above, not only humans are legal persons. Of course, not all entities protected by the law are legal persons: works of art may be protected without being persons. But there are legal persons who are not humans, such as companies, states or corporations. They are not metaphysical persons and, on most accounts at least, they cannot be considered moral persons either.[16] Some humans

[15] Steve Sapontzis also expresses this clearly when he writes, "[n]o matter how superior its behavior, a dog can never be a person$_d$ because it does not have a human body, and no matter how inferior the behavior of a human infant or a handicapped human, he is still a person$_d$ because he has a human body [Sapontzis uses the term person$_d$ to mean a metaphysical person]". Sapontzis, Steve F., "**A Critique of Personhood**", *Ethics*, 91, 1981, 607-618, p. 608. This is a widespread view among philosophers. See for instance Wiggins, *Identity and Spatio-Temporal Continuity*, Blackwell, Oxford, 1967, in particular p. 48.

[16] French has argued that corporations satisfy the criteria for being considered a moral person, on the basis

represent these legal persons and take legal actions on their behalf, but they are not the legal person they represent; rather, the company, state or corporation is.

Now, those who accept that these legal persons are real entities will have to accept that their existence shows that not only humans can be legal persons. Those who reject this but assume a legal positivist position will also accept that there can be nonhuman legal persons. Only those defenders of natural law who also claim that these legal persons are not really natural persons will have to reject this conclusion.

In addition, we need to note also that some humans who lack certain capacities and cannot act on their own behalf in any way are legal persons as well.[17] They are also represented by others. However, nonhuman animals are not considered to be persons (defenders of the moral consideration of nonhuman animals have thus understandably argued against this.)[18] This shows that the criterion by which legal personhood is granted is neither the possession of certain capacities nor metaphysical personhood. Rather, it is simply membership to the human species that matters here. It is the common language meaning of personhood that is taken as relevant. So we may think that at least when it comes to "natural" persons legal persons are humans. But then, this can only be accepted if we reject any grounding of the notion of legal person on metaphysical personhood or, in fact, on any non-definitional anthropocentric viewpoint.

that they are intentional systems we could say to have their own interests, and make their own decisions. See French, Peter A., "The Corporation as a Moral Person", *American Philosophical Quarterly*, 16, 1979, 207-215. In support of this view see also Weaver, William G., "Corporations as Intentional Systems", *Journal of Business Ethics*, 17 87-97. But this requires a conception of what is a moral person that many of us would find wholly implausible (see for instance Manning, Rita C., "Corporate Responsibility and Corporate Personhood", *Journal of Business Ethics*, 3, 1984, 77-84). We may claim that if no human being had any interest at all in the continuation of the existence of a corporation, we would have no reason at all to care for it, which we should have if the corporation were morally considerable in itself. If we consider that moral personhood should be attributed to them because they have agency we assume a view of agency that appears as very problematic. Of course, we may claim that the corporation can make decisions that none of its members individually would take (since its decisions may be the sum of their own decisions or the result of a deliberation among them). Yet if that is so, that would also be the case of any group of individuals who decide to act together. In addition, on many accounts this view cannot be right simply because agency is assumed to be a feature of individuals who act as a result of having certain mental states, and corporations do not have mental states.

[17] See Berg, Jessica "Of Elephants and Embryos: A Proposed Framework for Legal Personhood", *Hastings Law Journal*, 59, 2007, 369-406, in particular p. 377.

[18] Dunayer, Joan, *Speciesism*, Ryce, Derwood, 2004; Francione, Gary L., *Animals as Persons: Essays on the Abolition of Animal Exploitation*, Columbia University Press, New York, 2008. Some theorists have claimed (often simply for it being something easier to achieve) for the legal personhood of only some animals with certain capacities. See Cavalieri, Paola & Singer, Peter (eds.), *The 'Great Ape' Project: Equality beyond Humanity*, Forth Estate Limited, London, 1993; Wise, Steven M., *Rattling the Cage: Toward Legal Rights for Animals*, Profile, London, 2000; *Unlocking the Cage: Science and the Case for Animal Rights*, Perseus, Oxford, 2002.

5. WHAT ENTITIES ARE MORAL PERSONS?

We can now turn to the meaning of 'personhood' with which this paper is more concerned, moral personhood. We have seen that this term has been used with several different meanings. We will examine now which entities can be moral persons according to each of them.

5.1. Morally considerable entities

Let us start with the idea that claims that moral persons are morally considerable entities (or entities whose interests matter more than those of others). The fact is that there are many different views regarding what makes an entity morally considerable, and examining them all would require much space. However, we can follow a different approach here. Since we are primarily interested in knowing whether the different meanings of personhood are coextensive and if the common assumption is right, we can approach this problem by examining the different defenses that there are of the view that only humans are morally considerable, or at least if they are so in ways in which other entities are not. Critics of this view have claimed that it is an unjustified position, and that it must be rejected as an instance of speciesism, the discrimination of those who do not belong to a certain species.[19] So we need to carry out a careful analysis of the different arguments that can defend anthropocentrism.

There have been many different defenses of this idea. We can distinguish two main groups in which they can be classified. First, some of them are based on criteria whose possession can be verified and that appeal to further reasons beyond the actual desired conclusion (that is, that humans take priority over other animals). Second, others are based on criteria that do not meet these conditions.

Let us examine first those criteria of the latter kind. There are two types of positions that can be included here. First, there are those defenses of anthropocentrism that are purely definitional. That is, the ones that claim that it is the mere fact of being human that renders them morally considerable.[20] Second, there are those defenses of anthropocentrism that are based on religious reasons or on those that appeal to metaphysical criteria not related to verifiable criteria. Instances of this are the claim that

[19] See on this Horta, Oscar, "What Is Speciesism?", *Journal of Agricultural and Environmental Ethics*, 23, 2010, 243-266.

[20] Posner, Richard A., "Animal Rights: Legal, Philosophical and Pragmatical Perspectives", in Sunstein, Cass & Nussbaum, Martha (eds.), *Animal Rights, Current Debates and New Directions*, Oxford University Press, Oxford, 2004, 51-77.

humans are the chosen species by God,[21] or that only humans possess a high ontological status (without this being related to the possession of any actual physical features).[22]

The reason why these criteria fail to justify anthropocentrism is, simply, that they do not provide any verifiable justification of this view, but simply formulate it in a different way. However, there are other criteria that avoid this problem. Let us examine them now.

We can distinguish two different defenses of anthropocentrism here: the ones that appeal to intrinsic features which, it is supposed, only humans possess, and the ones that appeal to special relations allegedly maintained by humans.

The intrinsic features that are typically considered here are capacities. And the capacities that are usually considered relevant are those that are related to mental states, basically, the possession of complex intellectual capacities.[23] As for relations, defenders of anthropocentrism sometimes argue that humans have exclusive bonds of solidarity or affection, or power relations that determine that they should respect each other but not other animals.[24]

Now, there are reasons to claim that none of these criteria provide us with a sound justification of anthropocentrism. In particular, two reasons can be mentioned. First, we may claim that they are not based on what is morally relevant. Second, they do not draw a line between humans and nonhuman animals.

To start with the first reason, it may be pointed out that relations and cognitive capacities sometimes have to do with what is valuable for individuals. Not in themselves, however, but instrumentally or indirectly. For instance, if I have a close relationship with someone else, I may suffer or enjoy a great deal depending on how the life of that individual goes, which I will not if I do not know her at all. Also, having certain cognitive abilities may cause me to suffer or enjoy if I anticipate some future negative or positive experiences. Alternatively, it may spare me a great amount of suffering or joy if those capacities allow me to know that my present misery

[21] See for instance Reichmann, James B., *Evolution, Animal 'Rights' and the Environment*, The Catholic University of America Press, Washington, 2000.

[22] See for instance Machan, Tibor, *Putting Humans First: Why We Are Nature's Favorite*, Rowman and Littlefield, Oxford, 2004.

[23] See for instance Frey, Raymond G., *Interests and Rights: The Case against Animals*, Oxford University Press, Oxford, 1980; Leahy, Michael, *Against Liberation: Putting in Animals in Perspective*, Routledge, London, 1991; Carruthers, Peter, *The Animal Issue: Moral Theory in Practice*, Cambridge University Press, Cambridge, 1992.

[24] See for instance Becker, Lawrence C., "The Priority of Human Interests", in Miller, Harlan B. & Williams, William H. (eds.), *Ethics and Animals*, Humana Press, Clifton, 1983, 225-242; Midgley, Mary, *Animals and Why They Matter*, University Georgia Press, Athens, 1983; Goldman, Michael, "A Transcendental Defense of Speciesism" *Journal of Value Inquiry*, 35, 59-69.

or bliss will end immediately. However, it is quite clear that such relations or capacities cannot be what determine that I feel any kind of suffering or enjoyment in the first place. Rather, what can cause this is the fact that I have the physical wiring that allows me to do so. In other words, that I have the capacity to have positive and negative experiences. Accordingly, if we want to accept a criterion for moral consideration that is relevant for the very purpose that criterion will have, which is to determine which beings it may be right to affect positively and negatively, it seems that sentiency will be a fine one, since it is the one that determines that a being can be affected positively and negatively. Of course, we are not logically forced to accept this. We may perfectly accept that moral consideration is about who we should or should not affect in certain ways, yet reject that moral consideration be granted on the basis of whether one can be affected in those ways. But the idea that these two criteria should be connected seems a very sensible reason which many of us will surely find very compelling.

As for the second reason, we have seen already throughout this paper that there are many human beings who do not have complex cognitive capacities. In fact, given any non-definitional capacity or relation, there will be humans who lack it. There is just no non-definitional criterion that can grant moral consideration to all humans.

Consider, first, the mentioned capacities. There are humans (such as infants, some who have suffered serious brain injuries or some who have certain congenital conditions) that simply do not have them. Think now of what happens in the case of relations. It is simply not true that humans have universal relations of solidarity and affection among them. There are human beings who do not care for others, and there are human beings who do not have anyone who cares for them (we may think, for instance, of orphans who live in the streets, or people who live in conditions of slavery with no relatives). Also, if we consider power relations we can easily conclude something similar. Humans are often in a situation of power over other animals, that is clear. But then, many humans are also the victims of other humans, they are under their power without being able to do anything about that.

What follows from here is that such criteria not only exclude nonhuman animals from the realm of moral consideration, they also exclude a number of human beings. In fact, there is no non-definitional criterion that no nonhuman animals satisfy which may grant consideration to all humans.

Moreover, this entails that if these criteria are what define moral personhood then some humans cannot be moral persons. No matter how we define moral personhood, provided that we do so in a non-definitional

way, there will be humans who would fail to satisfy the requirements we establish.

So, given all this, we have to conclude that moral anthropocentrism is not justified.

This means that the idea that only humans are moral persons cannot be justified if by moral personhood we mean moral considerability, or special moral considerability. Due to this, several theorists have claimed that sentient nonhuman animals should be accepted as moral persons as well.[25] It may also be argued,[26] however, as I will do here, that this may be so if we accept the use of the concept of moral personhood, but that we may also oppose its use.

5.2. Entities with special interests

We have seen that according to some views moral personhood is not equal to moral considerability. Rather, persons are certain beings who have some features that determine that they have certain special interests that other entities lack, but that are not the ones that must determine the attribution of moral consideration as such. Most importantly this would allegedly mean they would have an interest in living. For instance, Peter Singer has claimed that all sentient beings must be morally considered, but that those self-conscious beings who have capacities such as being able to see themselves through time, to make long term plans and to engage in meaningful relationships with others are persons, and for this reason have special interests that other sentient beings lack.[27] Michael Tooley has also defended a similar concept of personhood which links its possession to an interest in not being killed.[28]

So, the question here is whether all and only humans are persons in this second sense. Given what we have seen thus far, we know already this cannot be so, since there are many humans who do not have the aforementioned capacities. Moreover, Singer himself, and other theorists too, claim that there are non-humans who are persons in this sense.[29] So

[25] See Aaltola, Elisa, "Personhood and Animals", *Environmental Ethics*, 30, 2008, 175-193; Sztybel, op. cit.

[26] See *Faria*, *Cátia*, *"Pessoas* não humanas: a consideração moral dos grandes símios e outras criaturas", *Diacrítica*, 25, 2011, 33-50.

[27] Singer, op. cit.

[28] See Tooley, "Abortion and infanticide"; Harris, op. cit.; McMahan, op. cit. Different theorists defend different arguments for personhood in this sense. It has been argued due to this that there is just no shared conception of what it is to be a person in this sense. See English, Jane, "Abortion and the Concept of a Person", *Canadian Journal of Philosophy*, 5, 1975, 233-243. See also Beauchamp, Tom L., "The Failure of the Theories of Personhood", *Kennedy Institute of Ethics*, 9, 1999, 309-324.

[29] See on this Cavalieri, Paola & Singer, Peter, op. cit.; Midgley, Mary, "Is a Dolphin a Person?", in her *Utopias, Dolphins and Computers: Problems of Philosophical Plumbing*, Routledge, London, 1996, 107-117; DeGrazia, David, "Great Apes, Dolphins, and the Concept of Personhood", *Southern Journal of Philosophy* 35, 1997, 301-

we have to reject the idea that moral personhood in this second sense can be identified with membership to the human species.

5.3. Entities with moral agency

We have seen that the term 'moral person' has also been used to name those individuals who can be considered moral agents. (Those who claim this typically argue that moral persons are those who are moral agents *and* are morally considerable.)[30]

Again, what we have seen before shows that not all humans can be moral persons in this sense either. Those humans who lack certain cognitive capacities simply cannot have responsibilities towards others. All this setting aside the fact that there are nonhuman animals who may be considered moral agents. So, again, moral personhood cannot be identified with humanity in this case either.

6. The common assumption is wrong

In light of what we have seen throughout the analysis carried in the previous sections, we have to conclude that the extension of the term 'person' in the different realms mentioned above does not coincide. The next reasons can be presented on the basis of what we have seen above:

(i) For any account of metaphysical personhood based on some verifiable criterion there are human beings who are not metaphysical persons.
(ii) It is plausible to claim that there are metaphysical persons who are not human beings.
(iii) There are legal persons that are not human beings, such as corporations and states.
(iv) For any account of moral personhood based on verifiable criteria there are humans who are not moral persons.
(v) It is plausible to claim that there are moral persons who are not human.
(vi) Due to (i) and (iv), there are human beings who cannot be considered to be persons *simpliciter*.

[30] This may seem somehow trivial, since in the world in which we are living those who are agents are also morally considerable; note, however, that it is open to question whether there could be, say, moral agents who, although conscious and able to think of how to act towards others, and to act accordingly, were unable to have positive and negative experiences, and would not mind what happened to them. On some accounts, they would not be morally considerable, since they would not need it, although they would still be moral agents.

This means that the common assumption regarding the identification of the different dimensions of the term 'person' must be rejected. Only in common language are 'personhood' and 'humanity' coextensive. But this provides us with no real justification to accept the common assumption. If anything, it is a reason to doubt it, since what may well happen is that language is deceiving, and we believe that persons are humans in the moral, the metaphysical and other fields because we usually call humans persons.

7. THE LINK BETWEEN SPECIESISM AND THE APPEAL TO MORAL PERSONHOOD

We have seen that the criteria for being a person in different realms are quite different, and can be satisfied by different individuals. Accordingly, the concept of personhood *simpliciter* is not a credible one. In addition to this, we have seen that there are strong reasons to conclude that (many) nonhuman animals are morally considerable, and that the appeal to moral personhood cannot undermine this claim. This idea need not be really connected to the previous one. Despite this, when we examine these problems closely we can discover that there is a relation between the failure of the concept of moral personhood and its anthropocentric root. Given what we have seen regarding the problems that the use of the concept of moral personhood has, we may wonder why it has been used so widely in moral philosophy. We may think that a reason for this is that it is comfortable and easy for us to think of the world as divided into different categories, and that the idea that there are persons and non-persons allows us to do this more easily than the idea that there are simply individuals with interests to respect. This claim appears to be rather reasonable. However, there is more to say regarding this. The concept of moral personhood seems to be used in moral thinking in particular (if not only) because of the common assumption.

Regarding this, it is worth noting something here. We have seen that anthropocentrism has been defended with many different criteria. It is interesting that, diverse as they are, they all end up drawing exactly the same domain. This strongly suggests that those who appeal to these different reasons do not evaluate impartially which criteria may be morally relevant and then draw the conclusions that they may imply regarding who we should take into account. Rather than that, we have strong reasons to believe that they do things just the other way round: they start with the idea

that moral anthropocentrism is right, and then try to look for convincing reasons to support that claim. This is not an impartial way to do ethics, rather, it is a biased one.

It is interesting to note that something very similar happens in the case of metaphysical personhood. Many different theories have been presented as to what kind of thing is a person. Some of them are anti-reductionist, others are reductionist; among the latter, some claim persons are flows of consciousness, others claim persons are agents, others claim persons are embodied minds, etc.

All these accounts of what persons are appear to be remarkably different. However, they commonly use the term 'person' to name the kind of entity they want to denote. Moreover, the philosophical topic which consists in what is the kind of thing we are is usually referred to as the problem of personal identity. This is so even if there are theorists who claim we are not persons but other things. Furthermore, a main discussion in the debates on this issue takes place between those views that claim we are basically organisms or bodies and those that claim we are basically psychological entities. Interestingly, the former family of views is sometimes referred to as 'animalism', while the latter is called 'personalism'. This seems to imply that humans may be similar to other animals if they are not essentially persons, and that nonhuman animals cannot be persons, even though, since many nonhuman animals are conscious beings, they may perfectly be considered metaphysical persons on several accounts. This fact is forgotten due to an anthropocentric bias. What is more, it is also interesting to note that the problem of what kind of things we are is almost always thought of as if the 'we' in the question denotes human beings. Even when the answer to that question is one according to which there are nonhuman beings who belong to the same kind of beings of which we are part, the question is almost always presented as having do with what humans are.

Furthermore, we also saw that legal personhood is granted, to individuals at least, not due to any metaphysical consideration or to the possession of certain capacities, but rather on the mere basis of membership to the human species. Again, in this field like in the other ones, it is assumed from the beginning that humans are persons and vice versa.

All this reinforces the idea that when these issues are considered there are strong anthropocentric biases conditioning the way we approach them. The common language meaning of 'person' conditions strongly the way we consider personhood in ethics as in metaphysics. In fact, the very use of the term 'person' in these fields is strongly motivated by its widespread use in common language. We not only think persons are humans, we also

think that the concept of personhood is a very relevant one because we assume persons are humans. But this is unjustified, since, as we have seen, membership to the human species is morally irrelevant.

8. MORAL PERSONHOOD: SUPERFLUOUS OR UNJUSTIFIED

Given that (a) the use of the concept of "moral personhood" in ethics is due to an anthropocentric bias, and that (b) anthropocentrism is an unjustified speciesist position (as we saw in section 5.1), we have reasons to doubt whether that concept is really a sound one. But there is more to say regarding this. There are further reasons to conclude that we should better get rid of the concept of moral personhood.

But then, if this is so, this concept adds nothing to moral decision making. We are left with a theory that claims that a certain criterion is morally relevant, and due to it is the one that grants moral consideration, and that those who are so regarded are persons. If this is so, the fact of being a moral person adds absolutely nothing to the way in which we should act towards someone, since even if we did not call that individual a moral person we would have to act towards her in exactly the same way (because of the moral consideration to which she is entitled anyway).

Therefore, moral personhood could be a meaningful concept only if qualifying as a person were a morally relevant feature not reducible to other criteria. That is, if it made a difference in moral decision making as to how to morally consider the interests of interests holders. But for this to be so, moral personhood would have to be ascribed on the basis of a criterion which would be different from the one that would grant moral consideration. But if this is so then something wrong is going on here. Moral consideration should be granted on the basis of a morally relevant criterion (or criteria). So if there is something else that is morally relevant that grants moral personhood, then that means that the criterion for moral consideration we initially had did not include all that it is morally relevant, and thus was not a sound one. Or, alternatively, if the criterion for moral consideration did include everything that is morally relevant, then there is simply no more room for extra criteria, which means that the criterion for moral personhood would not be really justified, since it would not be based on morally relevant reasons.

As a result of this, we have to conclude that moral personhood cannot be accepted as a meaningful concept. It can only be accepted as a superfluous concept. Otherwise, if moral personhood is not taken to be a superfluous synonym of moral consideration, then it is simply an unjustified device that distorts moral decision making.

In fact, most people are likely to reject that moral personhood can be a superfluous concept. Due to this, its acceptance will most likely generate a distortion that will just lead us to make wrong decisions.

Of course, we may prefer to use the term 'moral person' in a different way. We may think that moral persons are entities with some special interests, in particular an interest in living. But then, the same that was said above can be claimed here. If we just make our decisions on the basis of morally relevant criteria, then we will take into account justly the interests of different individuals. Fair consideration of interests is all that is needed here, without having to name those who possess a certain interest *i* with some name and those who possess another interest *j* with another one. Coining a term with such strong connotations such as 'person' to name those who possess a certain interest is likely, again, to distort moral decision making in cases in which it is used. The reasons for this are basically the same ones that we saw above for moral consideration as such.

Finally, when it comes to moral agency we might think that the same reasons could be pointed out. In this case, however, since moral agency is not about moral consideration, things would be different. Whether one is a person or not in this sense need not alter how we consider that individual. Nevertheless, even in this case the use of this term may be problematic. The reason for this has to do with the polysemy of the term 'moral personhood'. Since, as we have seen, it has been used to name both moral agents and morally considerable beings, even if we restrict our use to the former meaning confusion may arise. This may suggest the idea that moral agents are the entities that must be morally considered, or that they need to be considered in some special way. And this need not be so. In fact, I have claimed above, when I argued against moral anthropocentrism, that we should reject this view.

In light of all this, I have to conclude that the concept or personhood is problematic not only due to the common assumption being wrong. There are other reasons why it is highly questionable. In light of them, it seems we should better abandon altogether its use in moral philosophy. Moreover, in light of what we have seen, this conclusion is likely to apply in metaphysics as well, although in this paper I have focused on what happens in ethics.

THEORIES OF PERSONHOOD: GUILTY AS CHARGED?

Rui Vieira da Cunha

0. THE COMMON GROUNDS

The most common scheme on any introduction to the discussion on theories of personhood follows these steps[1], usually (although not necessarily) in this order: 1) to argue for a basic distinction that we make between persons and things (non-persons), 2) to claim that the characteristics/properties/qualities that make something a person are not exactly coincident with the ones that make something a human being/human animal, 3) to remember the etymology of the word, 4)to run through the various philosophical conceptions of personhood – typically starting with Locke[2] and ending with Parfit, and finally 5) to finish either arguing for the importance of the concept of person or personhood in virtually every area of our lives, from philosophy to everyday life, including law, ethics, religion, etc., or to finish by dismissing the concept altogether.

It is on the texts that finish with a dismissal of the concept that I want to focus my attention on this essay, whether those texts are introductory to the topic of personhood or not. There is a growing number of objections to the possibility and the practical use of personhood and theories based on it, particularly in bioethics. Chief on this attack has been the suspicion that many in the philosophical literature have expressed about the concept of person, both in its metaphysical and moral aspects[3]. The charges

[1] See, for instance, Newson (2007) or Goodman (2006).

[2] A controversial question here is whether there was a concept of person in ancient philosophy, even if the word, of course, was missing. On the topic, see the essays in Gill 1990. The religious perspective is also central, given the theological debates on the nature of Christ and the Trinity, and usually given its due relevance.

[3] I will postpone till the final section, where I discuss the bond between these aspects, the precise

leveled against this concept by different authors from different theoretical standpoints fall into four groups which I shall first try to discriminate – in doing so, I will count the over-simplification charge, the charge of vagueness/ambiguity, the cover-up/begging the question charge, and the irrelevance/superfluousness charge. In dealing with these objections, which will take me from sections 1 to 4, I will use Bert Gordijn's (1999) basic formulations of the charges, subsuming under the former many other similar objections from other authors. It is my claim that we can answer all the charges in a non-problematic way. My claim is that most of these charges are unjustified or, at least, that they could equally well be applied to many other concepts who play a fundamental role in philosophy, science, and in our own practical life. Of all the charges, the irrelevance charge will be given special consideration, because of its appearance of soundness. My ultimate goal, however, is that on a closer look that soundness proves to be merely an appearance.

My ultimate interest, however, is not so much on the objections themselves but on showing how they all share a deeper metaphysical *cum* moral question. In fact, we shall see that in each charge we will inevitably reach the point where the metaphysical and the moral aspects of personhood meet and that link is what will prompts us from one charge to the other, until we finally devote our full attention to it on section 5. To be precise, my crucial interest is on what kind of connection is there or should there be between metaphysical personhood and moral personhood or, to frame it in another possible manner, what is the link between the descriptive and the normative aspects of the concept of person. To pave our way into that discussion, let us begin with the charges put forth against the concept of person.

1. THE OVER-SIMPLICATION CHARGE

Putting this very simply (no pun intended), we are told that the use of the concept of the person is prone to simplifications. We are also told that such simplifications amount to the construction of "*all too simple black and white dichotomies like person/non-person or moral status/no moral status*". Central to the accusation is that this kind of dichotomies is unsuitable to the moral debates because "*Morality is too heterogeneous and varied to be fully grasped with the help of these simple dichotomies.*" (Gordijn 1999)

characterization of their nature, in order to render the reading more fluid. Until section 5, then, I will merely try to show that all the objections have a deeper grounding that entangles moral and metaphysical questions, but I will be somewhat succinct about them.

One kind of answering scheme that could be deployed here would be something along the following lines. Many other concepts over-simplify and are prone to dichotomies – think of organism, mind, knowledge, life, truth, justice, beauty, right, etc. – and yet we do not seem eager to relinquish those concepts. In fact, what some authors consider a weakness of the concept of person, its semantic richness, might be evidence of its long history and its layers of meaning and actually constitute an asset of the concept, as we shall see in sections 2 and 3.

One could also try another kind of reply, perhaps a less naïve one, and stress that the concept of person can admit of degrees of personhood[4]. There is nothing intrinsic to the concept of person to prevent us from sustaining that we can conceive of beings that are more or less persons than other such beings, depending on the qualities or properties or characteristics or attributes that flesh out the concept of person - Christian Perring, for instance, has interestingly argued that degrees of personhood follow from a Naturalist conception of personhood, such as the one revealed in Derek Parfit's views because whatever the criteria for personhood may be, "*they nearly all admit of degrees*" (Perring 1997: 181). Thus one can have criteria of general personhood, that is, criteria that distinguish between persons and non-persons, and criteria of particular personhood, distinguishing between one person and another.

As mentioned, this reply does make the concept of person a less simplistic, a less dichotomous tool, giving it some flexibility. The core of the objection, however, might not have been properly dealt with by this reply, in the sense that this still means that we would be carving the world at this certain specific juncture, between persons and non-persons, and only then would the degrees of personhood – general and particular -come into the picture. Maybe this is still not heterogeneous enough for moral purposes. The same could probably be said of another common theoretical strategy, that of distinguishing between potential and actual persons. And, of course, it could also be said that Perring's gradualist position[5], would be a slippery slope to politically and ideologically dangerous results: does it imply that those with long standing and irreversible psychological disability and mental illness have less personhood than persons with full

[4] In fact, Gordijn seems aware of this possibility - even if he doesn't pursue it, he mentions it on a footnote dedicated to Perrings' view (see Gordijn 1999:359, n. 26).

[5] I use the label gradualist to refer to a kind of position like Perring's, where degrees of personhood are postulated, by opposition to an absolutist theory, where there is no possibility for such degrees. This labeling was first brought to my attention from reading Miguens 2001: 141 e Miguens 2002: 392.

abilities?[6] The relevant point here, however, is that one can in fact carve the world between persons and non-persons and, simultaneously, not fall into a crude and one-dimensional moral view.

So one could add to the naïve reply the following: The over-simplification charge is itself based on a simplistic and erroneous assumption about the concept of person and personhood, namely, that its metaphysical dividing of the entities into persons and non-persons will always translate into a simplistic and dichotomous moral view. There is no necessary moral outcome of this kind of conceptual carving of the world – even if it may be true that we should carve it like this[7]. At this point we approach the deeper and main question that underlies this discussion: can we separate the descriptive aspect of the concept of person from its normative aspect? And if so, should we do it? Or to formulate it the other way around: is there any reason to sustain that metaphysical theories (in this particular case of personhood) have something to say about moral theories (of personhood)? A look at the second objection will lead us closer to the contours of these questions, about the moral implications of our metaphysical theories.

2. The vagueness/ambiguity charge

A previous note here: I'm obviously not claiming that ambiguity and vagueness are the same. I am lumping these accusations together because in my reading they stem from the same source: the historical richness of the reflection on the concept of person. A question like "When does a person begin?", pressing us for the answer that it is indeterminate, is meant to show that *person* is vague, because we have borderline cases or, to state it differently, we do not have a clear-cut boundary. "*George Bush and Fox News Incorporated are both persons she dislikes.*" could be a sentence meant to show that the word *person* has multiple meanings.

[6] A curious note here is that, for someone like Perring, this gradualist view might render our moral judgments even more complicated than they already are: how can we factor degrees of personhood into the already complicated calculus of utilities, if one is an utilitarian, for example? (1997: 191).

[7] Furthermore, authors such as Warren (2007) have clearly shown ways of construing the concept of moral status that don't rely on that simplistic person/non-person distinction, even if personhood is still a relevant criterion of moral status.

Because different authors have argued for different concepts of person, some authors, like Gordijn[8] and Beauchamp[9], have claimed that our research into the concept has left us with too many meanings of it and too many different kinds of beings in the world to which the concept can be applied.

In the case of Gordijn's accusation[10], the grounds are that there are too many (and perhaps too extensive) lists of necessary conditions for personhood for consensus[11] to arise. The accusation is not completely fleshed out but I believe one can understand the general idea and acknowledge the vagueness/ambiguity charge as one that simply states that 1) the concept of person can apply to different kinds of entities and 2) the concept of person admits of borderline cases.

Now, if this were all there is to it, we could answer it in the following way: many other concepts apply to different kinds of entities and admit of borderline cases – organism, mind, knowledge, life, truth, justice, beauty, right, etc. – and yet we do not seem eager to relinquish those concepts. Again, as in section 1, we would qualify this not as a weakness of the concept of person but as an asset, its semantic richness being evidence of its long history and its layers of meaning.

Amelie Oksenberg Rorty, on the other hand, is far more explicit in her presentation of the objection. On her view, and this is closer to the point *supra*, there is a philosophical dream "*that fundamental moral and*

[8] Gordijn (1999: 354-355): *"(...)a purely pragmatic use of the concept of the person as gathering the different qualities that transform an entity into a moral agent cannot be defended, since using the concept of the person only leads to confusion within the debate. This is, as I have already indicated, because the variety of lists of necessary conditions for personhood that the participants have in mind is so great, that the concept of the person is far from unambiguous. Therefore, using the concept does not contribute to mutual understanding and thus has no pragmatic use at all."* To be strict, Gordijn seems to be charging only moral personhood with that ambiguity that entails its unimportance even on a merely pragmatical attitude. After all, his aim is to show that *"The concept of the person is unsuited to be a central concept in bioethical debates"* (1999: 357).

[9] Beauchamp (2010: 256): "The vagueness of this concept is not likely to be dissipated by general theories of personhood, which will invariably be revisionary of the concept. Theories typically reflect the concept's vagueness and kindle more disagreement than enlightenment. They give us no more than grounds for a claim that there are alternative sets of sufficient conditions of personhood. The possibility of necessary and sufficient conditions of person in a unified theory now seems dim. The concept of person is simply not orderly, precise, or systematic in a way that supports one general philosophical theory to the exclusion of another. There is a solution to this problem of vagueness in the concept of person: Erase it from normative analysis and replace it with more specific concepts and relevant properties."

[10] Gordijn (1999: 352-353) identifies three Lockean influences on the concept of person and personal identity that are responsible for this confusion: 1) the absence of a clear ontological foundation in Locke's concept of the person; 2) the distinction between man and person; and 3) the grounding of personhood on consciousness.

[11] However, it is curious that Gordijn will also say that there is at least some consensus: that personhood is simply a matter of having certain properties, as we shall see later on.

political principles can be derived from the narrower conditions that define persons" [1990: 21]. One core element of this dream, in all its versions, is the goal of a single concept of a person. Such dream is nonetheless unachievable[12], since "the" concept of a person is also unachievable and three main factors concur to her conclusion: 1) historical factors (there are dramatically discontinuous changes in the characterization of persons), 2) anthropological-cultural factors (moral and legal practices heuristically treated as analogous across cultures differ so dramatically that they capture 'the' concept of personhood only vaguely and incompletely), and, most importantly, 3)functional factors - as Rorty puts it,

> "The various functions performed by our contemporary concept of person do not hang together: there is some overlap, but also some tension. Indeed, the functions that 'the' notion plays are so related that attempts to structure them in a taxonomic order express quite different norms and ideals." (1990: 22)[13]

It is true that Rorty ends her essay on an open note - stating that we are equally justified in denying one concept of person and accepting only "*highly regionalized functions that seemed, erroneously, to be subsumable in a structured concept*" (1999: 38) as we are in accepting one concept of person and concluding that its functions are at odds, rendering no decision procedures for resolving conflicting moral claims. However, one can only assume, from the overall tone, that she would agree with the general idea in Gordijn's criticism: that the concept of person has too many meanings and many of them are prone to confront us with borderline cases.

Specifying my assumption about Rorty's agreement: I would say that Rorty would endorse the charge of vagueness and ambiguity put forth by Gordijn in the sense that each function of the concept of person modifies the concept to the point that the kind of entities to which it applies are different from function to function and that, in some of these functions, at least, there are borderline cases (cases in which we are unable to determine,

[12] Notice that Rorty isn't exactly telling us not to draw moral conclusions from metaphysical theories. She is merely stating that *if*, as is often the case, our ambition of doing so depends on having only one concept of a person or having one meaning of it that achieves consensus, then we are doomed, because we'll never obtain that "one" concept. A similar warning, regarding some questions about the concept of person and the questions that relate to it, is made by Quante (2007).

[13] According to Rorty, there is overlap and tension between these various functions, each bearing a different relation to the class of persons and human beings, and each also with a different contrast class. The different functions aren't clearly named but we can summarize them in the following way: 1. moral, 2. legal, 3.agencial, 4. social, 5. narrative/existential, 6. biological, 7. mental/subjectival.

of the entity under consideration, whether it is or not a person).

Another example of someone who agrees with the basic tenets of this charge is Adam Morton:

> "What I doubt is that there is a single set of characteristics which would qualify a creature for intellectual and moral personhood. I think that we overestimate the simplicity and unity of the criteria we would apply in judging a creature's application for personhood, and I think that there are many more unclassifiable cases, in which there is no fact of the matter about whether or to what extent the creature in question is a person. (…) It is not just that I think the notion is vague, with wide fuzzy areas around many different edges. Its defenders would admit that. Rather, I suspect that there is no single concept there at all." (1990: 39-40)

For Adam Morton,

> "[What I will have shown is that] our use of the word is based on a set of rather indefinite family resemblances more than on simple and definite criteria [, and that the resemblances focus on parochial features of the human organism]." (1990: 41)

I believe many other philosophers (Harry Frankfurt[14], for instance) have at some point or another made this kind of observation: that we build the concept of person on the attributes (or qualities/properties/characteristics) of human organisms, somehow abstracting those attributes we deem worthy of higher valuing, conceptually detaching them from all the other characteristics of human organisms. Since those properties are usually able to be instantiated to different degrees, admitting of borderline cases, properties can be instantiated, at least conceptually, by other kinds of beings, it is easy to see how tagging the concept as anthropocentric can seem to be a definitive explanation of its vagueness and ambiguity and a decisive move to argue for its irrelevance. And yet there is still room of maneuver to recognize the anthropocentric ground of the concept, its vagueness and ambiguity and at the same time argue for its relevance, precisely because its anthropocentrism expresses our conceptual relation to the world and

[14] See Frankfurt 1971: 6.

what we value about it. At least conceptually, one can say that even if we start from our own parochial view, it is always possible to achieve broader horizons. The point is that we need to be aware of that connection between our conceptual tools and our evaluative practices. Again, like in the first objection, what is crucial is to recognize the link between metaphysical personhood and moral personhood. Is it the vagueness and ambiguity of the descriptive aspect of personhood that has moral implications or is it, on the other hand, the vague and ambiguous moral functions we wish to pursue with the concept of person that force our talk of vagueness and ambiguity of the concept?

3. THE COVER-UP/BEGGING THE QUESTION CHARGE

One of the strongest versions of this charge is made by Gordijn, who actually puts it in a very blunt way which can help us see the core of the objection:

> "[…] *a participant in a bioethical debate can simply choose a specific set of properties as being necessary for personhood in order to corroborate his own moral views* […] *This particular choice of certain qualities as being necessary conditions for personhood cannot be decisively criticized by his opponents, since there is no consensus on any ontology or metaphysics of the person that could deliver the necessary tools for such criticism*." (1999: 355)

The essence of this charge is that in the debates where the concept of person is called upon to decide on diverging claims about normative substantive issues, the party that calls on it is free to shape that concept at his own will, thereby free to choose a particular meaning (a particular set of properties or qualities), already imbued with normative assumptions. The subsequent use of the concept of person, although usually presented as value-neutral and so (expected to be) helpful in the debate, is consequently tantamount to a petition of principle. This "cover-up" use of the concept occurs because there is no independent external criteria of demarcation of qualities that are and those that are not necessary conditions for personhood – there is no consensus on any ontology or metaphysics of personhood.

The weakest version of this accusation is put forward by Rorty, who claims that it is possible that

> "*the various functions of the concept are sometimes at odds, that the concept of a person cannot function to provide decision procedures for resolving conflicts among competing claims for rights and obligations because it embeds and expresses just those conflicts.*" (1990: 38, my emphasis)[15]

Again, the same unsophisticated line of defense I have used to counter the first two charges can be deployed here: many other concepts lack independent external criteria of demarcation of their constitutive/essential qualities, many more long for consensus regarding their ontological nature (think of organism, mind, knowledge, life, truth, justice, beauty, right, etc.) and yet we do not seem eager to relinquish those concepts. In fact, what some authors consider a weakness of the concept of person, we can perhaps think of that as conceptual richness and so on… But is this enough? And, more importantly, isn't this accusation precisely pushing us to consider those deeper questions about the connection between metaphysical theories and moral theories?

Let us pause on the last period of that Rorty's citation. The claim there is that the concept of person embeds and expresses the clashes between competing claims for rights and obligations. Isn't this the same as saying that the uses we require of the concept somehow shape the very concept? And, since those uses are or have been mostly moral or normative uses, isn't this the same as saying – again – that the deeper question posed by this objection is that of the entanglement between the metaphysical/descriptive and the moral/normative aspects of personhood? Let us then move to the objection that brings us to that question in a more direct manner.

[15] Also: *"Both the arguments for excluding corporations and the left hemisphere of the brain and the arguments for including robots and Martians depend on normatively charged conceptual analyses."* (1999: 33). It should be noted, however, that, unlike Gordijn, Rorty's version of the story concedes the possibility that authors commit the error unconsciously: *"Indeed, because the classification has significant political and social consequences, we should not be surprised to discover that conceptual analyses of biological functions – particularly those presumed to affect intentional agency – are strongly, though often only implicitly and unself-consciously, guided by moral intuitions, ideology, and taste."* p. 33

4. THE IRRELEVANCE/SUPERFLUOUSNESS CHARGE

The core idea of this objection is that personhood is reducible to the beings in question or the entities under consideration possessing certain properties. After all, it is because of those properties that personhood is deemed morally relevant. Once one agrees on that (and, that, of course, is a metaphysical question that makes this objection harder to work on), we can probably do a better job in ethics by shifting our efforts to those properties directly and leaving personhood aside, or at least the objection goes.

The way Gordijn frames it, the charge takes the form of a conditional whose antecedent is deemed true: *If* the concept of person is reducible to the instantiation of those properties, *then* the use of the concept of person becomes unnecessary for moral purposes. This objection seems sound and appealing. One possible way of countering the objection is to try to reject the idea that personhood is reducible to these capacities[16], arguing that somehow being a person is more than being the conjunction of certain properties. It is true that this would preserve the importance of the concept of person (or of personhood) but at the expense of making it rather mysterious.

Another possible defense would be to accept reductionism about personhood but to reject the other underlying assumption in this objection: the supposedly necessary connection between the person-making capacities and moral status. So, we could accept that being a person is nothing over and above those properties, relinquishing any person-talk, deeming it gibberish, while still rejecting the idea that the moral status is grounded on those capacities. Whether this would allows us to preserve the conceptual and metaphysical importance of the concept of person I do not know. What is certain is that it would demand an additional theory on the grounds of moral status. We seem to be running in circles, seeing as this was already approached when discussing the first objection. My contention is that we are in fact running in circles, because all of these objections are grounded on the entanglement between the metaphysical and the moral aspects of personhood.

[16] The objection presupposes some sort of reductionist position – but what kind of reductionism is at play here? This is unclear in Gordijn's view, although I would argue that, pretty much like Beauchamp (2010), to whom I will turn latter, what we find here is an explanatory reductionism – see Johnston 1997.

But let us try to fill in what Gordijn means with this general idea of bypassing persons and looking directly into the capacities, in order to see if it seems at least promising in practical terms. The practical translation of this move, in building bioethical theories, for instance, would amount to asking questions like:

> "What is the moral significance of conception and nidation? How does the commencement of the nervous system influence the moral status of the foetus? Does the completion of the embryogenesis or the ability to survive independently of the body of the mother change the set of moral attributes of the unborn? What is the moral meaning of birth? What, if any, are the moral implications of being a human foetus instead of, for example, a chimpanzee foetus?" (Gordijn 1999: 356).

Thus, Gordijn claims, "*by analyzing bioethical problems concerning moral status without the concept of the person or a somehow disposed substitute*" these problems can be "*better and more clearly*" understood, just by focusing directly on "*the question of which properties and capacities within a being are a sufficient or necessary condition for which kind of moral status*" (1999: 355-357).

I believe however that there is a difference between 1) *we can do ethics or bioethics without the concept of person* and 2) *we can do a better job at ethics or bioethics without the concept of person*. Gordijn moves indistinguishably between these two meanings and although 1) seems plausible, 2) would demand more qualification. What would be a better and clearer analysis of bioethical problems? What would be our criterion for that judgment? In such contentious matters, I see no guarantee that removing personhood from the picture would yield wider consensus on discussions about abortion or euthanasia, for instance. Instead of discussing if a fetus is a person, we would be discussing if it has sentience, for instance, and when exactly does it start and, probably, what is sentience, and, even more likely, why is sentience morally relevant.

One author that has expressed some doubts about removing personhood out of the picture like that is Perring (1997: 188-189)

> "However, I think it is not so obvious that personhood "factors out" in such a simple way. The origin of the value of people's

> lives, their positive rights, their negative rights, and their moral standing is a very controversial matter. We at least need to consider some of the complexities of these issues in order to see how personhood and degrees of personhood are relevant to them. Even if it is possible to approach some of these moral issues without bringing in personhood, it may be simpler or more intuitive to start off from considerations of personhood and move on from there. For some issues, it is most natural to start off by considering personhood. I do not believe that we can tell in advance, with some general rule, where it is profitable to bring in personhood, and where it is simpler just to focus on the psychological properties of the persons involved, and bypass consideration of personhood. We have to go on a case by case basis. I cannot attempt an exhaustive review of the implications of the idea that personhood comes in degrees for all of medical ethics, but I will consider some of the more obvious ones." [1997: 188-189]

Truth be told, there is nothing different about this objection that would prevent us from answering it with the scheme we've used before to answer the other ones - many other concepts are reducible to certain qualities/properties/characteristics (again: organism, mind, knowledge, life, truth, justice, beauty, right, etc.) and yet we do not seem eager to relinquish those concepts. In fact, what some authors consider a weakness of the concept of person… You get the picture. We can of course sum this up by agreeing that being reducible does not by itself render the concept in question as unnecessary. But this would be too easy a fix, perhaps. The main crux is one that Gordijn does not pursue, at least not consistently. We need to turn to Beauchamp to understand the deeper problem here, one that started with Dennett's distinction between moral and metaphysical personhood and that is likened to my asking of the moral implications of our metaphysical theories.

5. PERSONHOOD: METAPHYSICAL AND MORAL (AND A PROVISIONAL CONCLUSION)

I have intentionally withdrawn from characterizing what I have interchangeably been referring to as metaphysical personhood or descriptive aspect of personhood and moral personhood or normative aspect of personhood. Nonetheless, I have tried to show that in each objection we inevitably reach the point where the metaphysical and

the moral aspects of personhood meet – about the over-simplification charge, I have argued that it is based on the erroneous assumption that the metaphysical dividing of the entities into persons and non-persons will always translate into a simplistic and dichotomous moral view; on the vagueness and ambiguity charge, along with its alleged anthropocentric foundation, I have argued for the relevance of the concept of person on the hypothesis that its anthropocentrism expresses our conceptual relation to the world and what we value about it and thus one is at troubles to separate the normative and the descriptive aspects of personhood; on the cover-up/begging the question charge, it was my assertion that perhaps the uses (mostly normative) we require of the concept of person somehow shape the very concept that we pretend to be merely descriptive; and, finally, on the irrelevance/superfluousness charge, I have repeatedly claimed for the need to recognize its roots in the descriptive/normative question.

It is now time to turn specifically, even if tentatively, to the said question of the connection between metaphysical personhood and moral personhood. One way of introducing the distinction and the possible connections is to resort to Dennett's words:

> "Does the metaphysical notion – roughly, the notion of an intelligent, conscious, feeling agent – *coincide* whit the moral notion – roughly, the notion of an agent who is accountable, who has both rights and responsibilities? Or is it merely that being a person in the metaphysical sense is a necessary but not sufficient condition of being a person in the moral sense? Is being an entity to which states of consciousness or self-consciousness are ascribed *the same* as being an end-in-oneself, or is it merely one precondition?" (1976: 176).

Another way is to resort to Beauchamp's perspective:

> "The different objectives of theories of persons can be clarified by a distinction between metaphysical and moral concepts of persons. Metaphysical personhood is composed entirely of a set of person-distinguishing psychological properties such as intentionality, self-consciousness, free will, language acquisition, pain reception, and emotion. The metaphysical goal is to identify a set of psychological properties possessed by all and only persons. Moral personhood, by contrast, refers to individuals who possess properties or capacities such as moral

> agency and moral motivation. These properties or capacities distinguish moral persons from all nonmoral entities. In principle, an entity could satisfy all the properties requisite for metaphysical personhood and lack all the properties requisite for moral personhood." (2010: 247):

Now, it is rather telling that to both of these authors the metaphysical/descriptive aspect of personhood consists exclusively of psychological properties. I see no principled reason to assume a psychological perspective right from the start. Even if we tend to think that metaphysical personhood is all about psychological properties, as I suspect many of us do, we should not carve that assumption into the discussion. Conceptually, metaphysical personhood (or the descriptive/metaphysical aspect of personhood) might very well rely on purely bodily properties. The question, if we are to take person as a serious ontological contender, is not one about the precise cognitive or psychological properties that beings that fall under the concept have. If we really have an interest in metaphysical personhood (or rather, in the metaphysical nature of persons), we would do better in leaving aside, for the time being, the question on what exactly are the psychological properties, and begin by asking if there is indeed a special category of beings in the world that are picked out by that concept.

To put the question in a different form: are persons the sort of things that appear in a list of the basic items of the world? Or are there better candidates, like events, minds, bodies, etc.? Again in a different manner: are persons the sort of entities that we can call primitives or can they be reduced to more basic things? This is the real metaphysical question about persons and it is one which, although will probably have to tackle with the lists of properties, need not be confused with it. Likewise, I see no need to assume from the beginning of the inquiry that *person* is what beings like us (human animals or human beings) most essentially are nor that only beings like us qualify as persons (nor, needless to say, the opposite of any of these views) – all of this is open to discussion, even if, again, it is very likely that we will tend to agree with those assertions.

Now, is there a constraint on metaphysical theories of personhood that they *have to* say something about moral personhood? Or is there a negative constraint, that they remain neutral about their moral implications? Beauchamp has argued that metaphysical personhood entails neither moral personhood nor moral status. If this is true, we might perhaps be free to defend that the personhood theorist can adopt a gradualist position, without any concern for the politically and ideologically dangerous results

that we discussed above, when answering the first charge and reviewing Perring's perspective. But would we still be interested in the metaphysical nature of persons, if we would somehow come to a consensus that it has no bearing on moral issues?

To conclude with Rorty (1990: 38):

> "For all practical and theoretical purposes it doesn't matter whether the concept of a person has multiple and sometimes conflicting functions, or whether there is no single foundational concept that can be characterized as *the* concept of a person. As long as we recognize that such appeals are, in the classical and unpejorative sense of that term, rhetorical, we can continue to appeal to conceptions of persons in arguing for extending political rights, or limiting the exercise of political power."

Throughout this essay I have discussed authors who support a dismissal of the concept of person. I discriminated the objections leveled against this concept by different authors into four groups of charges - the over-simplification charge, the charge of vagueness/ambiguity, the cover-up/begging the question charge, and the irrelevance/superfluousness charge. In dealing with these objections, I have claimed that we can answer all the charges in a non-problematic way, since most of them unjustified or, at least, they can equally well be applied to many other concepts who play a fundamental role in philosophy, science, and in our own practical life. Of all the charges, I have given special consideration to the irrelevance charge, because of its appearance of soundness, and particularly to show that it is but an appearance.

My ultimate interest, however, was not so much on the objections themselves but on a deeper metaphysical *cum* moral question that I argued underlies all of them. My crucial interest on this essay was on what kind of connection is there or should there be between metaphysical personhood and moral personhood or, to frame it in another possible manner, what is the link between the descriptive and the normative aspects of the concept of person. I have not put forward any substantive claim on this subject, other than methodological warnings about presuppositions one should avoid unless one whishes to skew the inquiry from the start.

REFERENCES

Beauchamp, Tom. 2010. The Failure of Theories of Personhood. In Tom Beauchamp, *Standing on Principles: Collected Essays*, 229-260. Oxford University Press: New York (first published in the Kennedy Institute of Ethics Journal 9:4, pp. 309–324, 1999).

Dennett, Daniel. 1976. Conditions of personhood. In *The Identities of Persons*, ed. Amelia O. Rorty, 175-196. University of California Press: Berkeley.

Frankfurt, Harry. 1971. Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68 (1): 5-20.

Gill, Christopher. 1990. *The Person and The Human Mind – Issues in Ancient and Modern Philosophy*. Clarendon Press: Oxford.

Goodman, Michael F. 2006. Persons. In *Encyclopedia of Philosophy: Oakeshott-Pressupposition*, 2nd edition, Donald M. Borchert, 237-244. 2nd edition, Macmillan: Detroit.

Gordijn, Bert. 1999. The Troublesome Concept of the Person. *Theoretical Medicine and Bioethics* 20: 347-359.

Johnston, Mark. 1997. Human Concerns without Superlative Selves. In *Reading Parfit*, ed. Jonathan Dancy, 149-179. Blackwell: Oxford.

Miguens, Sofia. 2001. Problemas de Identidade Pessoal [Personal Identity Problems]. *Revista da Faculdade de Letras: Filosofia* 18: 139-164.

Miguens, Sofia. 2002. *Uma Teoria Fisicalista do Conteúdo e da Consciência – D. Dennett e os debates da filosofia da mente* [*A Physicalist Theory of Content and Consciousness – D. Dennett and the philosophy of mind debates*]. Campo das Letras: Porto.

Morton, Adam. 1990. Why there is no Concept of a Person. In *The Person and The Human Mind – Issues in Ancient and Modern Philosophy*, ed. Christopher Gill, 39-59. Clarendon Press: Oxford.

Newson, Ainsley J.. 2007. Personhood and Moral Status. In *Principles of Health Care Ethics*, 2nd edition, eds. Richard Ashcroft, Angus Dawson, Heather Draper and John McMillan, 277-283. John Wiley & Sons – London.

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

Perring, Christian. 1997. Degrees of Personhood. *The Journal of Medicine and Philosophy* 22: 173-197.

Quante, Michael. 2007. The Social Nature of Personal Identity. In *Dimensions of Personhood*, ed. Arto Laitinen and Heikki Ikaheimo, 56-76. Imprint Academic: Exeter.

Rorty, Amélie Oksenberg. 1990. Persons and *Personae*. In *The Person and The Human Mind – Issues in Ancient and Modern Philosophy*, ed. Christopher Gill, 21-38. Clarendon Press: Oxford.

Warren, Mary Anne. 2007. *Moral Status – Obligations to Persons and Other Living Things.* Clarendon Press: Oxford.

JULIA DRIVER'S 'VIRTUES OF IGNORANCE'

Luís Veríssimo

Happiness (…) is something final
and self-sufficient, and is the end of action.
(Aristotle, NE, 1097b)

INTRODUCTION

The following paper pretends to discuss Julia Driver's account of virtue. Julia Driver is Associate Professor of Philosophy at Dartmouth College. In her work Uneasy Virtue she challenges the classical virtue ethics by refusing to accept that moral virtue must involve intellectual excellence. Instead of a classical approach, she adopts a consequentialist account of virtue – virtue is a character trait that systematically produces good consequences. I shall be arguing against some features of Driver's account of virtue, which arise largely from what I believe to be a misarticulated criticism of Aristotle. I shall object Driver's criticism of the traditional interpretation of the Aristotelian notion of virtue as involving a knowledge condition. Driver intends to show that there is a special class of virtues – 'virtues of ignorance,' that not only do not require that the agent possesses intellectual excellence, but even require him to be ignorant of some features of the world. The paradigm of this sort of virtues is Modesty. Modesty, she claims, is to be understood as an underestimation of one's real worth, so if modesty is a virtue, there are virtues that do not involve knowledge, but ignorance. I intend to demonstrate that either these character traits are not, at all, virtues, or they do not involve ignorance. Additionally, I will propose a different approach to the concept of modesty that allows this trait to be considered a virtue without the ignorance condition.

I. IN DEFENCE OF THE KNOWLEDGE CONDITION

In Chapter 1, of Uneasy Virtue, Julia Driver starts by highlighting the fact that traditional Aristotelian notion of virtue requires cognitive excellence. She shows that even contemporary virtue ethicists, such as John McDowell and Martha Nussbaum, subscribe this central feature of Aristotle's account of virtue – virtue is "correct perception", it implies "getting things right" in each occasion. This view is associated with another important feature of virtue ethics – particularism. What makes an act right or wrong is a matter of such complexity, due to the variety of situations one might face throughout life, that it cannot be fully captured by a system of abstract rules. The virtuous person must be aware of the morally relevant features in each case where a moral decision is called for. In order to acquire this special sensitivity, the individuals must train themselves through the actual exercise of virtues. Once you acquire this ability, you will have your eyes opened to 'reasons for action', that would otherwise be veiled to you and you will know what to do, at the same time you'll be motivated to do it. This is important, because when describing virtue ethics as evaluational internalist or externalist, Driver considers Aristotle's view hybrid (mixed), because it involves internal, as well as external, states. To be fully virtuous, one must not only know what he is doing (knowledge, which is acquired only through performance, just like riding a bike), but also to be succeeded in what he does. In Aristotle's words:

> (…) an act is not performed justly or unjustly or with self control if the act itself is of a certain kind, but only if, in addition the agent has certain characteristics as he performs it: first of all, he must know what he is doing; secondly, he must choose to act the way he does, and he must choose it for its own sake; and in the third place, the act must spring from a firm and unchangeable character. (Aristotle, NE, 1105a28-34)

So, one might ask, "Is knowledge (namely, "practical knowledge", or phronesis) a necessary condition for virtue?" Aristotle's answer to this question would surely be – Yes, if an agent is virtuous, then virtuous actions will spring from what he knows to be true. This affirmative answer will be the first target of Driver's criticism of traditional virtue ethics. Knowledge is not only disposable as a sufficient condition for virtue, she claims, but even as a necessary condition for it. Driver's move consists in presenting counterexamples to this Aristotelian thesis, thus she must find character traits that are considered virtues, though they do not involve knowledge, but ignorance. She christened this class of virtues 'virtues of ignorance' and

describes what she considers to be the paradigmatic expression of this sort of virtue, Modesty, in the following terms:

> Modesty has at least two senses. There is the sexual sense of modesty, usually considered a womanly virtue, which primarily consists in a chaste and unassertive countenance. There is also a more usual sense that is associated with self-deprecation and underestimation of one's self-worth. It is this later sense that concerns me (…) (Driver, 2001, p. 16)

And later, in the same paragraph she claims:

> A modest person underestimates self-worth. (...) Modesty is dependent upon the epistemic defect of not knowing one's own worth. (...) modesty is a virtue, therefore, undermines the view that no virtue is crucially connected to ignorance. (ibid., p. 16)

The first interrogation to strike me as I read these lines was: "Why should underestimation be a virtue?" And I saw no reason for this judgment, at all. On the contrary, there would be things I would be able to do, were I to believe in my true capacities, that would not only bring benefits to others, but also to myself. Driver provides the following answer:

> One thing that indicates to me that these traits [modesty, among other 'virtues of ignorance'] are virtues is the fact that, when recognized, they are valued by others as traits that morally improve the character possessing them. (ibid., p. 36)

However, this seems rather insufficient, in at least two ways: 1. There are societies where these traits are not valued by others as traits that morally improve character; and 2. Whatever characteristics that just happen to be valued by others would count as virtues – which seems rather arbitrary. Other problems spring from such a definition: 3. How many of those 'others' would be necessary to give the trait the status of virtue? and 4. Couldn't they just happen to be wrong, about the trait in question?

A more elaborate answer is sketched in these lines:

> In the case of modesty, the modest agent is modest because he underestimates himself, and this leads to some good that is valued by those he interacts with (e.g., an easing of tensions,

> lack of jealousies). (...) So what explains why a given trait is a virtue is simply that it is conducive (more conducive than not) to the good. (ibid., p. 26)

What makes these traits moral virtues is their tendency to produce beneficial effects. Though it would be controversial (and at this point premature) to claim that good effects are definitive of virtue, good effects are strong evidence for the presence of virtue. (ibid., p. 36)

Though more promising this answer is, as Driver admits, controversial, since it defines virtue in a consequentialist way, and this idea will not be argued for until Chapter 4, later on the book. But, for the purposes of this Section, I will work with this definition. It should be enough for me to disentangle virtue and ignorance, in Driver's own terms.

Drives states that:

> A desired feature of any account of modesty is that it explains the oddity of
> 1. I am modest. (ibid., p. 17)

The problem of (1) is that it "seems to be oddly self defeating", in other words "I can be modest, but I cannot know it" (ibid.). After spelling out this requirement of any account of modesty she proposes four approaches to the concept of modesty and picks the one she considers to deal with statement (1), without generating other complicated issues.

The first suggestion identifies modesty with a certain type of behaviour, namely, the careful avoidance of boastfulness. This deals with what sounds strange in (1), since anyone who said (1) would be bragging. This approach has other issues to solve: if behaviour is a sufficient condition for modesty, someone that doesn't brag would be modest, even if he overestimates his self-worth (as long as he doesn't have the opportunity to show signs of bragging, e.g., because he is alone in a desert island).

The second account of modesty presented by Driver is that the modest person knowingly understates his self-worth. The problematic issues of (1) are dealt with, because by uttering (1) while knowing himself to be modest, the agent would not be understating his self-worth, and, therefore, not being modest. Julia Driver refuses this view, for it provides an account of 'false modesty', rather than 'genuine modesty'.

The problem here is that given the above quoted definition of virtue ("what explains why a given trait is a virtue is simply that it is conducive (more conducive than not) to the good" (ibid., p. 26)), 'false modesty' can be considered a virtue. Unless additional reasons are presented, the virtue of modesty is not necessarily connected to ignorance, since one might be aware of his own worth but still understate it, in order to, for example 'ease of tensions', 'encourage others to believe in their own worth' or simply 'avoid envy'. Driver calls this trait 'false modesty', for it implies lying about one's estimation of self-worth. But she provides no additional reason to explain why this trait – 'false modesty', should not be considered a virtue, namely, the virtue of modesty. We might concede that there are other additional reasons to avoid lying to others (e.g., the fact that lying is often not conducive to the good), and thus there are reasons to create some sort of aversion to lying under normal circumstances in a systematic fashion, maybe this could exclude 'false modesty' as a virtue, but Driver does not provide any additional reason of this sort.

The third analysis of modesty, presented by Driver is the 'underestimation' account of modesty. In this view, modesty consists in a disposition to underestimate one's self-worth. Driver picks this sense of modesty because it allows her to create the connection between virtues and ignorance, and thus criticize the traditional view that virtue is knowledge. In this view, a "modest person is ignorant, to a certain degree, with regard to his self-worth" (ibid. p. 18), so knowledge is not to be considered a necessary condition for virtue. She goes even further, by stating that this ignorance must not be occasional or accidental, it has to be a disposition to ignore certain facts about self-worth. Otherwise anyone who seems modest could end up overestimating or accurately estimating self-worth given the right sort of evidence.

Another account of modesty Driver rejects consists in considering modest someone who doesn't take full credit for his achievements, because he realizes that if it were not for luck, or the efforts of others, he would not have been nearly as successful as he is. Driver finds this identification between the virtue of modesty and the recognition of luck factors or the help of other's rather unsatisfactory, because that sort of recognition is not a sufficient, nor necessary, condition for modesty. It's not sufficient, because we can imagine cases like the following:

> [a] criminal who recognizes that, brilliant criminal mastermind, though he may be, it's sheer good luck that he hasn't ended up in jail. (ibid., p. 22)

Although he recognizes that luck factors are involved, the criminal is not, necessarily, modest. On the other hand, it's not a necessary condition for modesty, because genuine cases of modesty, may or may not involve recognition of one's good luck, I may believe that my achievements rest upon my skills, and still underestimate, both, my skills and my accomplishments.

I suggest an account of modesty entirely different from the alternatives explored by Julia Driver. First I don't accept the assumption that any reasonable account of modesty must deal with the oddity of "1) I am modest". I do not find (1) odd or self-defeating. One might say "I am modest" without committing a performative contradiction. This happens because, in my account, the virtue of modesty consists in the disposition to frame one's self-worth in a wider picture, and thus relativize his achievements. This does not involve ignorance, but an awareness of at least one of the following:

1. Awareness of full potential (or Perfectionism): I may be good at X, but still I could be much better (e.g., I may be a good piano player, but still I could be much better);

2. Awareness of the talent of others in the same area: I may have been distinguished as good (or the best) in some area of expertise, but I know that there's a huge probability that someone as good as I, or even better than me exists somewhere in the vastness of the world (e.g., I may have been distinguished as one of the greatest European piano players, but I'm sure there are lots of great piano players in anonymity (or ignored by the critics));

3. Awareness of luck and train factors: I may be good at X, but that's due to the combination of personal dedication with the right sort of opportunities (e.g., I may be a relatively good piano player, but that's just because I spent some hard work, time and resources with piano playing (anyone with the opportunity to do the same as I did, would very likely become such a proficient piano player as I am));

4. Awareness of the variety of talents: I may be good at X, but not good in lots of other areas (e.g., I may be a very good piano player, but I'm not nearly as good philosopher as I could be (as you are)).

In sum, the virtue of modesty involves the ability to recognize that however extraordinary I may appear in the eyes of others, if we consider the variety of talents and of talented people around the world, then I am just an ordinary person. The only ignorance involved in such a conception of modesty is the same kind of ignorance expressed in the famous Socratic maxim: "All I know is that I know nothing", this sort of knowledge became known as Docta Ignorantia or learned ignorance, since what it really means is that one's aware of what lies beyond his present knowledge. Something similar happens with modesty, since it consists, in fact, in some sort of awareness of what lies beyond our present capacities (not only as far as our area(s) of expertise is(are) concerned, but also in what concerns other areas that we have not yet, or not at all dedicated ourselves to).

This account might be objected in the same way that the 'recognition of luck factors account' (previously analyzed) is, since (3) above refers to something similar to what's advocated by that account. However, since i) I am considering this awareness among others, and ii) it includes reference to personal effort and training, it does not state exactly the same as the 'recognition of luck factors account (or help of others)', for it does not state that the recognition of luck factors (or the help of others) constitute a sufficient, or necessary, condition for modesty. In my account, (3) is just an instance of what might be properly identified with virtue, which is the ability to relativize self-worth, because, although one's aware of his own worth, one's also aware of what lies beyond his present capacities.

This account of modesty can be easily accommodated by Driver's definition of virtue, since the effects of such a disposition would bring benefits to others and my-self, without implicating the waste of talents due to lack of self-confidence (as Driver's 'underestimation' account does). As Driver puts it, modesty could actually produce negative effects, since it could give us the frustrating feeling that the modest person is somehow neglecting his potential, we might be lead to believe that it is unfair that such dispositions are possessed by someone who doesn't esteem having them. Instead of easing tensions or avoiding envy, this could generate the opposite effect. My account, on the other hand, explains why modesty can, among other things, 'ease of tensions', 'encourage others to believe in their own worth', 'motivate individuals to develop their talents', or even, 'avoid envy'. Let's consider the instances of the 'relativization' of one's self-worth previously presented: (1), as well as (2), can motivate the individual to pursue further development of his present skills; (3) can encourage the individual, as well as others around him, to persist/engage on training

and focusing on opportunities; (4) may avoid envy and ease tensions, by leading others to believe in their own worth, as well as it might motivate the modest person to seek training in other areas.

This approach to the notion of modesty, as the paradigmatic 'virtue of ignorance', may be transferred to other examples of this class of virtues, such as 'blind charity', 'blind trust', 'the disposition to forgive and forget' and, last but not least, 'impulsive courage', either they are not virtues, or they don't involve ignorance.

Let's start by blind charity. Driver defines it in these terms:

A person who is in blind charity with others is a person who sees the good in them but does not see the bad. *Blind* charity differs from charity in that it is usually the case that when is merely charitable toward another, one favours that person in some respect, *in spite of* perceived defects or lack of desert. Blind charity is a disposition not to see the defects and to focus on the virtues of persons. (ibid., p. 28)

I cannot understand why blind charity is to be considered a virtue in the light of Driver's standard – "what explains why a given trait is a virtue is simply that it is conducive (more conducive than not) to the good" (ibid., p. 26). One might legitimately ask: What good is produced by blind charity? The only answer I feel temped to present is that blind charity may, often, be useful, because when someone fells under its target tends to change his flawed dispositions in order to conform to that distorted 'idyllic' view of his character. However, this answer does not fully satisfy me, because we can easily imagine that, at least quite as often as that happens, one might reinforce his flawed dispositions because it seems that people around him are not aware of them, which allows him to take advantage of others. This shows that blind charity may bring, at least as much harm to others, as it may benefit them, and thus this trait does not qualify as a virtue at all. One might say that the same applies to plain charity, so it mustn't be considered a virtue either. That's not entirely accurate, and a comparison between blind charity and plain charity shall be useful in spelling out why this assumption is mistaken. Charity is a virtue because it corrects the ignorance condition of blind charity. Someone who displays the virtue of charity is able to relativize the bad of others, because he is aware of their flawed dispositions, at the same time he is aware of their good dispositions. So the charitable person produces the good by reinforcing the good dispositions of others, leading them to believe in their goodness, instead of focusing on the

negative. One might object that, as in the case of blind charity, this could also lead to the reinforcement of flawed dispositions, because charitable persons will always focus on the best of others, allowing them to keep taking advantage of this sort of person. This is not true though, because the genuine charitable person has this special awareness to the relevant features of other's characters to know exactly when charitable behaviour is called for. This also explains how trust and forgiveness can count as virtues. The virtuous person knows when to trust (or to forgive) someone that has betrayed him is likely to reinforce his trust-worthy behaviour, or not. This doesn't display ignorance, or a belief against the evidence, it just shows that the virtuous person has his eyes opened to other evidence that outweigh other available evidence. Anyway, it's a profound degree of awareness, and not ignorance, that is required. In as far as forgetfulness is concerned, I don't see why it must be deeply connected to forgiveness, or to virtue. When discussing an example in which Jones betrayed his friend's trust, Driver says that "forgetfulness is crucial to the sort of forgiveness that Jones aspires to. (…) The person who feels that another as forgotten the harm is far more likely to feel comfortable around that person and to feel really forgiven" (ibid., p. 32). This might be true, but it's also true that that person is also more likely to be disposed to repeat that unpleasant behaviour. As I see it, if I cannot recall what harm as been done to me, then I cannot fully forgive anyone. Forgiveness involves knowing that someone has done something wrong, but still be able to relativize the harm done, or to frame it within a wider picture of the subject's character. The mistakes we make shape our future dispositions, we shall keep them in mind, because they taught us that we should avoid repeating them, so forgetting about them does nothing to produce the good.

Finally, let's dedicate some time to the analysis of 'impulsive courage'. I shall be discussing its status as a virtue. Here's what Driver has to say about it:

> Impulsive courage is an interesting example of a virtue of ignorance because it seems to involve inferential ignorance alone. The impulsively courageous person possesses certain relevant facts of his situation, yet fails to put these facts together in order to reach the conscious conclusion that he himself is in danger.
>
> A good illustration of this sort of person is one who, perhaps, fears for the persons trapped inside a burning building but does

> not fear for himself, since he fails to perceive any danger to himself, he isn't overcoming any fear or sense of danger. He is acting impulsively. (ibid., p. 33)

These lines are accompanied by the following end-note:

> James Wallace would argue that this is not a true case of courage, since it violates a condition he considers necessary, i.e., that the agent believe that the action he or she is performing is dangerous to the self. See his Virtues and Vices (Ithaca, NY: Cornell University Press, 1978), 78 ff. (ibid., p. 116)

I agree with Wallace in the fact that if there's no awareness of the danger, the agent cannot be considered courageous (thought this is not inconsistent with an evaluation of the act itself as being courageous). If we still consider that having a disposition to enter a burning building to save others, ignoring the danger to one's self is a good trait (since it brings about good consequences), we must think if it is the ignorance of salient features about the situation that we have in mind or simply the disposition to help others who happen to find themselves in dangerous situations. I believe the ignorance of the salient features might endanger the agent, preventing him to bring about as much good as he could, had he represented the danger to himself, as well as he represented the danger to others. So what we consider a good conducting character trait, and thus a virtue, in these cases is not the ignorance of salient features about the situation, but perhaps the disposition to help others, who happen to find themselves in dangerous situations (or other character traits related to the behaviour in question), it's does have to be necessarily connected to ignorance of any kind.

Julia Driver quotes Aristotle, in order to demonstrate that even he would recognize this trait to be a virtue.

> (...) it is a mark of even greater courage to be fearless and unruffled when suddenly faced with a terrifying situation than when the danger is clear beforehand. For the reaction is more prone to be due to a characteristic, since it is less dependent on preparation. When we see what is coming we can make a choice based on calculation and guided by

> reason, but when a situation arises suddenly our actions are determined by our characteristics. (Aristotle, NE, 1117a17– 22)

This passage allows her to conclude that:

> When faced suddenly with a situation that calls for action, the agent may just act – without pausing to register salient facts of the situation and weigh alternatives. That is, he acts without due regard to the danger. In such situations, this is the courageous way to proceed. It is really the only fruitful way to proceed even though it is fraught with more risk, perhaps, than situations where deliberation is possible and is taken advantage of.
>
> (...)
>
> On Aristotle's view, choice involves deliberation and "search." Thus, in making a choice, the agent weights alternatives. Yet in this particular case, though the man has other options (e.g., run away), he doesn't consider them. He does not "... make a choice based on calculation and guided by reason ..." He simply responds to the situation. (Driver, 2001, p. 34)

I agree with pretty much everything in the first paragraph of this quotation, but I don't infer from these lines that it is the ignorance involved in such behaviour that is worthy of esteem. We can conciliate the view that this sort of behaviour is chosen, and thus "involves deliberation and "search"", with the idea that at the moment, no activity of alternative weighing occurs. The deliberation took place before being confronted with the situation. The virtuous agent has voluntarily cultivated the right dispositions, so that when the time comes he sees what is required of him almost immediately, without wasting time in useless calculations. There really is no actual weighing of alternatives, because he adjusted his dispositions in a way we might say that his eyes are opened for reasons for action that simply silence all reasons to act otherwise. This doesn't mean that this is the case in every situation where opposite dispositions, or reasons for action, present themselves. There is a difference between silencing all reasons to act otherwise and simply outweighing them (which may also occur).

The difference may be spelled out with an analogy with what happens with a scale. If we put different weights on each side of a scale, the scale will tilt towards the heavier weight. The lighter weight will be 'outweighed.' This may also occur with reasons for action. I might, sometimes, have opposite reasons for action, and one of them outweighs the other. In this case I might feel sorrow for neglecting one of them, although I know it was the right thing to do. Aristotle's example of the captain who is forced to throw into the see the cargo of his ship in order to save the crew, is a fine illustration of what outweighing is all about. Now let's imagine that same scale, but this time let's pretend that certain weights, when used, require that all weights on the other side of the scale must be removed. This happens for instance when, although someone has a strong reason to avoid hitting others, this person finds herself in the situation of catching someone attacking his children and hits the attacker. The reason to save her children simply silences her reason to avoid hitting others. This person will not feel sorrow for ignoring his other reasons for action; they just don't apply in cases like this.

So, back to courage, what happens is that although the courageous person has a reason to avoid entering burning buildings, she also has a strong reason to help the persons inside of it, in a way that, perhaps, is able to silence her reason to avoid entering burning buildings. Since no ignorance is involved, we may conclude that impulsive courage fails to be considered a 'virtue of ignorance.' And since we have done pretty much the same to all other members of this class of virtues, we may also conclude that this is an empty class. If there are no virtues of ignorance, no counterexamples to the virtue as knowledge thesis have actually been presented. On contrary, we seem to have demonstrated that such as virtue is defined in Driver's account, the more awareness there is, the more likely good effects will be produced.

CONCLUSION

Throughout this paper I might sound highly critic about Driver's consequentialist account of virtue, and vividly advocating for a traditional Aristotelian account of virtue. But the fact is that Driver has called my attention to the possibility of developing a consequentialist account of virtue and this had a tremendous impact on the way I see normative ethics, because although I felt reluctant about modern moral philosophy, in both deontological and consequentialist traditions, I could not fully subscribe traditional virtue ethics. I agree with Robert Adams when he claims that:

> The subject of ethics is how we ought to live; and that is not reducible to what we ought to do, or try to do, and what we ought to cause or produce. It includes just as fundamentally what we should be for and against in our hearts, what and how we ought to love and hate. (Adams R. M., 1985)

The problem of modern moral philosophy lies in the fact that these theories seem to be more concerned with what we ought to do, than with what sort of person we ought to be. In my opinion, the answer to the first question is derivative from an answer to the second. To know what we ought to do, we must be concerned with knowing what sort of person we ought to be. As previously noted, this cannot be reduced to a system of abstract rules concerning what we should (should not) do, for it includes a deeper evaluation of one's character. This evaluation involves his character traits, or dispositions for action, his attitudes, emotions and desires, and these lie beyond the sphere of action. Both Kant's perfect moral agent – exclusively motivated to act by sheer respect to the rational moral law: the Categorical Imperative – and Mill's happiness maximizer can be so incredibly flawed in as far as their character is concerned, that would hardly be worthy of moral praise and admiration, even if their actions were immaculate. Imagine for instance a 'continent sadist', it is not his actions, but his character that is deplorable. This happens because morality is also about reliability. We are interested in knowing if we can count on others, as well as in ourselves, to act appropriately when suddenly facing a situation where a moral decision is called for. Thus, I believe that a normative theory must be primarily concerned with the notion of virtue.

However, traditional virtue ethics faces serious objections that seem irresoluble within its own framework, such as those posed by Schneewind (Schneewind, 1990), Louden (Louden, 1984) and consequentialist authors like Philip Pettit (Baron, Slote, & Pettit, 1997). In addition, there is a fundamental feature of virtue ethics that I find inconsistent, which is the fact that it recognizes happiness as the only one final end, and yet fails to consider it should be pursued in an universalistic, rather than individualistic way. Adopting the universalistic perspective would make the account consequentialist (see footnote 2), and this would not only solve this problem of the traditional account, but also numerous other problems frequently associated to virtue ethics. However, I don't believe that Driver has fully captured the potential of a consequentialist account of virtue, and that's what has taken me to advocate that knowledge is, at least, a necessary condition for virtue. In my opinion, if one doesn't know the meaning of his actions, then he fails to be virtuous.

REFERENCES

Adams, R. M. (1976). Motive Utilitarianism. The Journal of Philosophy, Vol. 73, No. 14, , 467-481.

— (1985). "Involuntary Sins". The Philosophical Review 94 , 3-31.

Aristóteles. (NE). Nicomachean Ethics. (D. Ross, Trans.) New York: Oxford University Press, 1980.

Baron, M. W., Slote, M., & Pettit, P. (1997). Three Methods of Ethics: A Debate. Oxford: Blackwell Publishing.

Bentham, J. (1789). Introduction to the Principles of Morals and Legislation. New York: Hafner Publishing Co., 1948.

Blackburn, S. (1998). Ruling Passions:a theory of practical reasoning. Oxford: Oxford University Press.

Brandt, R. (1992). "The Structure of Virtue". In Morality, Utilitarianism, and Rights (pp. 29-311). New York: Cambridge University Press.

Brink, D. (1989). Moral Realism and the Foundation of Ethics. New York: Cambridge University Press.

Crisp, R. (2004). "Pleasure is All That Matters". Think , 21-29.

— (1992). "Utilitarianism and the Life of Virtue". Philosophical Quarterly, vol. 42, N.º 167 , 139-160.

Damasio, A. (1995). Descartes' Error. New York: Avon Books.

Driver, J. (2001). Uneasy Virtue. Cambridge: Cambridge University Press.

Foot, P. (1988). "Utlitarianism and the Virtues". In S. Scheffler, Consequentialism and its Critics (pp. 224-42). Oxford: Oxford University Press.

Goleman, D. (1996). Emotional Intelligence: Why it Can Matter More Than IQ? London: Bloomsbury.

Hare, R. M. (1981). Moral Thinking: Its Levels, Method and Point. Oxford: Oxford University Press.

Hume, D. (1748). Investigação sobre o Entendimento Humano. Lisboa: INCM, 2002.

— (1751). Investigação sobre os Princípios da Moral. (J. P. Galvão, Trans.) Lisboa: INCM, 2005.

— (1740). Tratado da Natureza Humana. Trad. Serafim da Silva Fontes. Lisboa: Fundação Calouste Gulbenkian, 2002.

— (1960). Treatise of Human Nature. Oxford: Clanderon Press.

Hurka, T. (1992). "Virtue as Loving the Good". In E. F. ed. Paul, J. F. Miller, & J. Paul, The Good Life and the Human Good (pp. 149-68). New York: Cambridge.

Hursthouse, R. (1999). On Virtue Ethics. Oxford: Oxford University Press.

Jacobson, D. (2005). "Seing by Feeling". Ethical Theory and Moral Practice, N.º 8 , 387–409.

Kagan, S. (1998). Normative Ethics. Oxford: Westview Press.

Kant, I. (1785). Fundamentação da Metafísica dos Costumes. Lisboa: Lisboa Editora, 2003.

Louden, R. B. (1984). "On Some Vices of Virtue Ethics". American Philosophical Quarterly 21 , 227-36.

McCloskey, H. J. (1965). "A Non-Utilitarian Approach to Punishment". Inquiry , 249-63.

McDowell, J. (1978). "Are Moral Requirements Hipothetical Imperatives?". Proceedings of the Aristotelian Society, suppl. vol. 52 , 13-29.

— (1998). "Virtue and Reason". In J. McDowell, Mind, Value and Reality. Cambridge: Cambridge University Press.

Nozick, R. (1974). Anarchy, State and Utopia. New York: Basic Books.

Rachels, J. (2003). Elementos de Filosofia Moral. Trad. F. J. Gonçalves. Lisboa: Gradiva, 2004.

Railton, P. (1984). "Alienation, Consequentialism, and the Demands of Morality". Philosophy and Public Affairs , 134-71.

Schneewind, J. B. (1990). "The Misfortunes of Virtue". Ethics 101 , 42-63.

Searle, J. (1984). Mind, Brains and Science. Cambridge: Harvard University Press.

Slote, M. (1997). "Agent-Based Virtue Ethics". In M. Slote, & R. Crisp, Virtue Ethics (pp. 239-262). New York: Oxford Univesity Press.

Turing, A. (1950). "Computing Machinery and Intelligence". Mind .

Wittgenstein, L. (1953). Philosophical Investigations. (G. E. Anscombe, Trans.) Oxford: Blackwell Publishing.

AUTHORS

Guerreiro, Vítor, University of Porto, MLAG Researcher

Horta, Oscar, University of Santiago de Compostela

Jesus, Paulo, University of Lisbon, Philosophy Center Researcher

Locatelli, Roberta, University Paris I Panthéon-Sorbonne

Machado Vaz, João, University of Porto, MLAG Researcher

Magalhães Carneiro, Tomás, University of Porto, MLAG Researcher

Miguens, Sofia, University of Porto, Professor, Philosophy Department, MLAG Researcher, Director of MLAG

Morando, Clara, University of Porto, MLAG Researcher

Ramalho, Daniel, New University of Lisbon, Institute for Philosophy of Language Researcher

Santos, João, University of Porto, MLAG Researcher

Teles, Manuela, University of Porto, MLAG Researcher

Vaaja, Tero, University of Jyväskylä

Veríssimo, Luís, Universidade de Évora

Vieira da Cunha, Rui, University of Porto, MLAG Researcher