ISSN 2280-7853

PHENOMENOLOGY AND MIND

THE ONLINE JOURNAL OF THE FACULTY OF PHILOSOPHY, SAN RAFFAELE UNIVERSITY

7

IUSS Press
Istituto Universitario di Studi Superiori di Pavia

Registrazione del Tribunale di Pavia, n° 6 del 9/07/2012
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

n.7 - 2014

PHENOMENOLOGY AND MIND

THE ONLINE JOURNAL OF THE FACULTY OF PHILOSOPHY, SAN RAFFAELE UNIVERSITY

NATURALISM, THE FIRST-PERSON PERSPECTIVE AND THE EMBODIED MIND

Lynne Baker's Challenge: Metaphysical and Practical Approaches

Edited by Massimo Reichlin

Topics
Phenomenology and Social Ontology; Ethics and Political Theory; Cognitive Neurosciences, Philosophy of Mind and Language, Logic; Aesthetics, Metaphysics and History of Ideas.

Frequency
2 issues per year

Finito di stampare nel dicembre 2014 presso DIGIGRAPH
Str. Martinella, 96 - 43124 Loc. Alberi - Parma

IUSS Press
Istituto Universitario di Studi Superiori di Pavia

Phenomenology and Mind practices double blind refereeing and publishes in English.

CONTENTS

INTRODUCTION

Massimo Reichlin (Università Vita-Salute San Raffaele)
Introduction 10

SESSION 1. INVITED SPEAKERS

Lynne Rudder Baker (University of Massachusetts Amherst)
Cartesianism and the First-Person Perspective 20

Dermot Moran (University College Dublin and Murdoch University)
Defending the Transcendental Attitude: Husserl's Concept of the Person and the Challenges of Naturalism 30

Michael Pauen (Berlin School of Mind and Brain and Humboldt University)
How Naturalism Can Save the Self 44

Mario De Caro (Università Roma Tre and Tufts University)
Two Forms of Non-Reductive Naturalism 54

Katherine Sonderegger (Virginia Theological Seminary, Alexandria)
Naturalism and the Doctrine of Creation 64

Roberta De Monticelli (Università Vita-Salute San Raffaele)
Haecceity? A Phenomenological Perspective 74

Michele Di Francesco (Istituto Universitario di Studi Superiori, Pavia), Massimo Marraffa (Università Roma Tre), Alfredo Paternoster (Università di Bergamo)
Real Selves? Subjectivity and the Subpersonal Mind 90

Massimo Reichlin (Università Vita-Salute San Raffaele)
First-Person Morality and the Role of Conscience 102

CONTENTS

SESSION 2. CONTRIBUTED PAPERS

Alfredo Tomasetta (Istituto Universitario di Studi Superiori, Pavia)
We are Not, Fundamentally, Persons 114

Marc Andree Weber (Albert-Ludwigs-Universität Freiburg)
Baker's First-Person Perspectives: They Are Not What They Seem 120

Sofia Bonicalzi (Università di Pavia)
Does Reductivist Event-causal Compatibilism Leave Anything out? Lynne Baker's *Reflective-Endorsement* and the Bounds of the Traditional Analyses of Moral Responsibility 128

Alan McKay (The Queen's University of Belfast, Northern Ireland)
Constitution, Mechanism, and Downward Causation 136

Treasa Campbell (New Europe College, Bucharest)
A Humean Insight into the Epistemic Normativity of the Belief in the Self 144

Bianca Bellini (Università Vita-Salute San Raffaele)
Towards a Faithful Description of the First-Person Perspective Phenomenon: Embodiment in a Body That Happens to Be Mine 152

Patrick Eldridge (Katholieke Universiteit, Leuven)
Observer Memories and Phenomenology 160

Gaetano Albergo (Università di Catania)
The First-Person Perspective Requirement in Pretense 168

Giuseppe Lo Dico (Università Cattolica, Milano)
Introspection Illusion and the Methodological Denial of the First-Person Perspective 176

Valentina Cuccio (Università di Palermo)
The Notion of Representation and the Brain 184

INTRODUCTION

Massimo Reichlin (Università Vita-Salute San Raffaele)
Introduction

MASSIMO REICHLIN
Università Vita-Salute San Raffaele
reichlin.massimo@unisr.it

INTRODUCTION

The papers collected in this issue of *Phenomenology and Mind* were presented at the Spring School on "Naturalism, First-Person Perspective and the Embodied Mind" that was held at San Raffaele University, Milan in June 2014. As in the tradition of these philosophical schools, the meeting centred on the work of an outstanding living philosopher, namely, on Lynne Rudder Baker's philosophical views, and particularly on her recent book on *Naturalism and the First-Person Perspective*.

Beside the keynote speaker, there were seven invited speakers from four different countries, and ten contributed papers by scholars from five different countries, that were selected in a double-blind review process from a set of twenty-eight abstracts. The contributed papers subsequently underwent a double-blind review process when submitted in their full version.

Baker's 2013 book deals with a considerable number of important philosophical issues: most directly, the metaphysical one concerning the tenability of a scientifically-driven general worldview such as strong naturalism, but then also on many other topics: from the definition of our essential identity as persons to the specific characterisation of a robust first-person perspective in terms of "I-thoughts", from the criticism against eliminativist theories of the self, such as Metzinger's and Dennett's, to the discussion on Frankfurt-style compatibilism and moral responsibility. The papers presented at the conference discussed all aspects of Baker's proposal, and the presence and generosity of the author stimulated much lively discussion among senior and junior scholars.

This was not the first time that Lynne Baker came to visit the Faculty of Philosophy at San Raffaele: she had already been with us in May 2007 and, since then, relationships have strengthened, particularly with the main organiser of the School, *i.e.* Roberta De Monticelli. The papers collected in this volume are therefore a homage to the significant work of a philosopher and also an act of gratitude for an ongoing and lasting friendship.

In the first paper of this issue, Lynne Baker presents an overview of the main idea of *Naturalism and the First-Person Perspective*, particularly stressing the distance between her defence of an irreducible first-person perspective (FPP) against strong naturalistic and reductionist approaches, and the traditional Cartesian view according to which the mind is a separate substance, autonomous from the body. The difference between humans and non humans, she claims, lies in the possession of a robust first-person perspective by the human individuals who master language, and in the remote capacity to develop linguistic abilities that characterises human infants. Contrary to Descartes', this approach insists that persons are not isolated thinkers or non-social entities, but members of linguistic communities; it does not view persons as pure minds, but as necessarily embodied; it does not attribute to the FPP any epistemic primacy, since its aim is ontological and not epistemological; it does not claim to be without presuppositions; it is not dualist; it accepts that many of the primary kinds of things are intention-dependent; it does not postulate any inner transparent realm to which every individual has infallible access. The remoteness of Descartes' perspective from hers would be even greater, Baker claims, if we should accept that Descartes was committed to the goal of the absolute conception of reality, as claimed by Williams.

Baker defends a metaphysical view that she calls "quasi-naturalism". In the following paper, Dermot Moran defends a much more anti-naturalistic approach to the person or self: the phenomenology view, which is characterised not only by the content of experience, but mainly by the modes of experience. Phenomenology, in fact, is a non-naturalistic, transcendental approach, according to which objects reveal themselves from the standpoint of attitudes. All kinds of objectivity, therefore, are constituted accomplishments, reached by a certain kind of intentionality. This is why, according to phenomenology, persons cannot be wholly naturalised, for it takes a personalistic attitude to recognise and understand them, and a naturalistic attitude fails to do the job. From the personalistic attitude, persons can be seen as sense-makers and position-taking individuals, who have a relation to their history and are embodied, social and intersubjective agents. It is mainly the capacity to take a stance on oneself and one's life, according to Husserl, that characterises persons: the ego, as he said, is a centre of affections, actions, interests and habits, on which it exerts ownership and control. Persons are intentional agents and embodied sense-makers, who are involved in an intersubjective horizon of other persons. Persons, moreover, cannot be understood only as autonomous, rational beings, for, as embodied beings, they share a world of feelings and emotions.

On the opposite side of the spectrum of philosophical positions, Michael Pauen defends the view according to which naturalism need not endorse an eliminativistic position on the self: on the contrary, it can save its concept, analysing it in terms of the lower-level phenomena that contribute to its implementation. The main problem with the self, according to Pauen, is that every act of reflective identification presupposes self-awareness; this means that the self cannot emerge from reflection, but must be originally given in some kind of pre-reflexive self-awareness. This in fact happens, first, through the body-scheme of the core-self, *i.e.*, the pre-personal, affective capacity to recognise our body as our body, and to integrate its parts; second, through the theory of mind, that is, the capacity to adopt someone else's perspective and contrast it with our own. This perspective-taking strategy is much more cognitive than the body-scheme one, for it presupposes the ability to distinguish beings whose perspective you can take – *e.g.* humans and non-human animals – from those whose perspective you cannot take. Now, small children are able to make distinctions between the living and non-living, the human and non-human at a very early, pre-linguistic stage of development; and the same goes for mimicking con-specifics' behaviour and distinguishing emotions in facial expressions. All this comes well before twelve months of age, before the self-awareness evidenced by the traditional mirror test, before the capacity to master first-person pronouns and other forms of language, and before the capacity to correctly attribute beliefs to others. The conclusion is that a naturalistic defence of the self takes as central the capacity to recognise yourself as yourself: this is not a single ability, but a graduate one, progressively developing for emotions, perceptions and beliefs.

The metaphysical discussion between different forms of naturalism and other general philosophical approaches is tackled by Mario De Caro, who provides a very detailed overview of the different positions in the spectrum of general metaphysical worldviews, from strict naturalism to supernaturalism. Clearly sympathising with liberal forms of naturalism, De Caro explores the differences between this widely held philosophical position and Baker's proposal of "near-naturalism". He underlines several points of agreement between Baker and liberal naturalists such as Putnam, pointing to Baker's neutrality concerning the existence of supernatural properties as the main feature of genuine disagreement: this neutrality, he contends, is too liberal, and cannot be accepted even by liberal naturalists.

The metaphysical issue of supernaturalism also echoes in Katherine Sonderegger's paper that offers a theological discussion of the Biblical doctrine of creation in the light of modern and contemporary naturalistic approaches. She notes that a line of 'reductionism' concerning the conception of nature has always influenced the discussion on the interpretation of God's work in the creation: the ancient atomistic doctrine, trying to identify the deepest building blocks of reality, is mirrored by the attempt to understand God's work as the creation of basic particles or elements, from which all particular objects are derived. This reductionist approach is also echoed in the medieval notion of 'prime matter', as the simple element entering into the composition of every created entity, and largely influences the philosophies of the modern era and the contemporary thermodynamic conception of the cosmos. In the face of this all-embracing naturalism, Sonderegger contends that Christians have reasons to continue to talk of God's work as the creation of individuals, not of particles, force-fields or natural laws: this is because theology must not aim at harmonising the Bible with astrophysics, but at guiding humans in the acknowledgment of the grace and gift that comes from the richness and plurality of the natural world.

A different kind of metaphysical question is taken on by Roberta De Monticelli who discusses Baker's theory of personal identity. According to Baker, all informative theories of personal identity are third-personal, and therefore miss the importance of the FPP; this is why Baker's theory accepts circularity as a consequence of the fact that the conditions of personal identity cannot be stated in non-personal terms. De Monticelli, on the contrary, believes that a first-personal but informative theory can be formulated if the issue of personal identity is understood in the context of a wider account of personal individuality. De Monticelli's main point against Baker is that there is more to having a first-person perspective than a capacity for self-reference, since pure self-reference is uninformative about whose self it is referring to, and Baker's reference to haecceity as the decisive property for being a particular person is blatantly circular. Self-*knowledge* transcends self-*consciousness*, and aims at clarifying the individual 'whatness' of a person. De Monticelli argues for a different sort of 'haecceitism', according to which having an individual nature is just as much essential to one's personhood as having a first-person perspective. In the wake of Leibnizian 'superessentialism', she views haecceity as an individual essence, *i.e.* a constraint on possible (co) variations of the properties that a person may possess while remaining that same individual; accordingly, personal identity across time consists in sharing this substantial unity, or 'Scotistic heacceity'.

A very different perspective is embraced by Michele Di Francesco, Massimo Marraffa and Alfredo Paternoster who jointly author a paper on *Real Selves? Subjectivity and the Subpersonal Mind* that was presented at the School by Di Francesco alone. Their aim is to discuss the issue of subjectivity putting aside any metaphysical perspective, and adopting an epistemological and explicative attitude. Contrary to Baker's approach (but sharing her basic intention), they build their vindication of the self not on a metaphysical defence of the first-person perspective, but on a pluralistic reading of the nature of the science of the mental and on the assumption of pluralism at the explanatory level. Following the bottom-up approach common to contemporary cognitive science – an approach that moves from the automatic and pre-reflexive construction of representations of the external world, through the bodily self-monitoring, to self-consciousness – the authors suggest that a robust theory of the self must not understand the conscious subject as a primary subject, rather as emerging from the mechanisms of the neurocognitive unconscious. This, however, is not to accept its epiphenomenality; a robust self, emerging as the ongoing result of a narrative self-constructing process, is in fact necessary to explain the phenomena of intentional action and self-understanding presupposed by commonsense psychology and social science. Moreover, according to the authors this theory is fully consonant with contemporary (neuro)cognitive science, that acknowledges the psychodynamic component of the process of narrative self-construction and the stable internalisation of our narrative identity in the structures of our personality.

In my own paper, I explore some aspects of Baker's distinction between a rudimentary and a robust first-person perspective, and show that moral agency requires the second, more complex property. The failure to acknowledge the first-personal, reflective character of moral judgment accounts for the weakness of most contemporary naturalistic reconstructions of morality, that identify the automatic responses of our "sentimental brain" as the basic fact of our moral experience. I suggest that an appropriate view of morality should emphasise the genuinely first-personal element of possessing a conscience, as distinct from the possession of a moral sense, interpreted in a Humean fashion. I then proceed to criticise the neatness of Baker's distinction between the rudimentary and the robust FPP, suggesting that Baker excessively downplays the role of embodiment in her account of what it is for the same first-person perspective to be instantiated across time.

A variety of philosophical questions emerging from Baker's work is also faced by the ten contributed papers that follow. In the first of these, Alfredo Tomasetta tackles the metaphysical questions posed by Baker's contention that "person" is a primary kind and, specifically *our* primary kind. The thesis implies that we are fundamentally persons, and that we cannot fail to be persons without ceasing to exist altogether. If this were true, Tomasetta claims, human persons would have the same persistence conditions of God, the angels, and Cartesian souls, which allegedly are persons as well. But this implication is indefensible, since it is clear that these other entities cannot share our persistence conditions. Baker needs an argument to deny that the possession of a common primary kind implies having the same persistence conditions. However, the three arguments discussed by the autor fail, and this suggests that Baker's main thesis is unsubstantiated.

A different metaphysical point is raised by Marc Andree Weber, who argues that Baker's conception of the FPP is not a clear and natural view as it may seem. Firstly, she does not distinguish between synchronic and diachronic self-attributions of first-person reference: she clearly presupposes our persistence through time, but this is not necessarily implied by the FPP. Moreover, it is not clear that the capacity to make self-attributions guarantees the truth of this self-attribution, or that it implies indivisibility or unduplicability. In hypothetical scenarios of fission cases Baker suggests that there is a fact of the matter as to which person shares the original person's FPP (even though we may not know the right answer), simply presupposing that being the same person is having the same FPP; but in such cases, to decide which later person shares the original FPP is theoretically undecidable and practically unhelpful. Weber suggests a different account, according to which an FPP is predicated of a mereological sum of moments of consciousness, with no entity unifying them: this would be a reductive account, in that it reduces the persistent to the momentary, but would preserve Baker's irreducibility of the mental to the physical.

Two more papers are devoted to Baker's treatment of action. Sofia Bonicalzi discusses Baker's view concerning moral responsibility, suggesting that Baker's insistence on the first-person perspective improves on standard Frankfurt-style compatibilist accounts, which fall prey to the syndrome of the disappearing agent, *i.e.* make the agent a mere bystander of causal factors over which she has no control. However, Bonicalzi claims that, even though Baker's insistence on the FPP allows to refer opposing mental states to oneself, thus generating the impression of causing one's choices, nothing proves that this picture is not a *post-factum* illusory reconstruction. Also in Baker's reformulation, therefore, compatibilism cannot make sense of the concept of accountability, which is essential for an adequate understanding of responsibility. Responsibility implies that the agent has control on her actions and this seems to require the assumption of irreducible agential properties.

Alan McKay criticises Baker's view on downward causation between intention-dependent (ID) causal property-instances and the objects and properties of non-ID, physical world, suggesting that the idea that mental content, *qua* content, has effects in the physical world is incoherent. According to McKay, our manifest view of a physical causal relation implies a transfer of energy of some kind: this paradigmatic causation is norm-free, causally closed, productive, intrinsic, and involves the operation of mechanisms, whereas an ID causal relation presents none of these characteristics. Baker's insistence that ID causation is of the same basic kind as lower-level causation obscures deep differences between the two. This is not to deny our ordinary intuitions about the existence of ID causation: according to McKay, these intuitions can be defended by claiming that the causal *relations* between ID causes and effects are constituted by manifest physical causal relations in favourable circumstances. This means that the physical causal relations are transformed, in the context of a complex relational milieu, into a quite different causal nexus, constrained by such factors as inference, justification, purpose, and desire.

A peculiar, non ontological strategy for providing a justification of our belief in the self is explored in the paper by Treasa Campbell: it is the epistemic strategy that builds on Hume's descriptive account of "natural beliefs" to show that the belief in the self enjoys a peculiar kind of epistemic justification. Campbell shows that natural beliefs play the role of hinges, on which all our other questions and doubts turn; this is why, with Wittgenstein, we cannot but grant them non-evidential warrants. This strategy promises to develop adequate warrant for our belief in the self while circumventing the ontological domain.

Acknowledging the importance of Baker's defence of the phenomenon of the FPP from naturalistic attacks, Bianca Bellini stipulates three criteria for what she calls a faithful description of a phenomenon: consistency with the experience of the phenomenon, consistency with the phenomenon's appearance and transcendence, and consistency with the essential traits of the phenomenon, as considered from the viewpoint of the phenomenological reduction. Her discussion charges Baker's account for failing to satisfy the second and third criterion: indeed, the FPP, as reconstructed by Baker, does not embrace an *essential* trait of the first-person perspective phenomenon, that is, the *phenomenological* distinction between *Leib* and *Körper*.

A distinctive phenomenological approach is also at the heart of Patrick Eldridge's paper that builds on Husserl's phenomenology of recollection, and particularly on his distinction between intentional and inner consciousness, to tackle the problem of observer memories. Observer memories are ordinarily distinguished from field memories in that they are not recollections from the first-person point of view, but from the third person perspective, that is, memories in which we are spectators of ourselves. Philosophers like Husserl, who insist that the FPP is a necessary feature of mental phenomena, have a problem in explaining this kind of memories, and may be tempted to deny their existence. According to Eldridge, however, observer memories are genuine forms of recollection that involve an original and peculiar form of self-intention, which is self-objectification. Therefore, this phenomenon is not a counter-example to Husserl's view that self-identity and pre-reflective self-consciousness are vital structuring elements of mnemic experience. Notwithstanding, it shows that self-consciousness is displayed on a spectrum from immediate, immanent self-identification to quasi-exterior-representation.

A more empirical inclination can be found in Gaetano Albergo's paper, analysing the phenomenon of pretense play in children, which he considers as an early manifestation of the first-person perspective. In the wake of some points also stressed by Pauen, he suggests that the activity of pretense presupposes intentionality and is evidence of an early manifestation of self-awareness. In fact, the rich phenomenology of pretense and the priority of agency over both cognitive representation and the conceptualisation of the self-world dichotomy, suggest that a primitive self-consciousness is present in pre-linguistic stages of human development. According to Albergo, therefore, Baker's insistence on the central role of language for the acquisition of self-consciousness is not justified by the facts.

Also devoted to the empirical side of the debate on the FPP is Giuseppe Lo Dico's paper, discussing the naturalistic rejection of introspection as an unreliable method in psychology. A large part of the psychological literature, he reports, assumes the self/other parity account, according to which knowledge of one's own and of others' mental states is equally indirect – the argument for this conclusion being that most of our mental life is unconscious and that verbal reports are *post-hoc* theories of what is supposed to happen in the mind. Lo Dico reviews evidence showing that data coming from verbal reports, if adequately treated, cannot be defined as illusory or confabulatory and can be legitimately used in psychological theory. He concludes that subjects' introspective or verbal reports should be taken much more seriously than they presently are, and that the subject's ability to adopt a FPP should be considered as well. This probably means that the idea of psychology as a fully naturalised science must be seriously revised.

The last paper, also dealing with empirical issues, is Valentina Cuccio's discussion of the relationship between the mechanism of embodied simulation and the notion of mental representation. Embodied simulation is the activation of the neural circuits controlling certain actions and perceptions, when the subject is not actively engaged in them. The recently proposed notion of mental representations in bodily format suggests the identification of these representations with the activation of the mirror mechanisms that give rise to embodied simulation. According to the author, the definition of embodied simulation in terms of mental representations is problematic, because embodied simulation does not allow to clearly distinguish between the content and the format of the representation, or to identify the subject of the mental representation. Mechanisms of embodied simulation are sub-personal processes, crucially involved in our understanding of others; to define them in terms of mental representation presupposes a strongly reductionist view that, in the light of Baker's work on FPP, is unsubstantiated.

I wish to thank the Scientific Committee of the Spring School for their organising efforts: particularly, Roberta De Monticelli, Francesca De Vecchi, Elisabetta Sacchi, Emanuele Bottazzi. Francesca Forlé and Sarah Songhorian. I also would like to thank the many collegues who accepted to review the abstracts presented at the School and the full papers collected in this issue. Both the Spring School and this collection of papers would not have been accomplished but for the generous and tireless work of Francesca Forlè and Sarah Songhorian, who helped with more or less everything that was needed.
On behalf of the Research Centre *Persona*, and of the Faculty of Philosophy at San Raffaele University I warmly thank them for their invaluable support. Finally, many thanks also to Bianca Bellini and Laura Caponetto for their precious work with the editing of the papers, and to Silvia Tossut for the revision of the English form.

SESSION 1

SESSION 1

INVITED SPEAKERS

Lynne Rudder Baker (University of Massachusetts Amherst)
Cartesianism and the First-Person Perspective

Dermot Moran (University College Dublin and Murdoch University)
Defending the Transcendental Attitude: Husserl's Concept of the Person and the Challenges of Naturalism

Michael Pauen (Berlin School of Mind and Brain and Humboldt University)
How Naturalism Can Save the Self

Mario De Caro (Università Roma Tre and Tufts University)
Two Forms of Non-Reductive Naturalism

Katherine Sonderegger (Virginia Theological Seminary, Alexandria)
Naturalism and the Doctrine of Creation

Roberta De Monticelli (Università Vita-Salute San Raffaele)
Haecceity? A Phenomenological Perspective

Michele Di Francesco (Istituto Universitario di Studi Superiori, Pavia), Massimo Marraffa (Università Roma Tre), Alfredo Paternoster (Università di Bergamo)
Real Selves? Subjectivity and the Subpersonal Mind

Massimo Reichlin (Università Vita-Salute San Raffaele)
First-Person Morality and the Role of Conscience

LYNNE RUDDER BAKER
University of Massachusetts Amherst
lrbaker@philos.umass.edu

CARTESIANISM AND THE FIRST-PERSON PERSPECTIVE

abstract

Descartes's influence is so great that it is often assumed that any philosopher who emphasizes a first-person perspective, as I do, must be a Cartesian. I want to challenge the assumption that I am a Cartesian by setting out my view of the first-person perspective and its importance for being a person. Then, I shall enumerate the ways in which my conception of the first-person perspective differs from Descartes's. Finally, I shall consider an alternative interpretation of Descartes, proposed by Bernard Williams. According to the alternative interpretation, Descartes was aiming at a wholly objective "absolute conception" of natural reality. I shall argue that, because of the first-person perspective, no "absolute conception" can be a full account of natural reality.

keywords

Cartesianism, first-person perspective, Bernard Williams, absolute conception

René Descartes has a good claim to be the originator of first-personal philosophy. Descartes's first-person outlook permeates his philosophy. Indeed, Descartes begins his epistemological inquiries by examining his own beliefs to discover which ones might be false. Even today, Descartes's influence is so broad that it is often assumed that any philosopher who emphasizes a first-person perspective, as I do, must be a Cartesian.

I want to challenge the assumption that I must be a Cartesian by setting out my view of the first-person perspective and its importance for being a person. Then, I shall enumerate the ways in which my conception of the first-person perspective differs from Descartes's. Finally, I shall consider an alternative interpretation of Descartes, according to which he was aiming at a wholly objective "absolute conception" of natural reality, and I shall argue that such an absolute conception cannot be a full account of reality.

1. The First-Person Perspective

On my view, a human person begins existence constituted by an organism, but is not identical to the organism that constitutes her. For purposes of this paper, I shall leave the essential feature of embodiment aside. What matters here is another essential property of persons, a first-person perspective. A first-person perspective is a dispositional property that members of the kind *person* have essentially.

A first-person perspective is a complex property that has two stages, a rudimentary stage that a person is born with and a robust stage that a person develops as she acquires a language. At the rudimentary stage, a first-person perspective is a nonconceptual capacity shared by human infants and nonhuman animals. It is the capacity of a conscious subject to perceive and interact with entities in the world from a first-personal "origin". At the robust stage, a first-person perspective is a conceptual capacity displayed by language-users; it is the capacity to conceive of oneself as oneself from the first-person, without identifying oneself by a name, description or third-person demonstrative.

Born with a rudimentary first-person perspective and a remote (or second-order) capacity to develop a robust first-person perspective, a human person gets to the robust stage in the natural course of development. As she learns a language, a person acquires numerous concepts, among which is a self-concept that she can use to conceive of herself as herself in the first-person. At the rudimentary stage, she can do things intentionally; at the robust stage, she can conceive of herself as doing things. At the rudimentary stage, she can perceive things in the world; at the robust stage, she can conceive of herself as perceiving things. Although the robust stage of the first-person perspective requires language, it is exhibited throughout one's life in characteristically human activities – from making contracts to celebrating anniversaries to seeking fame by entering beauty contests.

So, a person with a robust first-person perspective can manifest her personhood in a much richer and more variegated way than can an infant who has only a rudimentary first-person perspective. What makes you now – with your robust first-person perspective – the same person that you were when you were an infant – with only a rudimentary first-person perspective – is that there is a single exemplification of the dispositional property of having a first-person perspective both then and now – regardless of the vast differences in its manifestations over the years. For example, an infant may manifest a first-person perspective (at the rudimentary stage) by drawing back from a looming figure, and an adult may manifest a first-person perspective (at the robust stage) by making a will. A human person from infancy through maturity until death (and perhaps beyond) is a single exemplifier of a first-person perspective – whether rudimentary or robust. Now in greater detail.

2. The Rudimentary First-Person Perspective

Let us start with the rudimentary first-person perspective. The stage of the rudimentary first-person perspective is shared by human and nonhuman animals; the rudimentary first-person perspective connects animals that constitute persons with other animals. A human infant is a person constituted by a human animal. An infant is born with minimal consciousness and intentionality, which are the ingredients of a rudimentary first-person perspective. A person comes into existence when a human organism develops to the point of being able to support a rudimentary first-person perspective. The person constituted by the organism – the new entity in the world – has a first-person perspective essentially.

The rudimentary first-person perspective does not depend on linguistic or conceptual abilities. The rudimentary first-person perspective is found in many biological species, perhaps all mammals, and seems to be subject to gradation or degrees. Among species, consciousness and intentionality seem to dawn gradually (from simpler organisms) and the rudimentary first-person perspective seems to become more fine-grained as it runs through many species in the animal kingdom.

Darwinism offers a great unifying thesis that “there is one grand pattern of similarity linking all life” (Eldredge, 2000, p. 31). Considered in terms of genetic or morphological properties or of biological functioning, there is no discontinuity between chimpanzees and human animals. In fact, human animals are biologically more closely related to certain kinds of chimpanzees than the chimpanzees are related to gorillas and orangutans[1].

Human infants, along with dogs, cows, horses and other non-language-using mammals, also have rudimentary first-person perspectives. So, my view recognizes the continuity between human animals that constitute human infants and higher nonhuman animals that constitute nothing. In this way, the biological continuity of the animal kingdom is unbroken.

But wait! If that is so, then why do I say that a person is only constituted by an animal and not identical to an animal? For this reason: Although there is no discontinuity in the animal world – no *biological* discontinuity – the evolution of human persons (perhaps by natural selection) does introduce an *ontological* discontinuity.

The ontological discontinuity between persons and animals lies in the fact that a human infant – who is not identical to the organism that constitutes her – has a remote capacity to develop a robust first-person perspective. A nonhuman organism that does not constitute a person may have a rudimentary first-person perspective (as chimpanzees do), but it has no remote capacity to develop a robust first-person perspective. And this remote capacity distinguishes persons from all other beings.

A remote capacity is a second-order capacity to develop a capacity[2]. For example, a healthy human infant has a *remote* capacity to ride a bicycle. She does not yet have the capacity to ride, but she does have the capacity to acquire the capacity to ride a bike. When the young child learns to ride a bicycle, she then acquires an *in-hand* capacity to ride a bicycle; that is, in certain circumstances (when she has a bicycle available and wants to ride), she actually rides a bicycle and manifests her capacity to ride a bicycle. She may never learn to ride a bike, in which case her remote capacity to ride a bike would not issue in an in-hand capacity to ride a bike. Similarly, even though a remote capacity to develop a robust first-person perspective is an essential property of persons, a person may never actually develop a robust first-person perspective (if, for example, the person had a case of severe autism).

The point is that an infant person has not only a rudimentary first-person perspective, but also has a remote capacity to develop a robust perspective; otherwise the entity would not be a human person.

So, the ontological difference between persons and animals lies in the robust first-person perspective and in the remote capacity to develop one. In pre-linguistic persons (like babies), the rudimentary stage of the first-person perspective brings with it the remote capacity to develop a robust first-person perspective. Nonhuman animals have no such remote capacity.

So: What makes persons unique is that only persons have robust first-person perspectives. (If dogs learned to talk and acquired the capacity to conceive of themselves in the first-person, a new kind of entity would come into existence, canine-persons. But the point would still hold: only persons have robust first-person perspectives).

To sum up: The rudimentary stage of a first-person perspective is a nonconceptual stage that entails consciousness and intentionality. The rudimentary stage is what ties us persons to the seamless animal kingdom; the robust stage is what makes us ontologically and morally unique. Now let us turn to what, exactly, a robust first-person perspective is.

3. The Robust First-Person Perspective

Unlike the rudimentary stage, which does not require language or concepts, the robust stage of the first-person perspective is a conceptual stage that entails the peculiar ability to conceive of oneself as oneself in the first-person. Conclusive evidence of a robust first-person perspective comes from use of complex first-person sentences like e.g., “I wonder how I will die,” or “I promise that I will stay with you”[3]. If I wonder how I will die, or I promise that I will stay with you, then I am thinking of myself as myself; I am not thinking of myself in any third-person way (e.g., not as Lynne Baker, nor as that woman, nor as the only person standing in the room) at all. Even if I had amnesia and did not realize that I was Lynne Baker, I could still wonder how I am going to die. Any entity that can wonder how she – she herself – will die *ipso facto* has a robust first-person perspective and thus is a person. She can understand herself from “within”, so to speak.

In order to have a robust first-person perspective, one must have a concept of oneself as oneself from the first-person – a self-concept. The second occurrence of ‘I’ in “I wonder how I am going to die” expresses a self-concept. It is impossible that two people have the same self-concept (cf. Kripke, 2011, p. 298). A self-concept cannot stand alone; it is a nonqualitative concept that is used only in tandem with other concepts[4]. If I promise that I will take care of you, then I manifest a robust first-person perspective by expressing a self-concept; but also I manifest mastery of empirical concepts like “promise” and “taking care”. And it is in learning a natural language that one masters these other common empirical concepts that one joins to a self-concept. (Hume was right that when I look inside

[1] Dennett, D.C. (1995), *Darwin's Dangerous idea*, Simon and Schuster, New York, p. 336. Dennett is discussing Jared Diamond's *The Third Chimpanzee*.

[2] I found the handy distinction between remote and in-hand capacity in (Pasnau, 2002, p. 115).

[3] Hector-Neri Castañeda developed this idea in several papers. See “He: A Study in the Logic of Self-Consciousness”, *Ratio*, 8 (1966), pp. 130-157, and “Indicators and Quasi-Indicators”, *American Philosophical Quarterly*, 4 (1967), pp. 85-100.

[4] I think that this point suggests that Hume's famous passage, “When I enter most intimately into what I call myself, I always stumble upon some particular impression” (*Treatise*, Book I, Part IV, Section VI, 253) does not imply that a self-concept has no extension, only that a self-concept can be deployed only with other concepts.

myself, I always stumble over an impression or – as I would say – a thought; but the moral to draw is that a self-concept cannot stand alone, but is always deployed jointly with other concepts [Hume, 1738/1968]).

On my view, the robust first-person perspective is much more far-reaching than thinking about one's mental states, or about oneself as the bearer of mental states. Applying for a job, making a contract, accepting an invitation all require a robust first-person perspective. If I wish that I were a movie star, I manifest a robust first-person perspective; but if I wish that LB were a movie star, I do not manifest one – even though I am LB. There is an important ineliminable contrast between my thinking about myself as myself in the first-person, knowing that it is myself whom I am thinking about, and my thinking about someone who is in fact myself without realizing it (Baker, 2013). And this contrast cannot be made without a robust first-person perspective.

To sum up my idea of the first-person perspective: Whereas a *rudimentary* first-person perspective is shared by persons and certain nonhuman animals, a *robust* first-person perspective – the conceptual ability to think of oneself as oneself in the first-person – is unique to persons. Human persons normally traverse a path from the rudimentary to the robust first-person perspective, from consciousness to self-consciousness.

4. Descartes and The First-Person Perspective

Now I want to explore some ways in which my conception of the first-person perspective differs from Descartes's own conception.

1. Descartes's allows for thinkers in isolation. Mine does not. Descartes envisioned the possibility that there existed a single person, with a sophisticated ability to entertain thoughts and reason from them. For example, Descartes said (something like), "I seem to be sitting in front of the fire in my dressing gown, but my senses have deceived me before. So, perhaps they are deceiving me now". I do not want to challenge the validity of the argument or its premises, but rather insist that it is conceptually impossible for a solitary person to have such thoughts. If Descartes had been the only finite entity in the universe, he could not have entertained such thoughts. Why not? Because he could not have acquired the concepts that are the constituents of such thoughts – e.g., *fire* and *dressing gown* – if he did not have a public language, and he could not have had a public language without a linguistic community.

Mastering a language requires a linguistic community. Wittgenstein pegged why one could not make up a language in isolation: If you did, then there would be no standards of correctness. If you categorized a new item that you took to be of a kind that you "named", there would be no difference between getting it right and getting it wrong. Wittgenstein avers: "One would like to say: whatever is going to seem right to me is right. And that only means that here we cannot talk about *right*" (Wittgenstein, 1958, par. 258). So, whatever one did in isolation, it would not be to invent a language (see Baker, 2007a; Baker, 2007b). So here is the first difference between my view and Descartes's: On my view, there can be no thinkers without a linguistic community.

2. Relatedly, Descartes's thinkers are nonsocial entities. Mine are social entities. On Descartes's view, there could be isolated thinkers; according to mine, we are essentially social entities. As I argued earlier, robust first-person perspectives are what distinguish persons from everything else in the universe: Although not every person must have a robust first-person perspective, every person must have at least a remote or second-order capacity for one: There could be no persons if there were no robust first-person perspectives. And since a robust first-person perspective requires having a language, and having a language requires that one has a linguistic community, and a linguistic community is a social community, it follows that persons are social entities – on my view, but not on Descartes's.

3. Descartes appeals to a pure mind. I appeal to nothing but whole embodied persons – to persons and the bodies that constitute them. Descartes thought that there was a pure mind, perhaps a center of an arena of consciousness; whereas on my view, there is no subpersonal mind, soul or self. All my view requires are whole persons constituted by bodies. Persons – whole, embodied persons – are the bearers of properties like anger, regret, belief, knowledge, seeing a parking place, feeling excited, and other so-called "mental" properties. Brains furnish the mechanisms that make the exemplification of these mental properties possible.

In my opinion, any appeal to a mind, soul, or self is just gratuitous. (Saul Kripke recounted a conservation he once had with a nonphilosophical friend about Hume's misbegotten search for a self. His friend said, "Well, Hume must never have looked in a mirror" [Kripke, 2011, p. 308]. In a way, I agree with the friend: What you see in the mirror is as close as you will get to a self).

4. In the *Meditations*, Descartes's aim was epistemological (What can I know with certainty?). His tool was his Method of Doubt: Suspend judgment about any of your beliefs that could possibly be false, until you get to beliefs (if any) that cannot possibly be doubted. Thus, Descartes is not only a foundationalist, but the foundation of knowledge is robustly first-personal. (Each of us is to inspect our own beliefs). For Descartes, what is discerned from the first-person perspective has epistemic primacy. One knows one's own mind better than she knows anything else, and justifies her beliefs about her environment by "inspecting" her own mind.

By contrast, my aim is ontological (What is ontologically required for reality to be as it is?). I do not believe that there is any single rigorous method for finding out what is genuinely real. I do not believe that the first-person always has epistemic primacy. I take it that an object of kind K belongs in ontology: If (1) objects of kind K are not reducible to objects of lower kinds, and if (2) elimination of objects of kind K renders the ontology incomplete.

Let me explain: Artifacts – like tables and chairs, bicycles and automobiles – are neither reducible nor eliminable from a complete description of reality. For example, a tractor (an artifact) – cannot be reduced to objects of lower kinds, because something is a tractor only in virtue of there being certain practices, purposes and uses of the thing. (Some tractor-like object that spontaneously coalesced in outer space would not be a tractor, however much it resembled one). And the relevant facts about practices, purposes and uses are not determined by *any* facts about objects of lower kinds than tractors (steering wheel, tires, etc.). So, artifacts are not reducible. Nor are artifacts eliminable. If the ontology left out artifacts, it would not be a complete description of reality.

My method, such as it is, for determining whether something is genuinely real and belongs in ontology is to determine whether it is irreducible and ineliminable. This "method", unlike Descartes's, is highly fallible. On my view, the first-person confers no epistemic justificatory primacy. We need not justify our beliefs about the ways that things are in terms of the ways that they seem or appear to us.

5. Descartes sought a level of reality that was wholly without presuppositions. I do not. On my view there is always a plethora of presuppositions, many of which are clearly empirical.

6. Descartes was a dualist – there are two kinds of finite things, immaterial thinking substances and material extended substances, minds and bodies. On my view there are countless kinds of finite things ("primary kinds"), from tomatoes to diplomas.

7. A related difference – at least between my views and those of some of Descartes's descendants – is that on my view, many of the primary kinds of things are "intention-dependent". That is, they could not exist or occur in worlds in which there were no beings with intentions. These include all sorts of manufactured goods like bedclothes, doorknobs, and eyeglasses. Social objects like passports and credit cards exist even though their existence depends on our intentions and practices. From this fact, it follows that the would-be distinction between things that are mind-independent and things that are mind-dependent is not fundamental; you can draw such a distinction where you would like, with artifacts on the mind-dependent or mind-independent side, but the distinction has no ontological significance. Dollar bills are as real as rocks.

8. Descartes draws a distinction between "inner" and "outer", according to which each thinker has

infallible access to an inner world – the world of experiences – known directly by "inspection"; the outer world is the world of physical objects, known indirectly only by inference. On my view, this distinction is misdrawn: There is no "inner" transparent realm to which I have infallible access. (Most of us are often mistaken about our own motivations). To say that we have inner lives on my view is just to say that we engage in silent speech.

This comparison of my views and Descartes's (mostly from the *Meditations*) yields two dissimilar pictures of reality. The only thing that the pictures seem to have in common is that they both countenance a non-objective aspect to natural reality: For Descartes, it is the mind or soul; for me, it is the first-person perspective.

5. An Alternative Interpretation of Descartes?

On the basis of Descartes's *Meditations*, it seems that Descartes holds that reality is not wholly objective. It seems obvious, does not it, that whereas material substances (e.g., physical objects) are objective, thinking substances (e.g., minds) are not? But maybe there is another interpretation of Descartes, one that would leave his ontology wholly objective. Does Descartes really take reality to include non-objective finite immaterial substances?

Although I take the interpretation that I have given of Descartes to be the standard interpretation, perhaps Descartes's first-person talk in the *Meditations* is just a ladder that can be kicked away after we climb up it.

Consider Bernard Williams' suggestion in his book *Descartes: The Project of Pure Enquiry*. Williams imagines that Descartes is a Pure Enquirer, a truth-gatherer, whose only desire is to maximize the truth-*ratio* of his beliefs. Descartes's "indefinitely well-informed and resourceful opponent" (Williams, 1978, p. 57), whose aim is to thwart Descartes in his pursuit of truth, is the fictitious Evil Demon. The Evil Demon gives rise to the "hyperbolical" doubt that there might not be a physical world at all – hyperbolical because it calls into question not just whether I am now dreaming, whether my present perception is veridical, but whether *any* perception is veridical. On Williams' view of Descartes, if we can get past this hyperbolical doubt, then we can come to "know truths about the world, and our conceptions of the world will not be systematically distorted or in error" (Williams, 1978, p. 61). This is so, because Descartes takes it to be self-evident that if any of my perceptions are veridical, then they are caused by things outside of me that the perceptions are perceptions *of* (Williams, 1978, p. 58). Once Descartes gets the certainty of his own existence and of the existence of a nondeceiving God (in Med. III and IV), he can count on the truth of his perceptions, his clear and distinct ideas. The aim of the project of pure enquiry, Williams suggests, is knowledge of the world, "knowledge of a reality which exists independently ... of any thought or experience. Knowledge of what is there *anyway*" (Williams, 1978, p. 64).

Each of us has experiences of the world and ways of conceptualizing it, which give rise to beliefs. Williams calls all this together a person's representation of the world (Williams, 1978, p. 64). Suppose that two people, A and B, have different representations of the world. In order to understand how A's and B's can be representations of the same reality, we must stand back and form a larger conception of the world that contains A and B and their representations. Then we add person C, and stand back again to include C and her representations with A and B and *their* representations. Suppose that we continue this process until we arrive at a conception that contains all the people in the world and all their representations of the world. Call this conception the "absolute conception".

If we cannot form such an absolute conception, then, says Williams, we have no conception of "the reality which is there *anyway*", no conception of any object of which we have knowledge. (According to Williams, Descartes was concerned with knowledge that physics uncovers). So, if "knowledge is possible at all, it now seems, the absolute conception must be possible too" (Williams, 1978, p.65)[5].

But notice: The absolute conception has no place for anything whatever that is irreducibly first-personal:

[5] Now Williams formulates a dilemma: On the one hand, the absolute conception may be specified only as "whatever it is that these representations represent". In that case, the independent reality "slips out of the picture, leaving us only with a variety of possible representations to be measured against each other, with nothing to mediate between them". On the other hand, we may have some determinate picture of "what the world is like independent of any knowledge or representation in thought. In that case, it seems that we are left only with only one particular representation of the world, "our own, and that we have no independent point of leverage for raising this into the absolute representation of reality" (Williams, 1978, p. 65).

Each person and his or her conception of the world is represented, but not from any first-personal point of view. The absolute conception is wholly objective. There is no place for a first-person perspective or for any first-person phenomena in the absolute conception.

This raises the question: What happens to the soul in the absolute conception? If souls are omitted from the absolute conception, but Descartes is committed to them, then the absolute conception is metaphysically incomplete. Well, Descartes may even agree. The motivation for the absolute conception is to map out a domain for knowledge that is produced by physics, but Descartes may think that there is no such knowledge of souls. (Williams suggests something like this at the end of his book. [Williams, 1978, pp. 299-302]). Moreover, Williams says that Descartes's interest "is as much, in fact, more in science as it was in metaphysics" (Williams, 1978, p. 276). In this case, if the soul is not knowable scientifically, then there is no loss in leaving it out of the absolute conception.

Recall: The point of the absolute conception is to have a conception of reality that is wholly independent of us (Williams, 1978, pp. 64-66). If a soul is private to each person, the absolute conception cannot be independent of us if it contains souls. So, let us leave souls out of the absolute conception, and return to Descartes's search for truth.

Perhaps Descartes did not think that there were any truths about the soul since physics does not deliver any knowledge of the soul. It seems to me incoherent to say that souls exist, truths exist, but there are no truths about the soul. (Would not the sentence "There are souls" express a truth?). To say that there are souls, but no truths about souls is tantamount to simply stipulating that there is no truth but physical truth.

In that case, we could interpret Descartes's use of the first-person as being only a stylistic choice, as Williams suggests. It is "a delicate question", Williams says, "at what point the first-personal bias, in any methodologically significant way, takes hold of Descartes's enquiry" (Williams, 1978, p. 68). The questions Descartes wants answered may just as well be of the form "What is true?" or even "What is known?" (*ibid.*) rather than "What can I know?"

Maybe so, but Descartes's method, the project of pure enquiry, still has a first-personal structure. As Williams says, Descartes's method "requires reflection, not just on the world, but on one's experience" (Williams, 1978, p. 69). So, even if Descartes's goal is objective, his method remains first-personal.

If what is presupposed by the possibility of knowledge is the absolute conception, why does it matter how the absolute conception came to be formulated? The absolute conception itself has no tie at all to the first-person: It is totally objective. Here is a mundane analogy: You walk to the store to buy some milk; if the aim is to obtain milk, what difference does it make whether you walk, ride a bicycle, or take a taxi to the store? It is the milk that counts. Similarly, if there is a wholly objective absolute conception, what difference does it make whether the method used to formulate it is not objective? If what counts is only the absolute conception, then Descartes's picture of the world, surprisingly, is itself wholly objective.

Perhaps Descartes's position was like that of the chemist Kekulé, who discovered the molecular structure of the benzene molecule (a hexagonal ring) while dozing in front of his fireplace in 1865 (Hempel, 1966, p. 16). The point here is that how Kekulé came up with the idea of a hexagonal ring is irrelevant to whether it is correct. Similarly, if Williams is right, maybe Descartes's method of hyperbolical doubt is irrelevant to how the absolute conception should be regarded[6].

Speaking now for myself, I do not think that Williams' interpretation of Descartes can succeed, for the reason that I do not think that the "absolute conception" can be a complete description of reality,

[6] Or perhaps Descartes is in a position like John Perry's. Perry says that self-knowledge (in the way that requires what I call the "robust first-person perspective") is just a way of believing things about myself, LB; it is knowledge about myself that I picked up in a typically self-informative way. Knowledge about myself that I would say does not require a robust first-person perspective is knowledge about myself that I picked up in some other (e.g., third-personal) way. Although Perry argues that indexicals like 'I' are essential for explaining action, he thinks that "all facts are objective". He says that he is "not very clear about what would make a fact not objective" (Perry, 2002, p. 239). Well, I think that I know: A fact is not objective if the fact's obtaining entails that someone has a robust first-person perspective. If Descartes thinks that the absolute conception is a complete representation of natural reality, then he, like Perry, should try to explain first-personal phenomena in a way that eliminates them from the ontology.

however anybody arrives at it. This is so, because among the representations to be included in the absolute conception will be representations whose existence entails exemplifications of a robust first-person perspective. For example, suppose that I believe that I am going to die young. (OK, too late for that). This thought would appear in the absolute conception as "LB believes that LB is going to die young". But that is not accurate; my belief is about *my* death.

To be accurate, the absolute conception would have to represent my belief as "LB believes that she (she herself) is going to die young". But to represent my thought in that accurate way would render the absolute conception not wholly objective, because "LB believes that she (herself) is going to die young" entails that the robust first-person perspective is exemplified. If there were no robust first-person perspective, there could not be such a thought.

So, even disregarding Descartes's first-personal method of arriving at the absolute conception, on my view, the absolute conception could not be a complete ontology. In order to be a conception of "what is there anyway", independently of any thought or experience, the absolute conception must leave out the dispositional property that is the first-person perspective in its robust stage, and hence, on my view, must be incomplete. So, I am even further from Descartes if you ally him to the absolute conception than I am on the standard interpretation.

6. Conclusion

In conclusion, even if the standard interpretation of Descartes is correct, and he agrees that there is a non-objective aspect of natural reality, his picture of reality is quite different from mine. Let me review some dissimilarities between a Cartesian approach and my approach.

There is first-person epistemic primacy for the Cartesian, but none for me; language is individual for the Cartesian, but language is social for me; thinkers are solitary beings for the Cartesian, but thinkers are social beings for me; pure minds exist for the Cartesian, but no subpersonal minds, souls or selves exist for me; the Cartesian endorses substance dualism, but I endorse an indefinitely broad pluralism; the Cartesian aims at a presuppositionless foundation, but I do not; the Cartesian takes a distinction between "what is there *anyway*" and what depends on us (a mind-independent/mind-dependent distinction) to be fundamental, I do not; there is an immaterial "inner" realm for the Cartesian, but no such immaterial "inner" realm for me; there is an infallible method of inquiry for the Cartesian, but no infallible method of inquiry for me. To me, that is quite a significant list of differences.

In short, while I affirm a robust first-person perspective – a capacity that sets mature persons apart from everything else in the world – my view is far from being Cartesian.

REFERENCES

Baker, L.R. (2007a), "First-Person Externalism", *The Modern Schoolman*, 84, pp. 155-170;

Baker, L.R. (2007b), "Social Externalism and the First-Person Perspective", *Erkenntnis*, 67, pp. 287-300;

Baker, L.R. (2013), "The First-Person Perspective and Its Relation to Natural Science", in M. Haug (ed.), *Philosophical Methodology: The Armchair or the Laboratory?*, Routledge, Oxford, pp. 318-334;

Castañeda H.-N. (1966), "He: A Study in the Logic of Self-Consciousness", *Ratio*, 8, pp. 130-157;

Castañeda H.-N. (1967), "Indicators and Quasi-Indicators", *American Philosophical Quarterly*, 4, pp. 85-100;

Dennett, D.C. (1995), *Darwin's Dangerous idea*, Simon and Schuster, New York;

Eldredge, N. (2000), *The Triumph of Evolution*, W.H. Freeman, New York;

Hempel, C. (1966), *Philosophy of Natural Science* (Foundations of Philosophy Series), Prentice-Hall, Englewood Cliffs, NJ;

Hume, D. (1738/1968), *A Treatise of Human Nature*, Book I, Part IV, Section VI, Clarendon, Oxford;

Kripke, S. (2011), "The First Person", in *Philosophical Troubles: Collected Papers*, Vol. 1, Oxford University Press, Oxford, pp. 292-321;

Pasnau, R. (2002), *Thomas Aquinas on Human Nature*, Cambridge University Press, Cambridge;

Perry, J. (2002), "The Sense of Identity", in *Identity, Personal Identity, and the Self*, Hackett Publishing Company, Inc., Indianapolis, IN, pp. 214-243;

Williams, B. (1978), *Descartes: The Project of Pure Inquiry*, Humanities Press, Atlantic Highlands, NJ;

Wittgenstein, L. (1958), *Philosophical Investigations*, Third Edition, The Macmillan Company, New York.

DERMOT MORAN
University College Dublin and Murdoch University
dermot.moran@ucd.ie

DEFENDING THE TRANSCENDENTAL ATTITUDE: HUSSERL'S CONCEPT OF THE PERSON AND THE CHALLENGES OF NATURALISM

abstract

The person is a concept that emerged in Western philosophy after the ancient Greeks. It has a multiple origination in Alexandrine grammar (first, second, third person), Roman Law (free person versus slave) and Latin Christian Trinitarian theology, epitomized by Boethius' definition – a person is an individual substance of a rational nature. In this paper I trace some aspects of the history of the concept of person and evaluate contemporary analytic approaches in the light of the Husserlian phenomenological account of the person.

keywords

Phenomenology, person, naturalism, Husserl, first-person, embodiment, Lynne Rudder Baker

The concepts of "person" and "personhood" have re-emerged as a central concern of contemporary philosophy of mind and action (Baker 2000, 2013). Persons *matter*. Their lives have significance for themselves and for others. There is broad agreement that personhood and agency are crucial for human social, moral and cultural life (Sturma 1997). Persons are intrinsically valuable and deserving of dignity and respect (Korsgaard 2009). The concept of the *person* is at the heart of morality and human rights; it is wrong to violate persons (e.g. by inhuman and degrading treatment). The person is fundamental to morality, law (human rights), the health and human sciences, and indeed to everyday life, yet it lacks theoretical definiteness. Charles Taylor in *Sources of the Self* calls the person "part of our moral ontology" (Taylor 1989). Daniel Dennett (1981) similarly recognizes the person as "an ineliminable part of our conceptual scheme", albeit he interprets persons as "roles" or functions and denies that they exist as real ontological entities.

Many questions arise about persons: what kinds of entity are they? Who or what are persons? What are the boundaries of personhood in human beings, e.g. embryo stage, implantation, capacity for awareness, sensitivity to pain (Becker 2000; Jones 2004)? Can personhood be diminished or lost, e.g. in patients in a coma or in advanced dementia? Peter Singer (2002), for instance, proposes removing personhood from certain human beings in persistent vegetative states, advanced Alzheimer's, or other forms of dementia (Kitwood 1997). Are there non-human persons (see White 2007; Francione 2008)? Dolphins? Great apes? Intelligent machines or genetically altered human beings? Robots? There are even personhood deniers. Others from a different standpoint reject humanism and propound a "posthuman" or "transhuman" condition that transgresses traditional boundaries of the human due to new bio-technologies (Bostrom 2003). The health sciences (person-centered medicine, nursing, personalistic psychiatry, geriatrics, end-of-life care) recognize the importance of persons (Thomasma, Weisstub & Hervé 2001; Kitwood 1997), but with little theoretical underpinning. Psychology examines "personality" rather than persons. Religion, theology, and humanistic psychology (Rogers 1961) advocate the value and integrity of persons but such traditional defenses are regularly challenged by those who do not share the underlying value system or its justification (Singer 2002).

The first point to note is that "person" is a specifically Western concept, although there are analogous conceptions of the unique worth of the human being in other cultures (e.g. the concept of *jen* or *ren* in Chinese Confucianism).

As the anthropologist Clifford Geertz writes:

> the Western conception of the person as a bounded, unique, more or less integrated motivational universe, a dynamic centre of awareness, emotion, judgment, and action organised into a distinctive whole and set contrastively against other such wholes and against its social and natural background, is [...] a rather peculiar idea within the context of the world's cultures (Geertz 1974, p. 126).

Confucianism employs the key concept of 仁, *Jen* or *Ren* ("benevolence" or "humaneness"). The Chinese character combines "human being" (人) and the number "two" (二) and carries in folk etymology the thought of humans involved with one another or caring for one another (see Chan 1955; Shen 2003) in mutually supporting roles (mother-daughter, father-son, husband-wife). Buddhism, on the other hand, with its doctrine of no-self, has often been seen to be hostile to the concept of personhood although it too can be seen as promoting a humanism which is informed by compassion (Tu & Ikeda 2011). But the debate with the East can begin only after the Western notion of the person has been clarified.

The concept of the person has a long history in the West – from ancient Alexandrine grammar, to Christian Trinitarian theology, to Enlightenment discussions. Unusually the concept of the person is one of the few still current philosophical concepts that did not find its first expression in ancient Greek philosophy (deVogel 1963; Sorabji 2006). The term "person" in Greek (πρόσωπον), in Latin (*persona*), means originally "face", "visage", and refers to masks worn by theatre actors expressing character. Clement of Alexandria complained of women who turn their "faces" (*prosopa*) into "masks" (*prosopeia*).

In fact, the first Western discussions of persons emerge in Alexandrine grammar (e.g. first, second, third "person") and in Roman Law which distinguishes *persons* "in their own right" as freemen (*liberus*) from *slave* (*servus*, "under the right of another"), see Long (1912). Roman law had a gradated series of conceptions of the person. The person with the fullest autonomy and authority over others, held the right to own and dispose of property, was the "head" (*capus*) of a household. All others had degrees of legal dependency.

Latin Christian theology in the fourth century CE and subsequently made a profound advance by attaching personhood to God and individuating three "persons" in the Trinity (see Kobusch 1997). The Roman philosopher Boethius' definition of a person as 'an individual substance of a rational nature' (*naturæ rationalis individua substantia*) in his *Contra Eutychen et Nestorium* emerges in this Christian theological context discussing the nature of the Trinity (Koterski 2004) and had enormous influence on Aquinas (Wallace 1995) and subsequent Christian thought (Braine 1992). Persons, on this account, are ontologically distinct rational individuals. Boethius' concept of the person depends on concepts such as *substantiality*, *rationality* and *individuality*. Aquinas discusses Boethius' definition in detail approvingly but with considerable transformation of meaning in his *Summa Theologiae* Part I Q. 29 Art. 1, where the person is understood as a *bearer* of rationality (see Braine 1992). Thomas defends the attribution of personhood to disembodied entities, e.g. God, angels. Indeed, medieval theology developed extremely subtle and sophisticated ways of talking about persons.

Persons have generally been understood in the Western tradition, then, as individual substances, as free agents, as rational animals, as worthy of infinite dignity and respect, and so on. Ancient accounts of personhood as found for instance in Panaitios of Rhodes (as reported in Cicero's *De Officiis* I §§30-32) tend to emphasize the rational character of the human person, free will, the unique individuality of persons and also their historical contingency. The problem is that the different sources of the concept of "person" suggest different underlying metaphysical conceptions and presuppositions.

In modernity, Descartes refines the concept to reflective *self-consciousness* (*cogito*). Enlightenment thinkers, including Locke and Kant, emphasized rationality, freewill and autonomy as the key characteristics of persons. Locke (1689), revived by Parfit (1984), proposed self-consciousness, memory and repeated ability to self-identify as necessary to the identity of the person. For Kant, all *rational* beings, not just embodied ones, are persons. Locke and Kant laid the groundwork for considering personhood as both a normative and a descriptive concept: to be a person is to be worthy of respect, but personhood also picks out individual, embodied beings in nature. Locke's definition is instructive because it encapsulates many of the concepts and indeed contradictions found in the current profile of the concept of person. Locke defined a person as a "thinking intelligent being [...] capable of a law, and happiness, and misery, [...] that has reason and reflection, and can consider itself as itself, the same thinking being in different times and places" (Locke, 1689, 2.27.9 and 2.27.26). Following Locke, Kant, in his *Critique of Practical Reason*, explains persons as: ": "nothing else than [...] freedom and independence from the mechanism of the whole of nature, regarded nevertheless as also a capacity of a being subject to special laws – namely pure practical laws given by his own reason, so that a person as belonging to the sensible world is subject to his own personality insofar as he also belongs to the intelligible world" (Kant 1787, p. 210). Note that both Locke and Kant identify this capacity to act not just in accordance with law but in recognition of the force of law on them. This conception re-emerges in recent discussions of normativity in Korsgaard, McDowell and others. For Kant, persons must be treated as ends in themselves because we must respect them as free and rational and not constrained by their embodiment in the world of nature. Thus Kant writes in his *Anthropology from a Pragmatic Point of View*: "The fact that the human being can have the 'I' in his representation raises him infinitely above all other living beings on earth. Because of this he is a *person,* and by virtue of the unity of consciousness through all changes that happen to him, one and the same person – i.e., through rank and dignity an entirely different being from *things*" (Kant 1798, p. 239).

Kant recognizes the bi-furcated nature of persons – as natural beings in the world and also as transcendental entities acting under their conception of the law. This bifurcation will continue in Edmund Husserl's conception of persons as being both in the world and for the world.

Current analytic philosophy includes diverse metaphysical accounts of persons as unique integral wholes, organisms, assemblages of objective temporal parts (Hudson 2001), even aggregates that are as loosely connected as heaps or swarms (Peter Van Inwagen's "mereological non-essentialism", Inwagen 1990), alternatively, metaphysical simples. Peter Van Inwagen writes:

> I suppose that such objects – Descartes, you, I – are material objects, in the sense that they are ultimately composed entirely of quarks and electrons. They are, moreover, a very special sort of material object. They are not brains or cerebral hemispheres. They are living animals; being human animals, they are things shaped roughly like statues of human beings. (When Descartes used the words '*moi*' and '*ego*' he was referring malgré lui to a living animal, a biological organism. When Hume looked within himself and failed to find himself, he was looking in the wrong place: like everyone else, he could see himself with his eyes open). It follows from this, and from well-known facts about animals, that it is possible for a material object to be composed of different elementary particles at different times. "Mereological essentialism" is therefore false (Inwagen 1990, p. 6).

One of the most influential recent movements is so called 'animalism' (Snowdon 2014; Olson 2007), that sees personhood as incidental to our essential animality or organic nature and identity to be constituted by bodily continuity. The *constitution* view (Baker 2000; 2013) defines persons in relation to the first-person point of view. According to the constitution view, human persons are constituted by human bodies without being identical to the bodies that constitute them.

Another contemporary approach that reformulates the traditional criterion of rationality presents human persons as possessing the power for second-order representations or *metarepresentation*, i.e. the capacity to *represent* their representations, e.g. to consider certain states as *having been* theirs ("I was in pain yesterday"). This latter example involves adopting a complex temporal stance towards one's cognitive states, something perhaps unavailable to creatures lacking language abilities. This view, often understood more generally as the capacity for *metarepresentation* (Sperber 2000), has been the subject of much critical discussion. The American philosopher Harry Frankfurt (1971) claims that human persons are capable not just of wants and desires but also of higher-order or *second-order* desires about their desires (I can desire to curb my desire for cigarettes). Frankfurt claims that the capacity to form higher-order desires is adequate to distinguish persons from non-persons. In some respects all the approaches listed are unsatisfactory because they do not take into account the complex ways in which persons live and engage with their lives and with other persons. The metarepresentation approach has real limitations in that it may also exclude certain infants and impaired reasoners who ought to be considered persons on other grounds. One can imagine a person being able to identify a reason as their own without being able to determine when they formed it or grasp it as something *having been held* by them for some time. Higher-order stances towards one's mental states is a powerful human (and arguably some mammals) ability but it needs to be considered in terms of the living of an intentional and affective life.

The *narrative* approach (Taylor 1989; Dennett 1990, Hutto 2007) sees the person as emerging in a story it weaves about itself. Charles Taylor writes in his *Sources of the Self*: "to ask what a person is, in abstraction from his or her self-interpretations, is to ask a fundamentally misguided question, one to which there couldn't in principle be an answer" (Taylor 1989, p. 34).

Elsewhere he defines a person as "a being who can be addressed and who can reply [...] a respondent" (Taylor 1985, p. 97).

The "no self" or "illusory self" view claims that selfhood (and personhood) are inventions or constructs of the brain (Metzinger 2009). There are competing ontological, instrumental and eliminative conceptions of persons, ranging from full realism about persons to a complete denial of their existence (Farah & Heberlein 2007). Many of these approaches seek to conform to naturalism. It is not until recently that other key features of human beings such as *feelings*, *emotions* (Goldie 2000; Prinz 2003) and the *bodily* sense of agency have been advanced as contributing to personhood, again often in a piecemeal manner and without a coherent map of how these capacities integrate in the full, concrete living person.

Lynne Rudder Baker's (Baker 2000, 2007, 2013) approach is much more promising because it recognizes persons as genuine ontological entities in their own right; the person is, in Aristotelian terminology, a "primary kind". Baker sees persons as uniquely defined by possessing essentially a first-person point of view. She writes: "what's unique about us are the features that make us persons, not just animals – features that depend on the first-person perspective (like wondering how one is going to die or evaluating one's own desires)" (Baker 2000).

And again: "What distinguishes person from other primary kinds (like planet or human organism) is that persons have first-person perspectives necessarily" (Baker 2007, p. 68).

Baker further clarifies what a first-person perspective is: "A first-person perspective is a very peculiar ability that all and only persons have. It is the ability to conceive of oneself as oneself, from the inside, as it were" (Baker 2007, p. 69). She goes on to say in her 2007 book *Metaphysics of Everyday Life*:

> A being may be conscious without having a first-person perspective. Nonhuman primates and other higher animals are conscious, and they have psychological states like believing, fearing, and desiring. They have points of view (e.g., "danger in that direction"), but they cannot conceive of themselves as the subjects of such thoughts. They cannot conceive of themselves from the first-person (Baker 2007, p. 70).

For Baker, the nonidentity of person and organism is based on the fact that organisms have different persistence conditions from persons. Human organisms have, Baker claims, third-personal persistence conditions: whether an animal continues to exist depends on continued biological functioning. Persons, on the other hand, have first-personal persistence conditions: whether a person continues to exist depends on its having a first-person perspective. Most recently, Baker has modified her view to distinguish between a "rudimentary" and a "robust" first-personal perspective. The rudimentary perspective is a metaphysical property possessed by human pre-linguistic babies and some animals. This rudimentary perspective includes the ability to perceive and act on environment from a particular spatiotemporal location, which, for Baker, necessarily requires consciousness and intentionality. The robust perspective, on the other hand, is a "remote" capacity that is acquired at birth but needs to be activated later. It seems to require the possession of language and the ability to refer to oneself as "I".

Baker's view is very rich and suggestive. I think she is essentially correct to recognize the ontological status of persons and their unique possession of a first-person perspective. She is also correct to acknowledge gradations or levels in the development of persons from the rudimentary to the robust stage. In my view her account is still too "third-personal" and perhaps too closely seeking to accommodate itself to naturalism. In the rest of this paper I am now going to sketch an alternative view – drawing on the rich resources of the phenomenological tradition and showing some comparisons with Lynne Rudder Baker's approach.

In contemporary European philosophy the phenomenological tradition (especially Husserl, Stein and Scheler) has much to say about persons, but this rich tradition has been relatively neglected until recently. The phenomenological tradition recognizes persons as embodied, intentional meaning-making historical beings, embedded in social contexts and acting on the basis of motivation rather than causation. The Husserlian phenomenologist Robert Sokolowski in his *Phenomenology of the Human Person* characterizes persons primarily as "agents of truth" and of disclosure (Sokolowski 2008).

I shall base my phenomenological account of personhood primarily on the writings of Edmund Husserl, but also, including insights drawn from some of the more neglected figures of the phenomenological movement, especially Max Scheler (1913-1916; 1973), and Edith Stein (1989; 2000). Martin Heidegger, in *Being and Time* (1927), deliberately rejects the Husserlian conceptions of consciousness and of the transcendental ego, as well as Scheler's "personalism", and instead introduces the notion of Dasein. There has been much controversy of the meaning of Dasein. Dasein picks out the transcendental conditions for the possibility of living in the disclosure of being. Is Dasein a person? Does it mean the way of existence of individual, embodied, historical human beings? Is it a kind of categorial picture of what human being essentially is? These are difficult questions. The later Heidegger becomes even more anti-humanist especially in his 1947 *Letter on Humanism* and the result includes Foucault's proclamation of the end of man – indeed of the birth of the conception of "man" in the classical age. Heidegger, on the other hand, claims the ancient tradition did not value human beings highly enough.

In fact, a very rich and still relatively unexplored phenomenological concept of personhood is developed by Husserl, especially in his *Ideas* II (Husserl 1952; 1989), unpublished during his life and which was assembled by his then assistant Edith Stein. This concept of the person is also taken up in Edith Stein's doctoral thesis *On the Problem of Empathy* (1917/1989) and in her subsequent important and neglected study *Contributions to the Philosophical Foundation of Psychology and the Human Sciences* published in Husserl's own *Jahrbuch für Philosophie und phänomenologische Forschung*, in 1922 and recently translated as 'Philosophy of Psychology and the Humanities' (Stein, 2000).

The phenomenology of personhood is closely wrapped up with the discussion of self-hood and this is something that one finds across throughout philosophical tradition. For Husserl, a person encounters itself in reflection as a self: "In reflection I therefore always find myself as a personal Ego. But originally this Ego is constituted in the genesis pervading the flux of lived experiences" (*Ideas* II § 58, Husserl 1989, p. 263; Husserl 1952, Hua IV 251).

Finally, despite the importance of Scheler who does invoke the concept of the person as a being oriented to value, neither Sartre nor Merleau-Ponty make much use of the concept of personhood, and it tended to fade out of phenomenological discussion, until relatively recently.

Phenomenology begins from the concept of functioning intentional life, of an embodied subject who is making sense of its world through intentional activity. Husserl writes in his *Crisis of European Sciences* (Husserl 1954/1970):

> Conscious life is through and through an intentionally accomplishing life [*intentional leistendes Leben*] through which the life world, with all its changing representational contents, in part attains anew and in part has already attained its meaning and validity. All real mundane objectivity is constituted accomplishment in this sense, including that of men and animals and thus also that of 'souls' (Husserl 1970, *Crisis* §58, p. 204; Hua VI 208).

Human beings, for Husserl, are essentially intentional meaning-makers. Moreover, despite his embrace of Cartesianism, Husserl was never solipsistic in his approach to human beings. They live in an intersubjective socially-constituted cultural life world. He writes: "The development of a person is determined by the influence of others" (Husserl, *Ideas* II § 58C, Husserl 1989, p. 281; Husserl 1952, Hua IV 268). Furthermore, persons simply do not appear in and therefore cannot be grasped by what Husserl calls the "naturalistic attitude" which sees things primarily as entities within nature as broadly understood within the natural, physical and biological sciences. Persons are recognized in what Husserl calls "the personalistic attitude" which, for Husserl, is prior to the natural attitude (and also to "the naturalistic attitude" which is even more derivative since it incorporates the outlook of modern science). It takes persons to recognize persons. Husserl writes: "[The personalistic attitude is] the attitude we are always in when we live with one another, talk to one another, shake hands with another in greeting, or are related to another in love and aversion, in disposition and action, in discourse and discussion" (Husserl 1989, *Ideas* II § 49, p. 192; Husserl 1952, Husserliana IV 183). Moreover, persons are in the world in a peculiar way in that they are also world-constituting and are being constituted in turn by their social and worldly relations (*Ineinandersein*). As the mature Husserl puts it, persons are "in the world" but also "for the world". According to Husserl's account of foundation, whereby there is a onesided dependence of one thing on another, persons are founded entities in that, in agreement with Baker, persons depend on corporeal living bodies but are not identical with their bodies (hence animalism is false). For Husserl, indeed, the conscious living self is *necessarily* embodied. This is an a priori, eidetic truth. Similarly, for him, consciousness is necessarily *egoic* (*ichlich*), that is ego-centered; all conscious acts and passions radiate from or stream into the ego or "I". An egoless consciousness is, for Husserl, also an a priori or eidetic impossibility. The pure I – the I of transcendental apperception – is, for Husserl, not a "dead pole of identity" (Hua IX 208), but rather is a living self, a stream that is constantly "appearing for itself" (*als Für-sich-selbst-erscheinens*, Hua VIII 189). It is sometimes described, in Hegelian language, as simply "for itself" (*für sich*). Husserl's terminology is wide-ranging. He speaks of "human-I" (*Ich-mensch*), "ego-body" (*Ichleib*), "I-pole" (*Ichpol*), "I-life" (*Ichleben*), "animate body" (*Leib*), "living body" (*Leibkörper, Körperleib,* depending on the emphasis), "pure ego", "phenomenological ego", "transcendental ego", "soul" (*Seele*), "psychic life" (*Seelenleben*), my "psychic" or "soulful" being (*mein seelishes Sein*, I 129), the "egoic" (*das Ichliche*), the "sphere of ownness" (*Eigenheitssphäre*), my "self-ownness" (*Selbsteigenheit*, I 125), the "primal I" (*Ur-Ich* VI 188) of the *epoché*, and so on. But frequently he employs traditional terms such as person, personal subject, life, subjectivity and so on, often endowing these terms with a new meaning. Husserl draws on all these locutions to try to articulate his sense of the meaning of subjective life in its first person, individual consciousness with its many layerings (including those that might properly be described as 'pre-ego' (*Vor-Ich*) and 'pre-personal'), as well as in its connection with other selves and in its moral, social and rational nature, amounting to its communalized 'life of spirit' (*Geistesleben*). In fact, subjectivity understood as "primordial, concrete subjectivity", "includes the forms of consciousness, in which is valid nature, spirit in every sense, human and animal spirit, objective spirit as culture, spiritual being understood as family, union, state, people, humanity" (XV 559, my translation).

For the mature Husserl, furthermore, the ego is an ego of habits. It also develops a personal style. The person emerges slowly and develops attributes which accrue to it as permanent characteristics that form its 'character'. Husserl writes:

> That which is given to us, as human subject, one with the human body [*Menschenleibe*], in immediate experiential apprehension, is the human person [*die menschliche Person*], who has his [or her] spiritual individuality, his [or her] intellectual and practical abilities and skills [*Fähigkeiten und Fertigkeiten*], his [or her] character, his [or her] sensibility. This Ego is certainly apprehended as dependent on its Body and thereby on the rest of physical nature, and likewise it is apprehended as dependent on its past (*Ideas* II § 34, Husserl 1989, p. 147; Husserl 1952, Hua IV 139-40).

Husserl's starting point is that, as persons, we can take positions and occupy standpoints and this is very close to Baker's view of the subjective first-person point of view. Husserl writes:

> As a point of departure we take the essential capacity of human beings for self-consciousness in the precise sense of personal self-reflection (*inspectio sui*) and the capacity grounded therein of reflectively taking positions vis-à-vis oneself and one's life, that is, the capacity for personal acts: of self-knowledge, self-evaluation, and of practical self-determination (self-willing and self-formation) (Husserl, Hua XXVII 23).

The mature Husserl was undoubtedly influenced by the Kantian and Neo-Kantian conceptions (he was a close reader of Natorp and Rickert) of the self as person understood as an autonomous ("giving the law to itself"), rational agent, but Husserl never suggests that the person is *purely* a rational subject. At the centre of the person, for Husserl, is a *drive* for reason, but it is a drive sitting upon many other affective and embodied elements, including drives, "strivings", passively being drawn to things, and so on. In its full "concretion" (Hua XIV 26), a *self* has convictions, values, an outlook, a history, a style, and so on. As Husserl writes in *Cartesian Meditations*: "The ego constitutes itself *for itself* in, so to speak, the unity of a history" (Husserl 1999, p. 75; Hua I 109). It is present in all conscious experience and cannot be struck out (*undurchsteichbar*). As the Husserl scholar Henning Peucker has written:

> The ego as a person is characterized by the variety of its lived experiences and the dynamic processes among them. According to Husserl, personal life includes many affective tendencies and instincts on its lowest level, but also, on a higher level, strivings, wishes, volitions, and body-consciousness. All of this stands in a dynamic process of arising and changing; lived-experiences with their meaningful correlates rise from the background of consciousness into the center of attention and sink back, yet they do not totally disappear, since they are kept as habitual acquisitions (*habituelle Erwerbe*). Thus, the person has an individual history in which previous accomplishments always influence the upcoming lived-experiences (Peucker, 2008, p. 319).

Given that Husserl sees persons as constituted in specifically personal acts and in intercommunication with other persons in mutual recognition, his approach to the person is resolutely anti-naturalist. Husserl rejects the naturalization of consciousness as one of the great counter-senses or contradictions of the age. He writes: "A univocal determination of spirit through merely natural dependencies is unthinkable, i.e. as reduction to something like physical nature [...]. Subjects cannot be dissolved into nature, for in that case what gives nature its sense would be missing" (Husserl 1989, *Ideas* II § 64, p. 311; Husserl 1952, Hua IV 297).

Generally speaking, Husserl regards naturalism as the reification of an outlook which is better understood as the natural attitude. He writes: "Naturalism is seduced by the spirit of unquestioning ('naïve') acceptance of the world that permeates the natural attitude, leading to the 'reification' (*Verdinglichung*) of the world, and its 'philosophical absolutizing' (*Verabsolutierung*)" (*Ideas* I, § 55, p. 129; Hua III/1, Husserl 1950, p. 107). Naturalism begins from the presumption of a given "ready-made world".

Transcendentalism, on the other hand, says: the ontic meaning (*der Seinssinn*) of the pregiven life-world is a subjective structure (*subjektives Gebilde*), it is the achievement (*Leistung*) of experiencing, prescientific life (Husserl 1970, *Crisis* § 14, p. 69; Husserl 1954, Hua VI , p. 70).

From *Ideas* I (1913) onwards, Husserl characterizes the ego as an 'I-pole' (*Ichpol*) or "I-centre" (*Ich-Zentrum*), "the centre of all affections and actions" (Hua IV 105). The I is a "centre" from which "radiations" (*Ausstrahlungen*) or "rays of regard" stream out or *towards* which rays of attention are directed. It is the centre of a "field of interests" (*Interessenfeld*), the "substrate of habitualities" (*Cartesian Meditations*, Hua I 103), "the substrate of the totality of capacities" (*Substrat der Allheit der Vermögen*, Hua XXXIV 200). This I "governs"; it is an "I holding sway" (*das waltende Ich*, Hua XIV 457) in conscious life (Hua IV 108), yet it is also "passively affected". In his *Kaizo* articles from the early 1920s, Husserl expands on the notion of personhood to speak of the character of social groups and peoples which he sees as 'personalities of a higher order' (*Personalitäten höherer Ordnung*, Hua XXVII 22) – made up of individual persons who are united into a culture and a communal consciousness. Communal achievements are not merely the aggregates of the achievements of individuals, Husserl points out. A people (*Menschheit*) can be understood as "an individual human writ large" (*Mensch im grossen*, Hua XXVII 22). There can be a common will or group properties that are not possessed by individuals.

Husserl gives an a priori account of personhood. The essential capacity for self-consciousness and what Husserl calls *inspectio sui* is important (Hua XXVII 23). But equally important is the idea of being able to "take positions" (*Stellungnehmen*) regarding their lives. This involves the capacity for uniquely personal acts, what Husserl often calls "I-Thou acts" (*Ich-Du-Akte*, Hua XXVII 22). Persons evaluate their actions, motives, goals, and values. The person is not just a rational agent but also built up on capacities, dispositions, skills, and what Husserl often refers to as *praxis.* Husserl also speaks of a *habitus* (Hua XXVII 23).

Husserl speaks of human person's ability to act freely from the "I-centre" outwards: thinking, evaluating, acting. Persons can curb their inclinations and what passively affects them. The subject is an "acting subject". A lot of this Husserl puts under the category of position-taking (*Stellungnehmen*). We can alter, take up or modify or negate position takings. We can affirm or reject previous decisions made freely. Husserl emphasizes that not only can we curb or alter position but we can reflectively renounce a position. It is important to emphasize that we can and do occupy positions pre-reflectively. We simply inhabit stances towards the world. This goes along with our personal habitus. But it also marks our individual "style".

In *Ideas* II, written roughly around the same time as *Ideas* I, Husserl begins from the experience of myself as embodied ego or *Ich-Leib*, as a special kind of physical entity in a physical world, or, to use Husserl's language of the 1920s, as a "world-child" (*Weltkind*, IX 216). The experienced body belongs to our "natural conception of the world". Husserl simply describes in phenomenological terms the manner of the givenness of the living body (*Leib*), which is first constituted in the stream of experiences (Hua XIII 5). The body is sensitive, reactive, responsive, but it also has freely willed movement, spontaneity, the basis for the autonomy that enables it to operate as a rational subject. There is a special kind of corporeality, embodiment or 'lived-bodiliness' (*Leiblichkeit*) belonging to the ego. It has its own kind of objectivity, its own peculiar mode of givenness. I am both a living organism (*Leib*) and a physical corporeal thing (*Körper*), an "external body" (*Aussenkörper*), a natural body, a spatio-temporal, material object (XIV 456). The body is unified with a psychic stratum (IV 25); it is a "psychophysical unity". The psychic or conscious stratum supervenes on the living body and is "interwoven" with it such that they penetrate (IV 94). The psychic, as Husserl understands it, is not an independent domain but one dependent on or "founded on" the physical (IV 310). This interpenetration of psychic with physical is personally experienced – I decide to raise my arm, my indigestion affects my mood, and so on. Husserl starts from my experience of myself and here there is a sense of "I" pervading the whole body, I *animate* my body from within, and physical body is only arrived at by abstracting from this animation (Hua IX 131). Moreover, peculiarly, I can experience myself both from the point of view of the purely physical (my body is subject to gravity, I fall down the stairs, or it can twitch under an electric shock, impulses that are other than self, *ichfremd*, XIV 89), or the psychic – I can move myself, I can leap out the window. The "body" in the sense of a Cartesian physical object is an abstraction that focuses on certain properties and ignores "practical predicates" (IV 25), rather I experience my own "innerness" (*Innerlichkeit*), my "inner flesh" (*Innenleib*), my alertness, relaxedness and so on. Only when we abstract from the essential "two-sidedness" of the animate body, do we experience the purely physical body (Hua IX 131). Husserl always stresses the bodily sensations and experiences that are "I-related", that are connected in some way with my will or in some way awake my interest.

Through my body I am an object in the world and also an actor in the world or as Husserl prefers to say "for" the world. My living body is the "organ of worldly life" (*Organ für das Weltleben*, XIV 456), and the world is the theatre where I display myself through my *Leib*. My body is primarily experienced as an instrument of my will, a "field of free will" (IV 310), it is the centre of a series of "I can"s, of my "being able to" (*Können*), of "powers" or "capacities" (*Vermögen*). I can move my eyes, head, limbs, alter my gaze, position, direction of attention. But not every bodily movement involves an explicit act or *fiat* of the will (XIV 447ff.). I may move my hand "involuntarily" because its position was uncomfortable (IV 260), I involuntarily reach for a cigar (IV 258). When I play the piano as an expert, I do not *wilfully* move my fingers but they do move voluntarily (XIV 89). They are doing what I want them to do, but I can still perhaps adjust my posture or press with greater pressure on the keys.

There is much more to be said about the complexity and variety of Husserl's thinking on the ego, the ego-body, the self and the person. But to clarify the manner in which Husserlian thinking developed I want now to turn briefly to two further phenomenologists – Max Scheler and Edith Stein. A person, for Scheler according to his *Formalism in Ethics and Non-Formal Ethics of Values*, is a "self-sufficient totality" (Scheler 1973, p. 390) and, moreover there is an individual world corresponding to each person (Scheler 1973, p. 393). For Scheler, a person is not part of the world but a *correlate of the world*. This essentially makes the concept of person a transcendental concept.

Edith Stein wrote her doctoral thesis under Edmund Husserl but was deeply influenced by Scheler's account of empathy, as well as by Hedwig Conrad Martius, Geiger, Pfänder and others. Stein completed her doctoral thesis with Husserl on the concept of empathy (Stein 1917/1989). Stein gives a very interesting characterization of "spiritual persons" in that work. Chapter Four of *On the Problem of Empathy* deals with what she calls "the spiritual subject" by which she means the human subject in so far as he or she is an agent attuned to values (a concept she found in Scheler). This attunement to values is, of course, a clear acknowledgement that the self and the person move in the space of reasons, meanings and values. The self and the person belong within the domain of normativity – but there is more in what Stein, following Husserl and Scheler, calls “spirit”.

As Stein puts it, “an “I” in whose acts an object world is constituted and which itself creates objects by reason of its will” (Stein 1989, p. 96). Spiritual acts are not simply separate rays streaming out from an ego but overlap, interpenetrate and build on one another. They are linked under the lawfulness of motivation. As she puts it, directly echoing Husserl in *Ideas* II: “motivation is the lawfulness of spiritual life” (Stein 1989, p. 96). Moreover, spiritual subjects operate within a general context of “intelligibility and meaningfulness”. A feeling, for example, may motivate a particular expression and define the range of expressions that can properly issue from it.

Stein is interested phenomenologically in the constitution of personhood. She emphasizes especially the role of feeling in the constitution of personality. Stein in general spends a lot more time on the feeling and emotive aspects of self-hood. There are different layers and dimensions to the self and different ways in which the ego is involved or at a distance from these feelings. The self is entirely permeated by emotions but even these can be at different *depth*. As Stein writes: “Anger over the loss of a piece of jewelry comes from a more superficial level or does not penetrate as deeply as losing the same object as the souvenir of a loved one. Furthermore, pain over the loss of this person would be even deeper” (Stein 1989, p. 101).

According to Stein, every feeling has a certain mood component “that causes the feeling to spread throughout the I from the feeling’s place of origin and fill it up” (Stein 1989, p. 104). A slight resentment can grow and consume me completely. There is not only “depth” and expanse (“width”), and “reach” in relation to emotions and feelings, but there is also duration. Emotions and feelings develop, evolve, change over time. Personhood can be “incomplete” – someone who has never experienced love, or who cannot appreciate art (Stein 1989, p. 111). Perhaps personality does not unfold and one becomes a “stulted” person. In this sense, there are aspects of the person that grow or can decline. There is a great richness of descriptive detail and psychological insight in the writings of Edith Stein on the nature of the person and it is very likely she influenced Husserl’s thinking on persons as much as she was influenced by him. Unfortunately, we cannot explore it further here.

In this paper, I have tried to show the depth and richness of the phenomenological approach to persons. According to the phenomenological approach – developing insights from Husserl, Scheler, and Stein – to be a person is minimally to be an *embodied intentional sense-maker*, involved in an intersubjective horizon of other persons, belonging to and acting and suffering in a world [*In-der-Welt-sein*], possessing a specific sense of a personal past and a future of possibilities that belong and in principle can be realized to it, living a *life* that is meaningful *for* that being, *for whom things matter* (but this “mattering” – i.e. the normative values need not be necessarily represented consciously). To be a person is always to be involved with other persons in a world. Furthermore, personhood is gradually acquired, grows and can be diminished. While Husserlian phenomenology acknowledges the paradigm case of the mature rational person who acts autonomously out of purely or mainly rational motives, phenomenology can also accommodate a much weaker notion of persons as beings who have a subjective point of view and live lives that have significance for themselves and for others. Personhood can decline although it is not easy to say if it can be completely lost. Phenomenology also recognises a “core” or “minimal self” (Strawson 2009), a consciousness of oneself as an immediate subject of experience, at the very heart of embodied human existence. This minimal self involves little more than a pre-reflective self-awareness that may be regarded as constitutive of consciousness as such. Part of phenomenology’s richness is that it can understand persons in a much wider context than that of autonomous rationality – there is the whole range of embodied selfhood, feeling, emotion and the apprehension and appreciation of value. Edith Stein offers a valuable insight also when she says that in the end persons always remain mysterious to one another.

REFERENCES

Baker, L.R. (2000), *Persons and Bodies. A Constitution View*, Cambridge University Press, Cambridge;

Baker, L.R. (2007), *The Metaphysics of Everyday Life: An Essay in Practical Realism*, Cambridge University Press, Cambridge;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, New York;

Becker, G.K. (ed.) (2000), *The Moral Status of Persons. Perspectives on Bioethics*, Rodopi, Amsterdam;

Bostrom, N. (2003), “Human Genetic Enhancements: A Transhumanist Perspective”, *Journal of Value Inquiry*, 37(4), pp. 493-506;

Braine, D. (1992), *The Human Person: Animal and Spirit*, University of Notre Dame Press, South Bend, IN;

Chan, Wing-tsit (1955), “The Evolution of the Confucian Concept of Jen”, *Philosophy East & West*, 4(1), pp. 295-319;

Dennett, D.C. (1981), “Conditions of Personhood”, in *Brainstorms*, Harvester Press, Brighton, pp. 267-285;

Dennett, D.C. (1990), *Consciousness Explained*, Little, Brown and Company, Boston;

de Vogel, C.J. (1963), “The Concept of Personality in Greek and Christian Thought”, in J.K. Ryan (ed.), *Studies in Philosophy and the History of Philosophy*, Vol. 2, Catholic University of America Press, Washington, pp. 20-60;

Farah, M.J. & Heberlein, A. (2007), “Personhood and Neuroscience: Naturalizing or Nihilating?”, *American Journal of Bioethics*, 7(1), pp. 37-48;

Francione, G.L. (2008), *Animals as Persons: Essays on the Abolition of Animal Exploitation*, Columbia University Press, New York;

Frankfurt, H. (1971), “Freedom of the Will and the Concept of the Person”, in *Journal of Philosophy*, 68(1), pp. 5-20;

Geertz, C. (1974), “From the Natives’ Point of View: On the Nature of Anthropological Understanding”, in R.A. Shweder & R.A. Levine (eds.), *Culture Theory*, Cambridge University Press, Cambridge, pp. 123-136;

Goldie, P. (2000), *The Emotions: A Philosophical Exploration*, Clarendon Press, Oxford;

Hudson, H. (2001), *A Materialist Metaphysics of the Human Person*, Cornell University Press, Ithaca, New York;

Husserl, E. (1913/1950), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Erstes Buch: Allgemeine Einführung in die reine Phänomenologie*, Walter Biemel (ed.), Martinus Nijhoff, Haag;

Husserl, E. (1950-), *Gesammelte Werke*, Husserliana, Springer, Dordrecht;

Husserl, E. (1952), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Zweites Buch: Phänomenologische Untersuchungen zur Konstitution*, Marly Biemel (ed.), Husserliana IV, Nijhoff, The Hague, reprinted Springer, Dordrecht;

Husserl, E. (1970), *The Crisis of European Sciences and Transcendental Phenomenology. An Introduction to Phenomenology*, D.C. Evanston (tr.), Northwestern University Press, IL;

Husserl, E. (1973a), *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Erster Teil. 1905–1920*, I. Kern (ed.), Husserliana Vol. XIII, Nijhoff, The Hague;

Husserl, E. (1973b), *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Zweiter Teil. 1921–1928*, I. Kern (ed.) Husserliana Vol. XIV, Nijhoff, The Hague;

Husserl, E. (1973c), *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Dritter Teil. 1929–1935*, I. Kern (ed.) Husserliana Vol. XIV, Nijhoff, The Hague;

Husserl, E. (1977), *Phenomenological Psychology: Lectures, Summer Semester, 1925*, J. Scanlon (ed.), Martinus Nijhoff, The Hague;

Husserl, E. (1983), *Ideas pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, First Book*, F. Kersten (tr.), Kluwer, Dordrecht;

Husserl, E. (1989), *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Second*

Book. Studies in the Phenomenology of Constitution, R. Rojcewicz & A. Schuwer (tr.), Kluwer Academic Publishers, Dordrecht;

Husserl, E. (1999), *Cartesian Meditations: An Introduction to Phenomenology*, D. Cairns (ed.), Martinus Nijhoff, The Hague;

Hutto, D. (2007), *Narrative and Understanding Persons*, Cambridge University Press, Cambridge;

Jones, D.G. (2004), "The Emergence of Persons", in M. Jeeves (ed.), *From Cells to Souls and Beyond*, Eerdmans, Grand Rapids, pp. 11-33;

Kant, I. (1787), *Kritik der praktische Vernunft*, in *Kant's Gesammelte Schriften*, Reimer, Berlin, 1900–, V, 87; English translation Kant, I., *Practical Philosophy*, M. Gregor (ed.), Cambridge University Press, Cambridge 1996;

Kant, I. (1798), *Anthropologie in pragmatische Hinsicht*, in *Kant's Gesammelte Schriften*, Reimer, Berlin, 1900–, VII, 127; English translation in G. Zöller & R.B. Louden (eds.), *Anthropology, History, and Education*, Cambridge University Press, Cambridge 2007;

Kitwood, T. (1997), *Dementia Reconsidered: The Person Comes First*, Open University Press, Buckingham;

Kobusch, T. (1997), *Die Entdeckung der Person. Metaphysik der Freiheit und modernes Menschenbild*, Wiss. Buchgesel, Darmstadt;

Korsgaard, Ch. (2009), *Self-Constitution: Agency, Identity, and Integrity*, Oxford University Press, Oxford;

Koterski, J. (2004), "Boethius and the Theological Origins of the Concept of Person", *American Catholic Philosophical Quarterly*, 78(2), pp. 203-224;

Locke, J. (1689), *An Essay concerning Human Understanding*, Clarendon Press, Oxford 1975.

Long, J. (1912), *Notes on Roman Law*, Michie, Charlotteville;

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

Metzinger, Th. (2009), *The Ego Tunnel: The Science of the Mind and Myth of the Self*, Basic Books, New York;

Olson, E.T. (2007), *What Are We? A Study in Personal Ontology*, Oxford University Press, Oxford;

Parfit, D. (1984), *Reasons and Persons*, Oxford University Press, Oxford;

Peucker, H. (2008), "From Logic to the Person: An Introduction to Husserlian Ethics", *Review of Metaphysics*, 62, pp. 307-325;

Prinz, J.J. (2004), *Gut Reactions: A Perceptual Theory of Emotion*, Oxford University Press, Oxford;

Rogers C. R. (1961), *On Becoming a Person: A Therapist's View of Psychotherapy*, Houghton Mifflin, Boston.

Scheler, M. (1913-1916/1954), *Der Formalismus in der Ethik und die materiale Wertethik. Neuer Versuch der Grundlegung eines ethischen Personalismus*, Gesammelte Werke, Band 2, Francke Verlag, Bern/München;

Scheler, M. (1973), *Formalism in Ethics and Non-Formal Ethics of Values. A New Attempt Toward a Foundation of An Ethical Personalism*, M.S. Frings & R.L. Funk. (eds.), Northwestern University Press, Evanston;

Shen, V. (2003). "Ren: Humanity", in *Encyclopedia of Chinese Philosophy*, A.S. Cua (ed.), Routledge, New York, pp. 643-646;

Singer, P. (2002), *Unsanctifying Human Life: Essays on Ethics*, H. Kuhse (ed.), Blackwell, Oxford;

Snowdon, P.F. (2014), *Persons, Animals and Ourselves*, Oxford University Press, New York;

Sokolowski, R. (2008), *Phenomenology of the Human Person*, Cambridge University Press, Cambridge;

Sorabji, R. (2006), *Self: Ancient and Modern Insights about Individuality, Life and Death*, Oxford University Press, Oxford;

Spaemann, R. (2006), *Persons: The Difference Between 'Someone' and 'Something'*, O. O'Donovan (tr.), Oxford University Press, Oxford;

Sperber, D. (ed.) (2000), *Metarepresentation*, Oxford University Press, New York;

Stein, E. (1917/1980), *Zum Problem der Einfühlung*, Verlagsgesellschaft Gerhard Kaffke, München;

Stein, E. (1989) *On the Problem of Empathy*, W. Stein (tr.), ICS Publications, Washington, DC;

Stein, E. (1922), "Beiträge zur philosophischen Begründung der Psychologie und der Geisteswissenschaften, Erste Abhandlung", *Jahrbuch für Philosophie und phänomenologische Forschung*, 5, pp. 1-116; English translation Stein, E., *Philosophy of Psychology and the Humanities*, ICS Publications 2000, Washington, DC;

Strawson, G. (2009), *Selves: An Essay in Revisionary Metaphysics*, Oxford University Press, Oxford;

Sturma, D. (1997); *Philosophie der Person. Die Selbstverhältnisse von Subjektiviät und Moralität*, Schöningh, Paderborn;

Taylor, Ch. (1985), "The Concept of a Person", in C. Taylor (ed.), *Human Language and Agency, Philosophical Papers I*, Cambridge University Press, New York;

Taylor, Ch. (1989), *Sources of the Self*, Cambridge University Press, Cambridge;

Thomasma, D.C., Weisstub D.N. & Hervé C. (eds.) (2001), *Personhood and Health Care*, Kluwer, Dordrecht;

Tu, W. & Ikeda, D. (2011), *New Horizons in Eastern Humanism: Buddhism, Confucianism and the Quest for Global Peace*, Tauris, New York;

van Inwagen, P. (1990), *Material Beings*, Cornell University Press, Ithaca, New York;

White, Th.I. (2007), *In Defense of Dolphins. The New Moral Frontier*, Blackwell, Oxford;

Wallace S. (1995), "Aquinas versus Locke and Descartes on the Human Person and End-of-Life Ethics", *International Philosophical Quarterly*, 35(3), pp. 319-330.

MICHAEL PAUEN
Berlin School of Mind and Brain and Humboldt University
michael.pauen@philosophie.hu-berlin.de

HOW NATURALISM CAN SAVE THE SELF

abstract

Skepticism regarding the self has been widespread among naturalist philosophers. Contrary to this view it is shown here that naturalism can provide a deeper understanding of the self. Starting with a phenomenology of the self it is argued that self-consciousness can be understood as an act of perspective taking. Thus, self-consciousness turns out to be a natural ability, which can be investigated empirically. These studies can further improve our understanding of the self.

keywords

Self, self-consciousness, perspective-taking

The problem of the self has been one of most important issues in occidental philosophy ever since it was discussed in Florentine Neoplatonism. However, many philosophers think that there is a basic incompatibility between naturalism on the one hand and a sufficiently strong idea of the self on the other. Here I would like to show that this is not the case. You can have the cake and eat it, too.

In order to show this, I will present a reductive account of the self in naturalist terms. Note that I will use the word "reduction" not in the ordinary sense of the word, according to which reduction means decrease. Rather, I am using it in the original Latin sense of *reducere* which is at issue when we talk about reducing an effect to its cause, say in order to understand how the effect came about. Obviously, such an explanation does not put the discussed phenomena at risk. Rather, it helps us understand them. Accordingly, I will try to reduce the higher level phenomenon of self-consciousness to psychological and neurobiological lower level phenomena – not because I am denying its existence but because I want to understand it.

In doing so, I will also discuss the role of language in the ontogeny of self-consciousness. I will refer to empirical evidence which seems to show that language might be not very important, particularly in the earliest stages of the development of self-consciousness. Also, I think that the development of self-consciousness in ontogeny tells us something about the elements which constitute the underlying abilities.

In using the word 'self', I do not mean some mysterious Cartesian entity outside the physical world. It is just shorthand for 'a person who has self-consciousness'. I will explain shortly in more detail what I mean by this.

The existence of the self has been put in doubt since long. Such doubts can already be found 1500 years before Christ in the Indian philosophy of the Vedas, as well – and much later – as in the German tradition, particularly in the philosophy of Schopenhauer and Nietzsche. More recently the idea that the self is just an illusion has been defended by Daniel Dennett: "The self turns out to be a valuable abstraction, a theorist's fiction rather than an internal observer or boss. If the self is 'just' the Center of Narrative Gravity, then, in principle, a suitably 'programmed' robot, with a silicon-based computer brain, would be conscious, would have a self (Dennett 1991, p.431).

A similar idea has been brought forward by Thomas Metzinger. In his view, there is no self, but only a self-model: “Metaphysically speaking no such things as selves exist in the world: the conscious experience of selfhood is brought about by the phenomenal transparency of the system model” (Metzinger 2003, p.627).

What Metzinger seems to have in mind is that we have a representation of ourselves, the so-called self-model, which we mistake for a real entity inside ourselves, that is, for something like the notorious homunculus. It seems that Dennett has a similar idea: what is really at issue when we talk about the so called self is the notorious homunculus. I do not think this is true. Obviously the idea of a homunculus is a very bad one, but I think that there is a very rational way of talking about the self that does not leave you with a homunculus. I would like to show what this alternative might look like. I will start with some classical views, particularly with some skeptical positions that I have already alluded to. Then I will refer to a discussion which was started by Fichte and which still has some followers in Germany particularly in the so-called “Heidelberg School” around Dieter Henrich (Henrich 1967; Henrich et al. 1966) and Manfred Frank (Frank 1991). Fichte pointed to a specific problem of any theory of self-consciousness which concerns recognizing yourself. To recognize an object such as a bottle of water you need certain criteria that enable you to distinguish the water bottle from other objects. The problem with self-recognition is that you already need self-consciousness if you want to find out which criteria can help to distinguish yourself from someone else: In order to determine whether a certain feature can serve as a criterion, you have to already know that this feature is one of *your* features. And this means that you need self-consciousness at this point already. Many philosophers have tried to solve this problem. I will present a solution proposed by the Heidelberg school, which I think goes in the right direction, but still leaves some questions unanswered. Finally, I will suggest an alternative idea.

1. Classical Views

First, however, I will talk about some classical views. Self-consciousness already played a role in pre-theoretical thinking of humans about themselves, for example in ancient religions and myths. Most typically the self was thought of as a kind of observable substance, say as a soul, which consisted of a particularly fine matter, or some kind of breath. The former idea is present in the platonic tradition, particularly in the *Timaeus*, the latter can be found in the Bible but it also persists in certain Greek and Latin terms like *flatus*, *spiritus*, *pneuma* or *psyche* that are used to refer to the self or the soul. Another idea was prevalent in ancient Egypt, where every person was thought to have various different souls each of which having a specific function. One of them represented a person’s individuality. This soul had the same look, size and shape as the persons themselves, and left the persons’ bodies when they died.

The ancient concept of a soul was much broader than most of the terms we use today. As a consequence, we need various different concepts in order to cover the entire meaning of the “soul”, among them “consciousness”, “perception”, “emotion”, but also the “identity”, “self”, and “self-consciousness”.

One of the most famous theoretical accounts of the self in classical philosophy has been called the *reflection theory* which was endorsed, among others, by Descartes. The basic idea here is that the self emerges from self-reflection. So in order to develop self-consciousness, you have to think about yourself. In this process you are simultaneously the subject and object of thought, and whatever thoughts you may have, they are immediately present to your introspection.

However – and this was Fichte’s idea – the problem is that this theory ends up in a vicious circle. Imagine that you want to recognize your fear as *your* fear in such a process of reflection. In order to do so, you already need some kind of self-consciousness regarding the feeling in question, otherwise you might end up ascribing your feelings to someone else or someone else’s feelings to yourself. The same holds when you try to recognize yourself in a mirror. In order to do so you already need some idea of what you look like – otherwise you might mistake someone else for you. But this shows that you already need self-consciousness. As a consequence, the self-reflection theory turns out to be false because it cannot explain what it is supposed to explain.

2. Skepticism

As I have already mentioned, there has be quite some skepticism about the self in the history of philosophy. One famous example is David Hume:

> For my part, when I look inward at what I call *myself*, I always stumble on some particular perception of heat or cold, light or shade, love or hatred, pain or pleasure, or the like. I never catch myself *without* a perception, and never observe anything *but* the perception. When I am without perceptions for a while, as in sound sleep, for that period I am not aware of myself and can truly be said not to exist. If all my perceptions were removed by death, and I could not think, feel, see, love or hate after my body had decayed, I would be entirely annihilated – I cannot see that anything more would be needed to turn me into nothing. If anyone seriously and thoughtfully claims to have a different notion of himself, I can’t reason with him any longer. I have to admit that he may be right about himself, as I am about myself (Hume 1978, p. 252; Book I, Part IV, Section VI).

Yet, I do not think that Hume really wants to deny the existence of self-consciousness. What he wants to deny is rather the existence of an immaterial substance as it had been postulated by Descartes. So what appears as an attack to the self turns out to be an attack to a specific theory of the soul. This is possible because – as already indicated – there has been no clear distinction between the self and the soul in huge parts of the philosophical tradition. Kant, for example, explicitly states that he uses the term “self” for the term “soul”[1].

Another strand of skepticism has been put forward by Marvin Minsky (1988). Minsky argues that there is not one self, but instead a multiplicity of agents within a person which cooperate with each other although none of these agents has more than a limited knowledge about what the others know. Today we might say that there are different systems within the brain or within the mind. Minsky claims:

> All this suggests that it can make sense to think there exists, inside your brain, a society of different minds. Like members of a family, the different minds can work together to help each other, each still having its own mental experiences that the others never know about. Several such agencies could have many agents in common, yet still have no more sense of each other’s interior activities than do people whose apartments share opposite sides of the same walls (Minsky 1988, p. 290).

What Minsky wants to deny is that there is one single agency, part, or system which exerts control and also has overall knowledge.

3. The Heidelberg School

These were only some famous examples for skepticism regarding the self. I have now mentioned the main points of some skeptical arguments about the self. Now I would like first to refer to one strategy of “saving” the self, which has been done in response to Fichte’s criticism of the reflection theory. Fichte argues that you cannot develop self-consciousness by reflecting on yourself (Fichte 1991, p. 11). The members of the so-called Heidelberg-School, Dieter Henrich and Manfred Frank, have developed an alternative idea. They think that the self is real and that we do have self-consciousness.

[1] “Wenn ich von der Seele rede; so rede ich von dem Ich in sensu stricto. So fern ich mich nun als einen Gegenstand fühle und dessen bewußt bin; so bedeutet dies das Ich in sensu stricto oder die Selbstheit nur allein, die Seele. Diesen Begriff der Seele würden wir nicht haben, wenn wir nicht von dem Object des inneren Sinnes alles Äußere abstrahiren könnten; mithin drückt das Ich in sensu stricto nicht den ganzen Menschen, sondern die Seele allein aus” (Kant 1821, p. 200).

However, they argue that the reflection model that was introduced above is misguided since it cannot explain the emergence of self-consciousness. Apart from this, it is also misleading to take the self as an internal object, since this would again invoke some kind of homunculus.

Like Fichte, Henrich and Frank claim that self-awareness cannot emerge from self-reflection, and the self cannot be a constellation of properties. The reason is simple: We would have to know for each of such properties that it is our own property. And this would require self-consciousness in the first place rather than explaining it. Let us assume that the feature that best identifies myself is the fact that I am the best aluminum welder in Berlin. So this feature would distinguish me from all other residents of Berlin. So why cannot I use this property in order to identify myself in acts of self-consciousness? The reason is that I already need self-consciousness in order to know that this feature belongs to me! So whatever the constellation of features or properties is, and however complex it may be, I cannot recognize that this constellation is my own or that it is the constellation that identifies myself prior to being self-conscious. The reflection theory therefore cannot explain the emergence of self-consciousness.

Henrich and his colleagues have concluded that because reflection does not explain the emergence of self-consciousness, this ability has to start with some sort of pre-reflexive self-awareness. So before you can begin to reflect on yourself you need some kind of direct, pre-reflexive access to yourself. Frank argues: "We cannot describe self-awareness as an awareness of something, if this 'something' represents a single object named 'self' (or 'I' or 'person'). Self-awareness is not object-like, its familiarity is not mediated by something else, its original instantiation is irreflexive, without criteria, and it's not based on observations, either" (Frank 1991, p. 5).

Frank makes it clear what self-consciousness is *not*, but he does not provide a positive account – which is an endeavor that the Heidelberg school never attempted. His idea of pre-reflexive self-awareness is not easy to grasp: If it is not object-like, irreflexive, without criteria and not based on observation, what, then, is it? We know that the reflection theory will not help us gaining an understanding of the self, but what *does* help is still an open question, even if we accept the account provided by Henrich and Frank.

4. An Alternative

As a consequence, the pre-reflexive self has to be made intelligible too. So we need additional explanations of what such a pre-reflexive self might look like, and this is what I would like to present in what follows. I will start with some minimal criteria for self-awareness, and then describe the phenomena involved. A reductive account of self-consciousness, after all, can only be successful if it captures the entire phenomenon.

So what do we mean when we talk about self-consciousness? What is the relevant phenomenon? It seems essential, first, that one is able to recognize one's body. This, in turn, means that one can distinguish it from things outside the body, including other persons' bodies. This is a basic ability which can be found in many animals, and it may be that even plants might have some rudiments of it. However, this ability is certainly not sufficient for self-consciousness. In addition, recognition has to be what I call *transparent*: you have to recognize your properties as *your* properties, your body as *your* body, and yourself as *yourself*. Many animals are able to perceive themselves in a mirror but most of them fail to understand that they are looking at themselves, so they do not perceive themselves *as* themselves. Something similar might happen when I talk about the best aluminum welder in Berlin without understanding that I am talking about myself. Given that I am, in fact, the best aluminum welder in Berlin, I am talking about myself. But I am not talking about myself *as* myself in the sense at issue here, as long as I do not understand that this is so.

So far we have only talked about individual acts of self-consciousness as they occur when I recognize myself in a mirror or fail to recognize my visual perception as my perception. But this is certainly not sufficient for a description of the phenomenon of self-consciousness, even if we focus only on the most essential features. We would not say that someone has self-consciousness in the full sense of the word unless he would have something like a persistent idea of himself, what he is, what his name is etc. I take it that this also requires autobiographical memories. So what is needed as a third criterion in order to capture the phenomenon of self-consciousness is something like a self-concept that is, a persistent representation of oneself. A person who sincerely claims to be Napoleon or George W. Bush would be considered to suffer from a severe self-disorder; the same would hold if someone told us yesterday that they are a very good aluminum welder but deny this very fact today. I assume that this is also where language comes in, even if language is not a necessary precondition for the first two criteria of self-consciousness.

The most important problem at stake here is transparency, which means recognizing yourself as yourself. This is also where the pre-reflexive self comes in according to the Heidelberg School. Philosophers have argued that this act of recognizing oneself as oneself leads to a paradox: On the one hand, there is no self-consciousness without being able to recognize oneself as oneself. Unfortunately, however, you need self-consciousness in order to recognize yourself as yourself.

In what follows, I would like to suggest a solution for this problem which also contributes to a deeper understanding of the emergence of self-consciousness. Though I think that the body and the "body scheme" play an important role for basic aspects of self-consciousness that have been described as the "core self" by Antonio Damasio (1994), I will focus on higher level cognitive phenomena like perspective-taking and theory of mind because I think that these abilities can give a very important contribution to our understanding of self-consciousness.

Let us first talk shortly about the body scheme. It implies the ability to recognize your body *as* your body, not on a personal, explicit level; rather, it appears as a direct feeling that your body is yours. For instance, you would react differently depending on whether someone threatens to injure your hand or some external object: Most likely, you would withdraw your hand only in the first case. This seems to show that we have a very deeply rooted, sub-personal access to our own body. Though I will not take a definite position here, it seems that this kind of body scheme is part of the immediate feeling of familiarity that we have with respect to ourselves. A body scheme also appears to be present in non-human beings, as it is non-cognitive, sub-personal and represented as a feeling.

Empirical support for this approach comes from the study of the disorders of the body scheme. Patients who suffer from somatoparaphrenia, a psychological disorder, really think that their own limbs are not parts of their body at all (Sacks, 1985). Merely telling the patients that this idea is irrational will not help them. It seems as if there is something wrong that is not affected by rational arguments. This is one reason why I think that the sub-rational body scheme may play a role in our immediate feeling of self-familiarity or self-acquaintance.

Additional evidence for these lower level aspects of self-consciousness comes from studies that show how the body scheme can be misled. Ramachandran (Ramachandran & Hirstein, 1997) treated a patient with phantom pain in his left hand, although it had been amputated. Ramachandran treatment consisted in showing the patient the mirror image of his right arm, thereby tricking the patient into believing that he was observing his left arm with a hand attached to it. Apparently the patient included the mirror image into his body scheme, thus allowing him to overcome his phantom pain.

Another example is the so-called "rubber hand illusion", which also shows that the body scheme depends on sub-personal stimuli, rather than being reactive to rational reasoning. In this paradigm, the experimental subjects have one of their hands covered so that they cannot see it. What they can see instead is a rubber hand which is visible exactly above their covered hand. The experimenter then touches the subjects' covered-hands and the rubber hands in exactly the same way; as a consequence, the subjects include the rubber hand into their body schemes and mistake the rubber hand for their own hand. For instance, if the experimenter threatens to injure the rubber hand with a hammer, the subject will retract his real hand. This demonstrates again that the body scheme is something subpersonal, and that it is at least a good candidate for constituting basic self-familiarity.

But there is a second possible candidate for an explanation of pre-reflexive self-consciousness namely perspective-taking, which takes a completely different route. It is cognitive, to some extent personal, and involves taking the perspective of a person with whom one is cooperating or communicating. There are certain requirements that have to be met in order to have this capacity. If perspective-taking is a good candidate for the development of self-consciousness, small children should be able to develop such abilities from a very early age on. In the following, I will give an account of the constitutive elements of perspective-taking, and show that children master these before they develop self-consciousness.

The first requirement in order to take someone else's perspective is to identify those beings whose perspective you can take. If you attempt to take the perspective of, say, a table or a camera, you will fail. However, the ability to distinguish between living beings and non-living beings is already present in infants at 2 to 3 months. Amanda Woodward (1998) shows that at 5 months, infants expect different behaviors from living and nonliving agents. In this study, a small child watched both a human and a mechanical arm reaching for an object. When the object was displaced the 5 months old child expected only the human arm but not the mechanical arm to account for this displacement and to reach for the object at the new location. Apparently, the children expected that the person had the intention to get the object, whereas they expected that the mechanical arm would just repeat its movement without any intention. This finding has been replicated several times and it shows that even at this early age, children can selectively attribute intentions to human agents. This capacity does not include perspective taking but it looks like a good starting point for developing this ability.

Between 7 and 9 months, infants are able to distinguish between human and non-human animals. This is important, as it seems to show that they can make such basic distinctions even before they are able to learn language (Pauen & Zauner, 1999). Of course, merely distinguishing human agents from non-human animals does not suffice for real perspective-taking. Children also have to be aware of, or be able to identify the behavior of their conspecifics. Studies conducted by Andrew Meltzoff (Meltzoff & Moore 1983; Meltzoff 1988a, 1988b) have shown that children have the ability to imitate soon after birth. If you stick your tongue out in front of a newborn, he will likely imitate your behavior. Interestingly, he will refrain to do so when a similar movement is performed by a mechanical object. Even very young infants thus seem to be aware of particular kinds of behavior that their conspecifics display. There is a theory about why they are able to translate the visual information into behavioral output, that does not assume complicated inferential processes happening in the infant's mind when he is imitating – such as identifying his tongue, finding out where it is, and how to move it. This theory suggests that there is a direct coupling between the motor system and the cognitive system, which draws on motor abilities to improve our understanding of observed behavior (Gallese & Sinigaglia 2014). It seems that something like this happens in the case of imitative behavior in newborns.

Finally, there is a last requirement for perspective-taking, which appears to be the most important one. So far, I tried to show that children are able to identify candidates for perspective-taking and that they can interpret their behavior and their movements from very early on. However, in order to be able to take someone else's perspective they should also be able to distinguish between their own mental states and others' mental states, more particularly between their own feelings and others' feelings as well as between their own beliefs and others' beliefs. This means that they have to make a perspectival distinction and recognize mental states in others that they do not experience themselves. There is evidence that children also develop at least rudiments of this capacity at an early age.

For instance, 6 weeks old infants are able to distinguish emotions in the facial expressions of their caregivers; and 4-months olds are able to use emotion expressions to assign a voice to a character in a movie: if they see a movie character and hear a voice that does not fit the emotional expression shown, they get irritated (Oerter & Montada 1995, p. 230). 9-month-olds can use facial expressions of their caregivers in order to assess special situations, a phenomenon termed "social referencing" (Feinman, 1982). When 9-month-olds are exposed to a potentially dangerous situation, for example an unfamiliar toy, they look at their caretaker: if the caretaker is relaxed, they take the toy; if the caretaker does not look relaxed, they do not take it. The assumption here is that the children themselves do not have this emotion; rather, they are able to interpret the emotion of their caretaker and therefore take his (or her) perspective in order to assess the situation. Generally, children develop basic forms of perspective taking as early as with 9 months when they acquire *secondary intersubjectivity*: "At 9 months of age infants begin to understand that other people perceive the world and have intentions and feelings toward it" (Tomasello 1993, p. 175).

Now let us look at data indicating when the first signs of self-awareness occur: when are children able to recognize themselves *as themselves*? An empirical test to shed light on this question is the so-called rouge test or mirror test, which children are typically able to pass between 15 and 21 months of age (Neisser 1993, p. 16). Of course, as with any empirical method, it can be debated what this test really shows (Loveland 1986). However, although far from perfect, it is still one of the best tests for the existence of self-awareness. In this paradigm, unbeknownst to the children, a red dot is placed on their nose. Subsequently, the children are put in front of a mirror. If the children do not recognize themselves, they will try to play with the children in the mirror; however, about the age of 15 to 21 months, they start to touch their own nose in order to clean it. This is taken to show that the children recognize themselves in the mirror and know that they are looking at their own nose. Thus, a form of self-awareness is displayed here: children passing the rouge test are aware of themselves as themselves; they understand that it is themselves what they see.

Between 18 and 24 months, children start using the expression "I" in the right way, they understand how the first-person pronoun works. At 24 months, children can do self-ascriptions of persistent features, such as their gender (Fogel 1997, p. 380). Arguably this is the time when they make the first steps to develop something like a self-concept. It is important to note that in the course of a child's development, the conceptual development comes after the development of the underlying non-verbal abilities – even if conceptual abilities are certainly essential for full-fledged self-awareness.

In any case, there seems to be quite some evidence that self-awareness, understood as the ability to recognize yourself as yourself, is based on perspective-taking. Small children who try to hide themselves by closing their eyes do not have self-consciousness regarding their own perceptions. They confuse their own perspective with what is going on in the world: when they cannot see anything, then no one can. In order to recognize their perceptions as their perceptions, they have to recognize the difference between their own and another person's perceptions. Note that all this can be done without referring to any criteria that identify yourself. So unlike the old reflection theory, understanding self-consciousness as perspective taking does not lead to a vicious circle. Even more important, it demystifies self-consciousness and makes it a subject of empirical research.

But what about the pre-reflexive self, this direct awareness of ourselves as it was assumed by the Heidelberg School? According to the present approach, two aspects are important: First, immediate access to our own emotions, perceptions, and beliefs comes for free, as soon as we are conscious. Having a feeling of pain means having direct first person access to this experience. What self-consciousness adds is, second, the ability to recognize this experience as your experience. And this is what is explained by perspective taking, that is, which helps us to recognize what is specific regarding our own emotions, perceptions and beliefs.

So being able to recognize yourself as yourself means being able to recognize the difference between yourself and someone else, and this ability develops in degrees and it can be investigated with empirical methods. As we have seen, it starts with recognizing a difference in emotions, which seems to work already at 9 months, and is followed by recognizing differences in perception. The next step is recognizing a difference in beliefs, which requires a theory of mind in the stronger sense – a theory about the other's mental states and beliefs. The question here is, how can a child understand her beliefs as *her* beliefs, and know that someone else might have different beliefs?

There are many studies investigating this capacity, of which I will only refer to one: the so-called Sally-Anne-test, or false belief test, which was introduced in the 1980s (Wimmer & Perner 1983; Baron-Cohen et al. 1985). Here children are shown a scenario where two characters, Sally and Anne, are in a room together. Sally places an object, a ball, into a box and then leaves the room; while she is gone, Anne moves the ball into a different box. The experimenter then asks the child observing this scene: "When Sally comes back, where will she look for the ball – in the first or the second box?". The right answer would obviously be the former: Sally still has the belief that the ball is in the first box, where she left it. However, children up to around 4 years say that she will look in the second box, because they are not yet able to distinguish between what is really the case and another person's, Sally's, false beliefs. This shows that the ability to recognize your beliefs as your beliefs develops at about 4 years. Of course, again, this ability develops in stages, and its precursors are instantiated much earlier. But a very stable finding of this experimental paradigm is that it takes around 46 to 49 months to develop.

The conclusion to take away from this is that recognizing yourself as yourself is not an ability that you either have or do not have; rather it comes in different degrees: You may be more or less proficient in this ability which develops at different stages of the ontogenetic development with regard to emotions, perceptions and beliefs.

I have tried to argue, against skepticism by Dennett, Metzinger and others, that self-awareness is not a fiction. It does not require reference to a self-object or some kind of property that happens to be specific for yourself. What it does require is the ability to recognize yourself as yourself, which can be explained as the ability to distinguish between your own experiences, beliefs, and other features and the related mental states and features of others. These abilities, particularly perspective taking, are not mysterious at all. They can be investigated in empirical studies which improve our understanding of this ability. So there is no reason to doubt that the self exists and naturalism can explain how it emerges[2].

[2] I would like to thank Laura Schlingloff for revising the manuscript.

REFERENCES

Baron-Cohen, S., Leslie, A.M. & Frith, U. (1985), "Does the Autistic Child Have a 'Theory of Mind'?" *Cognition*, 21(1), pp. 37-46;
Damasio, A.R. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam, New York;
Dennett, D.C. (1991), *Consciousness Explained*, Backbay Books, Boston New York Toronto;
Feinman, S. (1982), "Social Referencing in Infancy", *Merrill-Palmer Quarterly-Journal of Developmental Psychology*, 28(4), pp. 445-470;
Fichte, J.G. (1991), "Wissenschaftslehre Nova Methodo", in M. Frank (ed.), *Selbstbewusstseins Therien von Fichte bis Sartre*, Suhrkamp, Frankfurt, pp. 9-13;
Fogel, A. (1997), *Infancy. Infant, Family, and* Society, Third Edition, West Publishing, Minneapolis/St. Paul;
Frank, M. (1991), *Selbstbewußtsein und Selbsterkenntnis. Essays zur analytischen Philosophie der Subjektivität*, Reclam, Stuttgart;
Gallese, V. & Sinigaglia, C. (2014), "Understanding Action With the Motor System", *Behavioral Brain Science*, 37(2), pp. 199-200;
Henrich, D. (1967), *Fichtes ursprüngliche Einsicht*, *Wissenschaft und Gegenwart*, Klostermann, Frankfurt a. M.;
Henrich, D., Cramer, W. & Wagner, H. (1966), *Subjektivität und Metaphysik. Festschrift für Wolfgang Cramer*, Klostermann, Frankfurt a. M.;
Hume, D. (1978), *A Treatise of Human Nature*, L.A. Selby-Bigge (ed.), 2nd edition, Oxford University Press, Oxford, New York;
Kant, I. (1821), *Vorlesungen über die Metaphysik. Zum Drucke befördert von dem Herausgeber der Kantischen Vorlesungen über die philosophische Religionslehre*, Keysersche Buchhandlung, Erfurt (Metaphysik Pölitz);
Loveland, K.A. (1986), "Discovering the Affordances of a Reflecting Surface", *Developmental Review*, 6, pp. 1-24;
Meltzoff, A. & Moore, M. (1983), "The Origins of Imitation in Infancy: Paradigms, Phenomena, and Theories", in L.P. Lipsitt & C.K. Rovee-Collier (eds.), *Advances in infancy research*, Ablex, Norwood NJ, pp. 265-301;
Meltzoff, A.N. (1988a), "Imitation of Televised Models by Infants", *Children Development*, 59(5), pp. 1221-1129;
Meltzoff, A.N. (1988b), "Infant Imitation and Memory: Nine-Month-Olds in Immediate and Deferred Tests", *Children Development*, 59(1), pp. 217-225;
Metzinger, Th. (2003), *Being No One. The Self-Model Theory of Subjectivity*, MIT Press, Cambridge;
Minsky, M. (1988), *The Society of Mind*, Touchstone Books, New York;
Neisser, U. (1993), *The Perceived Self. Ecological and Interpersonal Sources of Self-Knowledge*, Cambridge University Press, Cambridge, New York;
Oerter, R. & Montada, L. (eds) (1995), *Entwicklungspsychologie. Ein Lehrbuch*, Beltz, Weinheim;
Pauen, S. & Zauner, N. (1999), "Differenzieren Kinder im vorsprachlichen Alter auf konzeptueller Eben zwischen Menschen und Säugetieren?", *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 31, pp. 78-85;
Ramachandran, V.S. & Hirstein, W. (1997), "Three Laws of Qualia: What Neurology Tells Us About the Biological Functions of Consciousness", *Journal of Consciousness Studies*, 4, pp. 429-457;
Sacks, O. (1985), *The Man Who Mistook is Hife For a Hat and Other Clinical Tales*, Summit Books, New York;
Tomasello, M. (1993), "On the Interpersonal Origins of Self-Concept", in U. Neisser (ed.), *The Perceived Self. Ecological and Interpersonal Sources of Self-Knowledge*, Cambridge University Press, Cambridge, New York, pp. 174-184;
Wimmer, H. & Perner, J. (1983), "Beliefs about Beliefs. Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception", *Cognition*, 13, pp. 103-128;
Woodward, A. L. (1998), "Infants selectively encode the goal object of an actor's reach", *Cognition*, 69(1), pp. 1-34.

MARIO DE CARO
Università Roma Tre and Tufts University
mario.decaro@uniroma3.it

TWO FORMS OF NON-REDUCTIVE NATURALISM

abstract

The debate on naturalism in the last years has developed around two main interconnected issues: the possibility of naturalizing the items of the manifest image of the world and the prospects of non-reductive naturalism. In this article, I will be concentrated on the second issue, by looking at two important proposals for a non-reductive naturalism: Hilary Putnam's liberal naturalism *and Lynne Baker's* near-naturalism*.*

keywords

Naturalism, liberal naturalism, scientific naturalism, Lynne Baker, Hilary Putnam

1. Naturalisms

Most philosophical labels are time-independent designators. During their respective times, Fichte and Hegel were considered idealists, Marx and Nietzsche atheists, Aquinas and Leibniz realists; and they are still considered as such. This is because the labels "Idealism", "Atheism", and "Realism" have not changed their meanings over time (or, if they have, only in minor respects). With the term "Naturalism", instead, the situation is quite different. In a very general sense, this term means that nothing can be accepted in one's philosophy that is beyond nature. Yet, the meaning of this definition is not a time-independent one. For example, Heraclitus, Jean Buridan, Francis Bacon, Giordano Bruno and Goethe can all be considered naturalist philosophers if we look at the cultural contexts of their respective times. Though, today no philosopher defending their views would be considered part of the naturalistic crew.

This happens because the meaning of the term "naturalism" is conceptually dependent on the meaning of "natural" and, indirectly, "nature" (from which the former derives); in turn, the meanings of these terms have changed dramatically over time. Giordano Bruno, for example, can be considered a naturalist as long as one looks to the Renaissance view of nature – which attributed a crucial role to vitalistic forces and secret non-causal correspondences among things –, but certainly today nobody could present views similar to Bruno's without being considered a supernaturalist (and a bizarre one, for that matter). Consequently, in the course of history different naturalisms have been developed, depending on the views of nature that each period has held. Thus, in discussing the nature of contemporary naturalism, one has to consider which is the conception of nature that the philosophers who label themselves naturalists are referring to. Still, the answer to this question is not univocal.

Most contemporary naturalists – the "strict naturalists" – take the term "nature" as referring only to the subject matter of the natural sciences, if not to the subject matter of physics alone[1]. But according to other naturalist philosophers – the "liberal naturalists" – while the subject matter of the natural sciences is certainly a fundamental component of the concept of nature, it does not exhaust it. This is because a "second nature" (to use the Aristotelian term revived by John McDowell [1994]) also exists, which is distinct from the nature that is investigated by the natural sciences. "Second nature" stands for the world of culture, into which we enter by way of education, and this is a world that is still

[1] See Papineau 1993 and 2007; Ritchie 2009; Baker 2013, part I.

"natural", even if it cannot be accounted for by the natural sciences.

In the first place, therefore, liberal and strict naturalists differ over their respective metaphysical views as to what nature is – that is, whether nature coincides with, or is broader than, the subject matter of the natural sciences. This metaphysical difference generates three other differences, respectively at the epistemological, the semantic, and the metaphilosophical level.

More specifically, the strict naturalists accept: (i) an *ontological tenet* according to which reality (that is, nature) consists of nothing but the entities to which successful explanations of the natural sciences commit us; (ii) an *epistemological tenet* according to which scientific inquiry is our only genuine source of knowledge or understanding; all other alleged forms of knowledge (e.g. *a priori* knowledge) or understanding are either illegitimate or reducible in principle to scientific knowledge; (iii) a *semantic tenet* according to which no truth-apt factual judgments exist that do not regard scientifically accepted entities, and are irreducible to judgments regarding such entities; (iv) a *metaphilosophical tenet*, according to which philosophy must be continuous with science as to its contents, methods, and purposes[2].

The main problem that strict naturalism faces has been called the "placement problem" (Price 2004) or the "location problem" (Jackson 1998, pp. 1-5). It concerns the items that are part of the common-sense view of the world (which the liberal naturalists connect with our "second nature"), but at least at their face value do not belong to the scientific view of the world – that is, are not part of "first nature". Examples of this category are moral features, free will, normativity, consciousness, and intentional properties. According to the strict naturalists, either these features are in principle reducible to the features accepted by natural science, and are thus investigable with the scientific means, or they are just fictions, in which case no judgment concerning them can be objective.

Liberal naturalism liberalizes the tenets of strict naturalism, by accepting: (i) a *liberalized ontological tenet*, according to which there may be entities that are both irreducible to, and ontologically independent of, entities whose nature and behavior are not explainable by science but are not supernatural either; (ii) a *liberalized epistemological tenet*, according to which legitimate forms of understanding (such as conceptual analysis, imaginative speculation or introspection) exist that are neither reducible to scientific understanding nor incompatible with it; (iii) a *liberalized semantic tenet* according to which there are truth-apt factual judgments that do not concern scientifically accepted entities or properties and are irreducible to judgments regarding such entities or properties; (iv) a *liberalized metaphilosophical tenet*, according to which there are issues in dealing with which philosophy is not continuous with science as to its content, method and purpose.

The main difficulty that liberal naturalism encounters may be labelled the "reconciliation problem". How can the common sense image and the scientific image be on a par with each other, i.e. without one being conceptually prior on the other? What kind of relation is there between the scientific descriptions of the world and those referring to our second-nature features? Is that a relation of supervenience (and in case, of which kind?), asupervenience, grounding, incommensurability, or what? In the next paragraph of this article I will discuss a prototypical form of liberal naturalism, proposed by Hilary Putnam in the last years. In the last paragraph I will instead discuss a different proposal that, *stricto sensu*, is not a form of liberal naturalism but has many similarities with it: Lynne Baker's *near-naturalism*.

2. Putnam's Liberal Naturalism

The *cliché* according to which Putnam is guilty of changing his mind too often is unfair for at least two reasons. One is that, in itself, there is nothing wrong – no guilt! – in changing one's own mind (unless the change is due to bad reasons or bad faith, which certainly is not the case for Putnam).

Another reason, more relevant here, is that there are many important issues about which Putnam has *not* changed his mind for many years. The fact/value dichotomy, conceptual pluralism and conceptual relativity, the externalist theory of meaning, a cognitivist and realist view of ethics, and the denial of Metaphysical realism are only some of these issues. Another, very important thesis about which Putnam has not changed his mind, with the exception of few very early publications, is that science is a fundamental source of knowledge but not the only source of knowledge. To paraphrase an excellent non-professional philosopher, for Putnam there are more things in heaven and earth than are dreamt of in our science; but still nothing can be truthfully said that would contradict science.

These views sound very much like liberal naturalism[3]. But this is no coincidence, of course, since Putnam – with McDowell and P.F. Strawson – is one of the founding fathers of that philosophical movement (whose grandfather is John Dewey). Let us now look in turn at the four tenets of liberal naturalism and see how Putnam has accounted for them.

Let us begin with the liberalized ontological tenet. According to this tenet, besides the entities assumed by the natural sciences, we should also admit the existence of other entities whose reality is presupposed either by the social sciences and/or by our non-scientific practices, and that (without being supernatural) are both irreducible to and ontologically independent of the entities whose nature and behavior are explainable by the natural sciences. According to a common strictly monistic view – advocated, among many, by Quine – the world is composed by exactly one domain of individuals (ontology) and one domain of properties attributed to those individuals (ideology) and science alone has the ability to determine what these domains are. This view characterizes strict naturalism, and Putnam strongly refuses it. However, differently from the antinaturalist thinkers, who typically defend one or the other among the antirealistic views of science (such as conventionalism, instrumentalism, or relativism), Putnam is a stern realist about science – *i.e.*, he believes that scientific theories can be (and often are) true or approximately true and that scientific terms refer to real entities also when those are unobservable. In this perspective, Putnam (1975, 2012b) has developed the famous "no-miracles argument", which advocates scientific realism by appealing to an inference to the best explanation. The core of that argument is that realism recommends itself insofar as it offers a much more convincing account of the great success of modern science than antirealism does – since for the latter view the success of science is nothing less than an unexplainable miracle.

But, even though he is a scientific realist, Putnam refuses the strict naturalists' monistic view for two main reasons. First, because of the phenomenon he calls "conceptual relativity", which means that some theories can be cognitively equivalent even if *prima facie* they appear incompatible. (Less equivocally, this phenomenon could be labelled "descriptive equivalence", since the other expression may suggest a connection with relativism and antirealism which is entirely inappropriate).

As Putnam convincingly argues, in some scientific fields, such as mathematical physics, this phenomenon is ubiquitous.

> To take an example from a paper with the title "Bosonization as Duality" that appeared in *Nuclear Physics* B some years ago, there are quantum mechanical schemes some of whose representations depict the particles in a system as bosons while others depict them as fermions. As their use of the term "representations" indicates, real live physicists – not philosophers with any particular philosophical axe to grind – do not regard this as a case of ignorance. In their view, the "bosons" and "fermions" are simply artifacts of the representation used. But the system is mind-independently real, for all that, and each of its states is a mind independently real condition, that can be represented in each of these different ways. And that is exactly the conclusion I advocate [...]. [These] descriptions are both answerable to the very same aspect of reality, [...] they are "equivalent descriptions" (Putnam 2012a, pp. 63-64).

[2] Differently from what I did in De Caro & Voltolini 2010 and De Caro 2010, in the list of the commitments of contemporary naturalism here I also mention a semantic tenet of liberal naturalism. I think this is especially important for understanding the originality of Putnam's view.

[3] For Putnam's latest views on these issues see his (2012c, forthcoming a, and forthcoming b).

The second reason why Putnam refuses the old monistic view about ontology is more interesting for our purposes. It consists in the fact that, in his view, the ontology of the world cannot be limited to the entities and the properties described by natural science.

> I do indeed deny that the world can be completely described in the language game of theoretical physics; not because there are regions in which physics is false, but because, to use Aristotelian language, the world has many levels of form, and there is no realistic possibility of reducing them all to the level of fundamental physics (Putnam 2012a, p. 65).

One of Putnam's favorite examples is that, depending on our interests, we can correctly and usefully describe a chair in the alternative languages of carpentry, furniture design, geometry, or etiquette. Each of these descriptions is useful in its own way, without being reducible to any of the others. There is no a fundamental theory of what being a chair is, so to speak. And this is valid with regard to a vast amount of entities (possibly all of them, with the exception of the entities of microphysics), since they can be described in different ways not just because of conceptual relativity, but also because things have different properties that belong to different ontological regions, to use Husserl's term.

In this pluralistic light, the old ontological project of providing a general inventory of the universe, which would supposedly encompass the references of all possible objective statements – a project of which strict naturalism is the last expression – has made us wandering in Cloud Cuckoo Land for too long (Putnam 2004, p. 85). And this means that Ontology with a capital "o" is a dead project. But another form of ontology (one with a lower-case initial) is still possible, *i.e.*, the search for the entities our best theories and practices commit us to. But this cannot be carried out if one is driven by an ideological bias that there is one, and only one, true theory of the world. Nor can it be carried out without noticing that reality has different levels. And it is a pragmatic question which level is relevant to a particular discursive practice.

However, the fact that reality is articulated in different levels raises a question about the relationship running between them. About this relationship, Putnam is straightforward: different levels of reality are linked by a relationship of supervenience (sometimes local, sometimes global) from the most basic to the less basic. In this sense, it is useful to mention a discussion between Putnam and Stephen White. White (2008) defends the idea that the "agential perspective" and the "objective perspective" are categorically different and, in fact, incommensurable, so that between them there is a relation of "asupervenience" (neither supervenience nor non-supervience). To this Putnam replies,

> I do think that all of our capacities, including "agential" ones (a category which, as Stephen White correctly argues, includes our perceptual capacities), supervene on the states of the physical universe, including, in a great many cases, past as well as present ones [...]. I am a naturalist – a non-reductive naturalist – and I don't see how any naturalist can deny global supervenience of human psychological states and capacities. (And appealing to the murky doctrine of "incommensurability" is no help). But there is no one simple answer to the question of whether our agential capacities are *locally* supervenient (supervenient on just the relevant brain-states) or *globally* supervenient on factors external to the brain, and even to the organism, because *it depends on which agential capacities one is talking about*, even if we restrict the issue to perceptual capacities (Putnam 2008, p. 29).

These ontological claims have of course important epistemological implications. In this respect Putnam holds what I have called a "liberalized epistemological" view, claiming that many cognitively non-equivalent and mutually irreducible conceptual schemes have to be used to account for the different levels of reality. And this means that, *pace* Quine, there is no such a thing as a "first-grade conceptual system" (*i.e.*, the natural sciences, if not physics alone), which is in charge of describing reality, while all the other conceptual systems are either reducible in principle to it or completely flawed. According to Putnam, we legitimately "employ many different kinds of discourses, discourses subject to different standards and possessing different sorts of applications" (2004, p. 22).

Putnam also endorses the "liberalized semantic tenet", in a very radical way. Not only he does say there are true judgments that do not concern scientifically accepted entities or properties, but he also adds that some of these judgments are objective even without describing anything; that is, there can be "objectivity without objects" (Putnam 2004, pp. 77-78), as in the case of ethical and mathematical judgments. For example, no *special* moral entities (such as free-floating values) exist that make our moral judgments true or false. This does not amount to saying that there are no *non-special* moral entities, since these certainly exist, and are the agents. Still when we say that someone is good, there is no ontologically autonomous "goodness" to which we refer.

Finally, as to his metaphilosophy, Putnam strongly refuses Quine's view of philosophy as a branch of science. According to Putnam, there certainly are legitimate philosophical issues that are not scientific in character and cannot be treated by using the methods of the natural sciences. To mention some of these issues: the ontological status of possible worlds, the conditions of a just war, the skeptical challenge to the existence of the external world, the ontological proof of God's existence, the conceptual link between free will and moral responsibility – and the list of specifically philosophical issues could go on for very long.

Putnam himself states this point with great clarity when he claims that philosophy has two faces: the *Theoretical face* which aims at clarifying "what we think we know and work out how it all 'hangs' together", as Wilfrid Sellars [1962, p. 37] famously put it in, and the *Moral face* (which "interrogates our lives and our cultures as they have been up to now, and which challenges us to reform both")[4].

It is clear that the moral face of philosophy does not depend on science as its primary source of inspiration, and even less as its foundation. This, however, does not mean that what it is said at the moral level can be incompatible with what science says about the world. If a defense of racism, for example, can certainly be criticized from a moral point of view, it is just refuted by the strong scientific evidence that human races do not exist.

Summarizing, as to his ontological, epistemological, semantical and metaphilosophical views, Putnam is undoubtedly a liberal naturalist.

3. Baker's Quasi-naturalism

A very useful distinction made by Lynne Baker in her important book, *Naturalism and the First-person Perspective* (2013), is that between the diverse forms of scientific naturalism, which depend on how its advocates respond to some crucial open issues. In particular, some of them (such as Philip Pettit) claim, and others (such as Hilary Kornblith and Philip Kitcher) deny, that all the sciences are reducible to microphysics. And some (the "disenchanted naturalists", such as Alex Rosenberg) maintain that the so-called "fundamental questions of life" disintegrate once they are framed within the scientific worldview, while others (the "optimistic naturalists", such as Philip Kitcher and Daniel Dennett) think that such questions are legitimate and can be understood (if not answered) with scientifically kosher conceptual tools. However, all advocates of scientific naturalism encounter serious difficulties when they try to naturalize – either by reduction or elimination – the most relevant features of the common-sense view of the world. In this regard, Huw Price has talked of a "Placement Problem": "If all reality is ultimately natural reality, how are we to 'place' moral facts, mathematical facts, meaning facts, and so on? How are to locate topics of these kinds within a naturalistic framework, thus conceived?" (Price 2004, p. 74).

A different route has been taken by influential philosophers such as P.F. Strawson, John McDowell, Jennifer Hornsby, Barry Stroud, and (as we have seen) Hilary Putnam, who have proposed different versions of

[4] Putnam 2010, p. 93. On this issue, see also De Caro & Macarthur 2004 and 2012.

a more liberal naturalism[5]. These authors aim at accounting for the common-sense features of the world at face value, without being at odds with the scientific view of the world.

Baker locates her view in the periphery of the liberal naturalism – a view with which she sympathizes, with an important distinction, as we will see. She explicitly sides with the liberal naturalists in claiming that what escapes naturalization is not necessarily ontologically unacceptable. When a phenomenon that is central in our lives appears impossible from the point of view of a particular philosophical conception, this is a kind of *reductio* for that conception: "We should not embrace a metaphysics that makes mundane but significant phenomena unintelligible" (Baker 2013, p. 73). Among the significant and arguably irreducible phenomena one cannot dispense with, there is a very important one that according to Baker has been unjustly neglected by both scientific and liberal naturalists: the first-person perspective of the world. According to her, genuinely first-person aspects of reality exist and they cannot be explained nor explained away by science. This, however, is not because science adopts a third-person perspective, as is commonly thought, but rather because "the so-called third-person perspective is centerless; it is [Thomas Nagel's] 'view from nowhere' " (Baker 2013, p. xix).

In this respect Baker makes an important distinction between "rudimentary first-person perspective" and "robust first-person perspective" (a distinction that is very promising, it could be argued, since it appears confirmed by massive evidence coming out of cognitive science). The rudimentary first-person perspective is a dispositional property that does not require language, allows phenomenal consciousness, and makes it possible for an organism to interact, consciously and intentionally, with the environment. The robust first-person perspective, which subsumes the rudimentary one, is the capacity that every person endowed with a language has of thinking of herself as the object of her own thought. This capacity is a dispositional property, which is expressed with I* thoughts – *i.e.*, "every thought, utterance, or action that exhibits self-consciousness" (Baker 2013, p. xx), such as 'I hope that I* will be able to write a fair review of Lynne Baker's book'. According to Baker, the robust first-person perspective is an emergent property that may globally supervene on the physical properties of the world, but can neither be explained by science nor explained away; consequently, the account of reality advocated by scientific naturalism, which is wholly impersonal, must be false.

Having a robust fist-person perspective is indispensable for self-evaluation, self-understanding, moral responsibility, agency, practical reasoning, and deliberation; and, of course, it is a necessary condition of self-consciousness. On the last issue, Baker strongly disagrees with most philosophers of mind – including Ned Block and David Chalmers – who do not find it scientifically or metaphysically puzzling.

Baker also defends a detailed non-Cartesian account of the first-person perspective, intended as an irreducible but not supernatural feature of reality. Her defense is based on two "unpopular views" (Baker 2013, p. 220), ontological emergence and downward causation. Against the mainstream, she argues that higher-level properties do not locally supervene on lower-level properties but are constituted by them – in the technical sense of "constitution" that Baker has explored at length in her past work. Still, she notices, property-constitution is compatible with global supervenience (and this may make her views less alarming for some philosophers).

Finally, Baker advocates "near-naturalism", a view that in her opinion can adequately account for the first-person perspective. Adapting Dan Dennett's famous phrase, one could say that for Baker near naturalism can give what is worth wanting in naturalism without committing us to the ineffective reduction and elimination strategies of the common-sense features of the world. In fact, on the one side near-naturalism "does not take science to be the exclusive arbiter of reality" (Baker 2013, p. 208); on other side, it is not committed to supernaturalism. Therefore, Baker seems to have a point when she claims that near-naturalism is palatable for liberal naturalists. But one thing should be noted: the suffix "near" in the expression signals the fact that this view is neutral regarding the possible existence

[5] For a general presentation of the issue, see De Caro & Macarthur 2010 and De Caro & Voltolini 2010.

of supernatural entities. Even naturalists of a liberal tendency (but not all of them: see for example Robert Audi 2000) would probably disagree with such neutrality; yet this does not change the fact that they could be happy with the positive part of Baker's conception.

I said above that Baker's view could be located in the periphery of liberal naturalism. Indeed, even if regarding its positive stances, this view could certainly be considered a form of liberal naturalism, it does not incorporate a refusal of supernaturalism, as the standard liberal naturalist views do. This is because, as we have seen, Baker is in fact neutral as to the issue if one should also accept supernatural entities in our ontology: her near-naturalism is thus compatible with both liberal naturalism and supernaturalism.

Most, if not all, liberal naturalists would disagree with this part of Baker's view, considering it too liberal – or, which is the same, not naturalistic enough. However, these philosophers would split as to the reason for disagreeing with Baker. Some, as Putnam, would accept her idea of giving a metaphysical interpretation of anti-reductive naturalism, but would refuse to broaden this view to the point of incorporating entities that would not obey the laws of nature. Other philosophers, such as John McDowell, Akeel Bilgrami, David Macarthur, and Stephen White tend instead to be quietists regarding metaphysical issues, such as the relation between physical and personal entities, whereas Baker aims at working out the framework of a unified metaphysical view of the world, which could encompass both scientific and common sense entities.

Lynne Baker's partial opening towards the possibility of supernaturalism would then be refused both by metaphysically oriented liberal naturalists and by quietists liberal naturalists. And it is indeed an open question whether liberal naturalists should prefer a quietist or a metaphysical approach – and certainly one that will be debated for many years.

REFERENCES

Audi, R. (2000), “Philosophical Naturalism at the Turn of the Century”, *Journal of Philosophical Research*, 25, pp. 27-45;

Baker, L.R. (2013), *Naturalism and the First-person Perspective*, Oxford University Press, New York;

De Caro, M. (2010), “Varieties of Naturalism”, in G. Bealer & R. Koons (eds.), *Waning of Materialism*, Oxford University Press, Oxford, pp. 365-374;

De Caro, M. & Macarthur, D. (eds.) (2004), *Naturalism in Question*, Harvard University Press, Cambridge (MA);

De Caro, M. & Macarthur, D. (eds.) (2010), *Naturalism and Normativity*, Columbia University Press, New York;

De Caro, M. & Macarthur, D. (eds.) (2012), “Hilary Putnam: Artisanal Polimath of Philosophy”, in H. Putnam (2012c), pp. 1-35;

De Caro, M. & Voltolini, A. (2010), “Is Liberal Naturalism Possible?”, in M. De Caro & D. Macarthur (eds.), *Naturalism and Normativity*, Columbia University Press, New York, pp. 69-86;

McDowell, J. (1994), *Mind and World*, Harvard University Press, Cambridge (MA);

Jackson, F. (1998), *From Metaphysics to Ethics*, Clarendon Press, Oxford;

Papineau, D. (1993), *Philosophical Naturalism*, Blackwell, Oxford;

Papineau, D. (2007), “Naturalism”, in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/spr2009/entries/naturalism/;

Price, H (2004), “Naturalism without Representationalism”, in M. De Caro & D. Macarthur (eds.), *Naturalism in Question*, Harvard University Press, Cambridge (MA), pp. 71-88;

Putnam, H. (1975), “What is Mathematical Truth”, in Id., *Philosophical Papers*, I, *Mathematics, Matter and Method*, Cambridge University Press, Cambridge, pp. 60-78;

Putnam, H. (2004), *Ethics without Ontology*, Harvard University Press, Cambridge (MA);

Putnam, H. (2008), “Reply to Stephen White”, *European Journal of Analytic Philosophy*, 4(2), pp. 29-32;

Putnam, H. (2010), “Science and Philosophy”, in M. De Caro & D. Macarthur (eds.), *Naturalism and Normativity*, Columbia University Press, New York, pp. 89-99;

Putnam, H. (2012a), “From Quantum Mechanics to Ethics and Back Again,” in Putnam (2012c), pp. 51-71;

Putnam, H. (2012b), “On Not Writing Off Scientific Realism”, in H. Putnam (2012c), pp. 91-108;

Putnam, H. (2012c), *Philosophy in an Age of Science. Physics, Mathematics, and Skepticism*, M. De Caro & D. Macarthur (eds.), Harvard University Press, Cambridge (MA);

Putnam, H. (forthcoming a), *Naturalism, Realism, and Normativity*, M. De Caro (ed.), Harvard University Press, Cambridge (MA);

Putnam, H. (forthcoming b), “Realism, Naturalism, and Normativity”, *Journal of the American Philosophical Association*;

Ritchie, J. (2008), *Understanding Naturalism*, Acumen, Stocksfield;

Sellars, W. (1962), “Philosophy and the Scientific Image of Man”, in R. Colodny (ed.), *Frontiers of Science and Philosophy* (University of Pittsburgh Press, Pittsburgh); reprinted in *Science, Perception and Reality* (Routledge & Kegan Paul Ltd, London 1963); reissued in 1991 (Ridgeview Publishing Co., Atascadero);

White, S.L. (2008), “On the Absence of an Interface: Putnam, Direct Perception, and Frege’s Constraint”, *European Journal of Analytic Philosophy*, 4(2), pp. 11-28.

KATHERINE SONDEREGGER
Virginia Theological Seminary, Alexandria, USA
ksonderegger@vts.edu

NATURALISM AND THE DOCTRINE OF CREATION

abstract

Christians hold to a distinctive view of the natural and of naturalism. In the Doctrine of Creation, Christian theologians set out the natural realm as a cosmos, a world, which has its origin and unity in God, and in this way is radically dependent. Such a world prompts Christian theologians to ask about the scope and aim of the Divine act of creation: What is it that the Lord God creates when 'in the beginning, He creates the heavens and the earth'?. Whole objects, or 'substances', have been taken as the traditional answer to that question, and some pre-modern views are set out as examples. But modernism has raised fresh objections to the traditional account. Perhaps God intended an underlying substance or noumena that cannot be known by human creatures? Perhaps, more radically. God intended only particles or force fields or energy? Contemporary views of the natural and of naturalism exert pressure on Christian teaching about the created realm, and I note some examples. Finally, I sketch a positive Doctrine of Creation in light of these modern developments, affirming traditional elements, amidst some changes, and the irreducible status of 'moderate-sized dry goods' in the world Almighty God has made.

keywords

Doctrine of Creation, Modernism, Naturalism, Anti-reductionism

In a poem of rare mystery and power, the English poet William Blake muses about a creature enigmatically titled, the Tyger. Brawny and terrible, the Tyger's 'fearful symmetry' is forged in some terrible, unearthly furnace, sinews and eyes and heart hammered out in the fire of some 'distant deep'. Looking back on his delicate *Songs of Innocence*, Blake asks the haunting question about this fearsome Tyger, 'Did he who made the lamb make thee?'. Is this Maker, who, Prometheus-like, 'seized the fire' to bring to life this terrible thing, the very Creator of Heaven and Earth? Was his fiery act of manufacture a creation? And like the Lord God of Genesis, did this Maker 'smile his work to see?'. Blake does not answer his own questions; perhaps he considered the *Songs of Experience*, from which Tyger is drawn, to raise questions unanswerable within the life of sorrow and fear and sin we know all too well.

Blake could not answer these questions, it seems, but perhaps he could lend us his framework to explore the doctrine he adumbrates so finely: the Doctrine of Creation in the modern age. A child and architect of the modern, Blake sensed in his poetic imagination the elements of the modern Doctrine of Creation in the west. The Tyger sets out a vision, dark and brooding, of a world that is at once natural and distorted, familiar and alien, and a Maker at once Lord of Grace and a Stranger, a Creator and a Terrible Power. Here we see the themes that will carry us from the brink of the modern – the late Enlightenment and burgeoning Romanticism – to the suffering heart of the 'terrible century', the 20th, and the dawn of our day, the 21st. We can summarize these elements this way: the theme of the natural and naturalism; the theme of the artifact; and the theme of genesis, the absolute beginning of all things.

The doctrine of creation most broadly and traditionally treats the absolute origin of all things from God, and the prominence of our third theme, 'genesis', marks out the modern era as fully traditional in the midst of its many innovations. To be sure, 'genesis' in the 19th and 20th centuries could hardly speak with the confident tones of earlier eras. From the rise of modern astronomy and particle physics, to the carbon-dating of our earth, and the development of present-day animal species, the genesis of all things from God has found itself in the midst of pitched battles over the place and cogency of Christian doctrine in an intellectual climate dominated by the exact sciences, and the fear of them. It will take all our concentration to set out this element in the modern doctrine of creation without falling prey to the old, and discredited, story of 'religion against science', on one hand, and the newer, but hardly more persuasive story, of science as the confident and supreme champion of the entire field. In sum, we will see the theologians of our modern and post-modern age strive to confess the doctrine of creation in a world that remembers Blake's natural Lamb of Innocence but cannot forget the Tyger that roams freely in our day, the natural Lamb and the artificial Tyger, each in its own way mysterious and demanding, each in its own way dependent upon the genesis of the Almighty Maker of heaven and earth.

1. The Genesis of the Natural World

Both the higher unity, but the deep divisions, too, in the understanding of God's genesis of creaturely nature point to an original and rather unexpected question that will begin our entire investigation: when God created all that is, just what is it he made? Aspects of this question are not new, of course. As we will see, traditional elements will emerge throughout the discussion of this topic. But in the main, this is a modern question, raised by modern science and the philosophy that accompanies it. Although this topic will return in different guise when we discuss the modern conception of nature and the natural, it belongs here as the *precondition* for any discussion of creation itself. The identity of that reality God created – its fundamental character – sets the terms of the debate over creation, and no exchange among modern theologians of creation can be intelligible apart from this conceptual underpinning. Just as Descartes' analysis of the nature and relation of body and mind set the terms of all modern debate about the mind and its relation to the brain, whether Cartesian or no, so the fundamental analysis of the creaturely sets the terms of debate about the natural, whether Christian or no. It is the lens through which all modern, western theologians, and their opponents, see the world.

So, just what is it that God creates in the beginning of all things? Now, the instinctive response of most Christians through the centuries has been rather straight-forward and filled with sturdy common sense: God makes all the things we see on our earth, and all that belongs to the starry heavens that stretch out beyond our earthly sight. It is just this insight that is quietly affirmed in a straight-forward or 'plain' reading of *Genesis*. Greater and lesser lights; waters beyond the heavens and on the earth; swarming creatures of all sorts and winged birds; fruit-bearing trees; men and women, and all animals; and light itself: all these are made by the Lord God, and fashioned into a Garden fit for the human creatures to tend and to flourish within. As we will see, this stout affirmation of God's will to create *things*, animate and inanimate, will lead to complex and painful encounters with the science of the present age. Yet it is the plainest, and to many, the most compelling answer to the question before us: Just what did God make when he created the world? And it is not without defenders of a very sophisticated sort in this age and in the past. To express this common-sense insight in more scholastic and philosophical language we would say: God created, without any prior material or aid – *ex nihilo* – complete or 'whole substances', the inert and living matter, the animals and organisms, the planets and stars in their courses, the measureless galaxies that make our world a cosmos. It is this language of 'whole substance' which will find its way into the documents of Vatican I (1869), and its controversial definitions of nature and grace. In the first decree of Vatican Council I – the Canons of the Dogmatic Constitution of the Catholic Faith – we read the following firm affirmation of a traditional Latin Doctrine of Creation: "If anyone does not confess that the world and all things which are contained in it, both spiritual and material, were produced, according to their whole substance, out of nothing by God, let him be anathema"[1]. But the roots of this Council reach much further back, back to the greatest scholastic theologian in the west, Thomas Aquinas.

Thomas gives voice to the common-sense tradition in his Doctrine of Creation, relying on Aristotle's notion of substance – itself a complex concept with its own history – when he asserts that God created the world in one, simple, motionless act, bringing out of nothing whole substances, both matter and form (*Summa Theologiae I*, q. 45, a. 2, ad 2). Now, Thomas knew perfectly well that the cosmos was filled with more than *objects*, living or inert; he knew that the world of things was qualified by innumerable properties or characteristics, and he recognized that certain immaterial realities – ideas, values, numbers, and time itself – governed much of what we call our world. These were also created by God, Thomas firmly concludes, but they receive a special delimitation: they are 'con-created' by God, as these properties or *qualia* accompany all that is.

To advert to more modern terminology, and putting J.L. Austin's phrase to rather other purposes, we could say, in this common-sense reading, that God creates "moderate-sized dry goods" (Austin 1962, p. 8) when he turns outward to make finite reality. Now, notice how such a conception affects the doctrine of creation in all its parts. When we encounter debates over Darwin's theory of evolution, say, or Heisenberg's theory of thermodynamics, or in another dimension, astrophysical accounts of the Big Bang, we see modern theologians of the "moderate-sized dry goods" school attempting to square their doctrines with these scientific accounts *as they relate to visible, tangible objects*. In their Doctrines of Creation they are 'anti-reductionists'. The *scopus* or goal of God's creative will, that is, is directed toward *objects*; the debate, for these theologians, *assumes* this goal and from this presumption, turns to the questions that remain on the composition of objects and their creaturely origin and destiny. For this reason, the evolution of the species posed the greatest threat to these theologians' doctrine: Natural selection concerns and pre-supposes medium sized objects in all its varying interpretations.

Of lesser danger to this school are the theories of modern quantum mechanics or astrophysical origin and collapse, for these theories are seen only to touch on the parts or elements of physical reality which compose objects, and not the objects themselves. A kind of 'instrumental cause' is assigned to these theories of subatomic or cosmic physics: God may make use of these particles and their behavior to achieve his goal, the creation of medium-sized objects in a harmonious universe. Just as a carpenter may make use of a hammer or level to set out the framing for a house, so God may make use of these physical elements and laws to create all animate and inanimate things, and in both cases, the instruments drop out of sight when the finished house, or cosmos, is complete. (Martin Heidegger made a similar point about the metaphysical status of instruments, *die Wären*, in *Being and Time*; and in a different key, Ludwig Wittgenstein made an analogous point about objects embedded in practices or 'language games'). In all cases, the *telos* or goal of God's creative will is the finite object, and the Creator's sustaining and judging and governing of the cosmos will be measured by the Divine decree concerning the things, not the elements, of this world.

Not so do others argue. For these other theologians, the *scopus* of God's creative will is the *fundamental particle or law* that will then result in visible and finite objects. God's intention or aim, we might say, is toward the *infinitesimal* and the medium-sized objects which emerge from these particulars and their relations are the out-working or, more daring still, the *epiphenomena* of these deep realities. As with the 'medium-sized dry goods' school, so with this school – we might call it, 'reductionist' – there are both ancient and modern philosophical and scientific correlates. To find our ancient corollaries we must reach back to the very roots of the western philosophical tradition, to the 'pre-Socratic' philosophers of Attic Greece.

Ancient indeed is the human impulse to discover the deepest reality or fundament of the world. Many of the earliest philosophers – as do their modern counterparts – held that the foundation of things could be uncovered by going *deeper*, moving down through the layers of the known and visible world to a hidden and truer element that is the basis of all things. For such thinkers, all things are *in reality*, one thing or one kind of thing; though the cosmos appears diverse it can properly be reduced to one element, one particle or kind. Heraclitus taught that the world, in its deepest and truest sense, was fire. As there is no flame without motion, so the cosmos as a whole at its deepest reality is change –ceaseless motion and alteration. To be sure, many things in this cosmos appear static, permanent, unshakeable; but it is just this appearance to eye and common-sense experience or measurement that must be set aside and seen through to its deeper identity. (A parallel but inverted schema can be seen in Parmenides, for whom Being is eternal and all change an illusion). Common to reductive accounts

[1] The Decrees of the First Vatican Council, Dogmatic Constitution of the Catholic Faith, Canon 1, found at http://www.papalencyclicals.net/Councils/ecum20.htm, the Vatican official web site.

of creation, ancient and modern, is this appeal to the *conceptual* reality of all things, at once deeper and higher than anything our senses and instruments can record. Democritus began a long line of analysis in western thought when he sought the fundament of reality in 'atoms', those parts of whole objects that could be divided no longer. These were 'simples', the deepest and truest building blocks of reality.

A second form of simplicity proved far more troublesome for the doctrine of the world's creation in time, however. In its wake this form of simplicity provided a ground for holding that the world is eternal. We might think of this as a second root of the reductionist impulse in the medieval concept of nature itself. Much to the dismay of modern interpreters such as Colin Gunton (1998, esp. Ch. 2, pp. 14-40), the notion of the simple carried over into Christian doctrines of created or 'material' objects, in Thomas and other Augustinian theologians. All created things, these medievals said, were composed of an utterly simple, yet utterly inferior kind of stuff titled 'prime matter': without form or definition, it was "close to nothing" (Conf. XII, 6, 6), in Augustine's fateful phrase, and just so could enter into the composition of every created thing. We should be quick to note, however, that such reductionism remains an impulse only, as created objects, for these thinkers, are far more than their matter – indeed their reality lies not in matter at all but in their definition or 'form'.

When we turn to our era, however, we see the strong resurgence of full-throated reductionism in philosophical and scientific circles, and its downward pressure on the Doctrine of Creation. Consider that architect of English Enlightenment, John Locke. In his *Reasonableness of Christianity*, Locke affirms – though in passing – the Doctrine of Creation and God as Almighty Creator to be the bedrock of rational religion. Locke's Christianity is hardly traditional or dogmatic, however, despite this conventional nod toward the Doctrine of Creation. Famous to the *Reasonableness*, after all, is Locke's confident assertion that nothing more is required of the Christian than to assent to the teaching that 'Jesus is the Messiah', an assertion considered 'reductionistic' already in Locke's own day. 'Rational religion' was certainly reductionist in just this sense – the fewer dogmas the better. Yet the reductionist commitment of modern philosophy does not properly pertain to elements of Christian creedal belief. Properly, reductionism in its full power pertains to worldly ontology, to the theory which enumerates the kinds and qualities of finite, created things. For John Locke, a certain form of reductionism in creaturely substance makes our knowledge of creation, and the aims of the Creator, deeply mysterious. Locke-interpretation is notoriously vexed, so we must treat carefully here. But his positions – however interpreted – are so vital to modern conceptions of epistemology, metaphysics and religion that we must hazard a reading all the same.

In his *Essay in Human Understanding*, Locke draws a famous distinction between the substance or, literally, the underlying reality of a thing – "it is that which I know not what", on one hand – and its appearances to our eyes and thought, on the other, its host of primary and secondary qualities. Objects are congeries of *qualities* as they strike our senses and awaken our intellect. Two sorts can be intellectually discerned: primary qualities, which inhere in their substance apart from our sensing them; and secondary qualities, which depend upon our encountering the object, and judging it. Locke seemed to think that extension – Descartes' great property of matter – belonged to primary *qualia*, but added solidity and impulse as 'objective' properties of things. Secondary *qualia* consist in the properties we most common-sensibly associate with things: color and taste and texture, even utility. Already, the notion of 'primary quality' reduces objects to elements – 'atoms' or perhaps, 'corpuscles' – that fall outside human sight and touch. But the deepest reality of an object lies far deeper, in the unifying concept of 'substance', a reality so metaphysically hidden that we can know just nothing about it. A great gulf is here fixed between our experience of the world and its deepest reality, a gulf that will in time be known as philosophical 'Idealism', though its earliest advocates, Locke and George Berkeley were considered to be classical empiricists. The stern transcendence and hiddenness of substance in Locke's reflections lead him to teach that God's providence fits us out to see the world after a human and creaturely fashion, and graciously shields us from sensing the world as would a powerful telescope or perfect microscope – a monstrous and debilitating power for a finite, human creature. Yet the goal of God's creative will is the substance, with its primary properties – the deep and true unity of all things – and it is just this we can conceive through reflection but never encounter or know.

Such ideas live on in the 'philosopher of Lutheranism', Immanuel Kant. For Kant, the world of medium-sized objects could be understood and known only by retaining clearly in the mind a distinction or schema that separates the truest reality of a thing from its appearance to the senses. That distinction is Kant's celebrated contrast of the *Noumena* from the *Phenomena* in every act of knowing. Kant does not deny that we have certain and trust-worthy knowledge of the world; indeed his Critical Philosophy presses every lever to achieve such certainty under the conditions of modern scientific thought. Yet, like Locke, Kant's distinction between 'things in themselves' – of which we can know strictly nothing – and 'things for us' forces Kant to radical positions that threaten the foundations of his very campaign. So radical is Kant's denial of the experience of, and so, the knowledge of the deep underlying substratum of objects that it is not clear whether in the *Critique of Pure Reason* Kant affirms a particular substance underlying each object, or whether in the end, he must affirm that there can be only one *Noumenon*, an utterly uniform Simple or Prime Matter that supports each object and its qualities (Kant 1929, First Division, Ch. II, *The Deduction of the Pure Concepts of Understanding*).

Puzzles of this kind led Kant to express reservations about most traditional metaphysical and theological categories, from the doctrine of the soul to the doctrine of *Creatio ex nihilo*. Dogmatic doctrines of this sort must be relegated to a realm of moral and intellectual usefulness: these Ideas regulate and limit our thought, so that, in Kant's celebrated trio, we can recognize 'what we can know; what we can hope; and what we can do'. Kantianism, then, is reductive in a critical sense: the truest reality of the world lies underneath what we encounter and know; we cannot know it but can only infer and point to it; it may be in reality but one substance; and we cannot properly know but rather believe and postulate that God, as Creator, willed and sustains it in being.

Modern scientific accounts of the physical laws of finite objects do not stray so far from Kantianism, though in a strong and reductive sense. Consider the modern thermodynamic concept of the cosmos. Here we find a reductionism so thorough-going that to apply it directly to the Doctrine of Creation would entail a symbolic or 'mythic' reading of Genesis altogether. That is because the objects named in the Creation narratives could scarcely constitute the goal of an omniscient Creator; rather the Author of the physical laws of the universe would aim instead at the deep and universal reality of the cosmos, *energy*. Matter, for these physicists, is a form of heat or energy; from the largest visible object to the tiniest sub-atomic particle, energy constitutes the building block and deepest reality. Indeed, the very notion of an object or thing is revised in such quantum physics. All physical things are composed of atoms, these scientists tell us, and these atoms, far from representing Democritus' simples, are themselves divisible into particles, each bundles of energy. Atoms, molecules, compounds organic and inorganic, elements, minerals, gases and liquids: all are forms of energy, joined together by chemical bonds that are themselves forces of energy. To break down and decay, to cook and slice and boil; to eat and digest; to separate in nuclear fission – in all, energy is released and in the latter, tremendous, annihilating energy, a power that has re-shaped modern politics and modern war.

Parallel to such descriptions of the object as energy is the modern notion of force-field, a notion associated with Michael Faraday and put to great dogmatic use by the modern Lutheran theologian, Wolfhardt Pannenberg. For Faraday, the force-field expressed the unique properties of magnetism, a power fascinating to the early scientific naturalists. Magnets attracted iron filings in *patterns* drawn around the magnets' poles. These patterns marked the outer reaches of a field where magnetic force would register and attract. Later physicists generalized Faraday's findings to the cosmos as a whole:

the universe was an interlocking structure formed by the forces of energy in relation and repulsion to each other. The world of things is revolutionized. No longer free-standing or independent, no longer discrete substances, however counted and conceived, creaturely objects are now 'nodes' in a web of energy, places of density where energy has coalesced and become visible to the naked eye. This web of relation that gives rise to objects, scarcely conceivable to earlier generations, has now become the fundament of all finite reality, the energy that drives the universe in all its parts. Reductionism could hardly find a greater partisan than these theoretical physicists. The Heraclitan fire returns now under the idiom and concept of energy and its forces. One step remains.

In modern philosophy of science, or in metaphysics, reductionism is laid out as a complete theory of the cosmos and all things within it. For these philosophers, particularly in the Anglo-American analytic tradition, all objects, organic and inorganic, all artifacts and culture, every thought and hope and belief, all matter, living and inert must be in fact and reality a collection of sub-atomic particles. For metaphysicians such as W.V.O. Quine (1969) or philosophers of mind such as Jaegwon Kim (1998), every thing that is and every thought conceived and held must be traced back to these particles, either in element and molecule, or in brains and their chemical structure and state. The biochemical account of the physical universe, sketched above, has become in these philosophers a metaphysical *theory*, a complete doctrine of everything. In the philosophy of mind, such philosophers are 'physicalists'; in metaphysics, 'reductive' or 'eliminative materialists'. It is important, and difficult, to see just how radical this position is.

If we were to count up all the things in this world, these philosophers say, we would count no trees, no rocks nor birds, no kitchen chairs nor dessert plates, no Sistine Chapel, no Michaelangelo. That is not because such things do not *matter* to these philosophers; far from it! Rather, these beings and objects belong to a human, cultural, and linguistic world that we might call 'phenomenal', following Kant, or 'intentional', following the physicalist Daniel Dennett (1996). But such artifacts and conventions and practices, if they are to belong to a true and scientific account of the world must be seen for what *in truth* they are: a collection of particles 'arranged thing-wise'.

Should any such philosopher be a Christian, he or she would affirm that the Creator God – wholly omniscient, wholly immaterial and transcendent – would create this realm of quarks and positrons and electrons, and would decree the physical laws that would govern their ordering and organizing as the medium-sized objects human beings see and prize, love and fear. The cosmos such a God would create would be exhausted in the infinitesimal particles of energy that compose and structure and cause the universe and all its parts.

With the help of these philosophers we have reached the antipodes of our common-sense readers of Genesis and their anti-reductive kin. This array, from the theologians and philosophers of ordinary objects, to the scientists of modern quantum mechanics, to the philosophers that translate their findings into metaphysics, all contribute an answer to the background question: When God created the world, just what did he create? From medium-sized objects, to prime matter, to quarks and positrons, the notions of the natural and naturalism have guided the modern doctrine of creation, for good and for ill.

2. A Theological Proposal

So, I want to leave as much scope as I can for Christian theologians when they face the demands of naturalism. Yet I would not do justice to my own field, systematic theology, if I did not offer my own accounting of the relation of Creation to the natural and naturalistic. There is every reason, I believe, for Christian theologians to defend and take seriously the world of *objects*, of moderate-sized dry goods. With every scientific theory in place, from cosmology to particle physics to evolution, Christians may still with confidence hold that the Bible speaks without hesitation of the creation of *things*, not particles or force-fields or natural laws, if such there be; but of individuals and of kinds. Christians need not undertake the sorry endeavor of a *harmonization* of the Book of Genesis with modern day astrophysics; we can quietly, or gladly, place that effort on the shelf marked, 'false starts in dogmatic theology'. For religion and science do not enter the world of the real through the same gateway, nor do they work on the same floor – though they serve the same Master and aim at the same universe of the real. Rather, Christians should rightly expect that the Bible will give them an impulse, a guiding hand, a *telos* by which modern doctrine should be forged. The Book of Genesis, in just this way, points theologians to the proper *scopus* and goal of their concern: the environment and thought-world of whole, complex and real living beings. We need not *rank* such creatures, though to be sure, the door remains open to a hierarchy of forms of life. For my own part I will confess that I believe human beings to stand apart and distinct from other animals and plants; but I will confess too that I believe Almighty God is far greater pleased with other living creatures than with us and our kind, the great predators and destroyers on God's fair earth.

Far more significant, however, than the matter of ranking comes the place of diversity in Christian Doctrines of Creation. Once again, I believe that theologians have every reason to hear in Holy Scripture an underlying and persistent percussion of the diverse and multiple and richly complex. Christian theologians, that is, have good reason to resist the ancient pull of reductionism, of simplicity, and of uniformity. The integrity of the world – that we live in a *cosmos* not a disordered array – rests not on the conviction that at base everything is one and of one kind. Rather, the remarkable and irresistible conviction that we live in an integrated whole, a working universe and a home, rests not on its substance but rather on its *relation* to Another: unity is an external and relational property, an essential one. The entire world comes from God, the Creator, and in virtue of His work and gift, it is a whole. The rich diversity of this planet, and perhaps other planets and star systems as well, is an exceedingly good gift that need not be thought away in a mis-guided search for simplicity and coherence. The metaphysical wholeness and interconnection of this earth rests on its origin and destiny – in scholastic idiom, its exit and return to God. The world is natural, that is, but not alone. Or to speak once again in poetic diction, this time in the stately words of Gerard Manley Hopkins:

> And for all this, nature is never spent; / There lives the dearest freshness deep down things; / and though the last lights off the black West went / Oh, morning, at the brown brink eastward, springs — / because the Holy Ghost over the bent / world broods with warm breast and with ah! bright wings (*God's Grandeur*).

3. Conclusion

William Blake introduced our themes for a modern Doctrine of Creation: of the Tyger, burning bright, sinewy and terrible, an Artifact forged in an industrial age; and of the Lamb, innocent and mild, born in some distant garden when Nature was young. The modern Doctrine of Creation has encountered both animals in its complex journey through the thought-forms, philosophy, and science of our world. Christians have struggled to understand the very foundation of the world – its composition and character – and have sought, at times at great cost, to find God presence, design, and will in the ordering of Nature's laws, growth and creatures. They have witnessed over two long and often brutal centuries the godforsakeness of a world seemingly left to its own cruel self-destruction in war and famine and despoliation. Yet Christians have remained faithful to the Doctrine of Creation's central tenet: that God is the Absolute Origin of all that is; that what God has fashioned is wonderfully made, and rich in Divine benevolence; and that human life, however ordered and however wayward, receives from this natural world a grace and gift fresh each morning.

REFERENCES

Austin, J.L. (1962), *Sense and Sensibilia*, Clarendon Press, Oxford;
Dennett, D. (1996), *The Intentional Stance*, MIT Press, Cambridge;
Gunton, C. (1998), *The Triune Creator*, Eerdmans Publishing, Grand Rapids;
Kant, I. (1781/1929), *The Critique of Pure Reason*, N.K. Smith (trans.), St Martin's Press, New York;
The Decrees of the First Vatican Council, Dogmatic Constitution of the Catholic Faith, Canon 1, found at http://www.papalencyclicals.net/Councils/ecum20.htm.

ROBERTA DE MONTICELLI
Università Vita-Salute San Raffaele
demonticelli.roberta@unisr.it

HAECCEITY? A PHENOMENOLOGICAL PERSPECTIVE

abstract

The concern of this paper is the nature of personal identity. Its target is the account Lynne Baker gives of personal identity in terms of haecceity, or rather, in terms of that particular reading of Scotus' principle of individuation that has been widely accepted in a late 20th century debate on the metaphysics of modality (Plantinga 1974, Adams 1979 and others) and that Baker's account appears to share. I shall try to show that such "haecceitistic implications" (Baker 2013, p. 179) of her theory of personhood miss something essential to the very question of personal identity, such as the question emerges within the lifeworld, i.e., in the world of everyday encounters and ordinary experience. This "something essential" seems to be better accounted for by a different theory of essential individuation or haecceity, which, as it happens, turns out to be more similar to Scotus' original theory (prior to Occam) than modern haecceitism.

keywords

Personal identity, personality, individual essence

1. A Crucial Question

Baker's theory of personal identity is a completion of her deep and rigorous view of personhood. Yet it is far from obvious that the former is logically dependent upon the latter view, though I shall not raise this issue. I will presently only address Baker's theory of personal identity. What strikes the reader is its remarkably deflationary appearance. It appears to be a critical deconstruction of all "informative" theories of personal identity, that declines to present an alternative (informative) theory. And that is quite on purpose, for any "informative" theory, Baker thinks, is one more example of that "wholly impersonal account of the world" (2013, pp. xv) characteristic of (scientific) naturalism. "Impersonal", in this context, must be understood as "third personal". All informative accounts of personal identity – so goes Baker's claim – conceive of personhood in non-personal or sub-personal terms. And this is exactly what is supposed to make them informative. But if personhood *cannot* be understood third-personally, then we cannot give a non-circular condition for personal identity over time. Given that a persisting first-person perspective cannot be but the one of that persisting person, that person's sameness over time is presupposed in the identity condition. "*You,* a person, continue to exist as long as *your* first person perspective is exemplified" (2013, p. 144).

I wholeheartedly endorse the main point. *If* what makes personal identity theories informative is that personhood is accounted for in non-personal or reductive-naturalist terms, then those theories overlook the essential feature of being a person, and *a fortiori* that of being *this* person, one and the same, persisting over time. But I don't endorse the premise. It is true that all the recent examples of informative theories I am aware of do understand personhood in non-personal or sub-personal terms. But I believe that alternative ways of working out an informative theory of personal identity remain on the table.

Maybe such a non-reductive but informative theory, though, should be more ambitious than traditional ones. Maybe specifying *a condition of temporal persistence for persons* is only part of a wider problem concerning the very nature of *individuality*, the solution to which is thus key to solving the problem of personal identity over time.

Before cashing out these suggestions in greater detail, let me give a general idea of my perplexity about Baker's, by my lights, deflationary strategy.

Baker's theory of personal identity in terms of modern haecceitism (as opposed to Scotus' actual principle of individuation) is a brilliant solution to what I will call *the crucial puzzle of personal* *identity*. Yet it is a solution, if I may say so, *not (entirely) true to the sense of the puzzle*. It ultimately ought to answer to the intuitive, pre-philosophical sense of the problem of personal identity, which, incidentally, is one of the few philosophical problems with deep roots in the world of everyday life. It is one of those rare philosophical problems that sound quite intelligible in their naïve, pre-philosophical understanding.

Putting the point in Baker's own language, her view of personal identity does not seem to take the problem as seriously as a real and decisive question originating from the world of pre-theoretical encounters deserves. The question of personal identity is indeed one belonging *par excellence* to the "metaphysics of everyday life" (Baker 2007). Baker's theory of personal identity, I have said, is a deflationary theory. By that I mean that it takes the question to be deceiving or illusory if it asks for an informative answer. Because this expectation is unjustified or unreasonable, a circular answer will suffice as a kind of Wittgensteinian therapy for pseudo-problems. But does the deflationary strategy do justice to the metaphysics of everyday life? Should not we first try to unravel the implicit, often confused meaning of those basic questions which arise in the lifeworld across all cultures, rather than brushing them aside as logical or conceptual errors? Should not we attempt to clarify the desire for information rather than dismissing it as illusory?

There is a question that we raise all the time, crucial to our lives, values, interests, crucial to ethics, law, politics, friendship and love, and the question is, *Who* are you? *Who* am I? Because such a question is so significant, and so difficult, its meaning cannot be such that the general form of an answer to it turns out to be non-informative, or circular. For an account of personal identity, at least from the point of view of a philosopher taking the lifeworld seriously, should provide us exactly with a better understanding of *the general meaning of this basic question*, one that would *not* make the question hopeless or redundant. It should shed light on the general form of an answer to it, so as to tell us in which direction we might turn our gaze in searching for the answer in particular cases. And it should do so in such a way that explains why that basic question is so hard and so crucial for us, or how it is linked to what is so singular about us as individuals, about *each one* of us. Dealing with this "more ambitious task" by *first* addressing the very nature of personal *individuality* would *then* put us in a position to solve the narrower problem of personal *identity across time*, yielding a non-circular condition of temporal persistence for persons. Or so I shall argue.

2. What Exactly the Traditional Metaphysical Problem of Personal Identity is About

Let us first consider the terms of our problem more precisely. I shall defend a phenomenological perspective. Yet my purpose, like Lynne's, has nothing to do with what is called "narrative identity", that some (like Paul Ricoeur, Dan Zahavi)[1] take to be part of a phenomenological account of personal identity.

A brief clarification of what a phenomenological perspective amounts to is in order here. It is the perspective one has when adopting the phenomenological stance. Adopting the phenomenological stance toward any object is clarifying how that object appears from an appropriate first-personal perspective, e.g., a perceptual one, if it is a perceptual object, or an emotionally qualified one, if it is an object of emotional experience, and so on. In short, adopting the phenomenological attitude means putting oneself ideally in the place of the subject of some kind of intentional state (in the Husserlian sense of "intentionality", e.g., the basic subject-object structure of consciousness). One adopts this stance "ideally" by "bracketing" whatever is contingent upon an actual subject, e.g., as this person I am.

I endorse another qualification Baker makes about how to account for personal identity. As she explains, the problem is not how we re-identify a person, nor does it have to do with psychology, not directly at least.

[1] P. Ricoeur (1990, pp. 137-198); D. Zahavi (2005, pp. 106-114).

Indeed, the problem is metaphysical. Let us recall a recent rephrasing of the problem by Harold Noonan: "The problem of personal identity over time is the problem of giving an account of the logically necessary and sufficient condition for a person identified at one time being the same person as a person identified at another time" (2003, p. 16).

Lynne Baker distinguishes Simple and Complex Views of personal identity in this sense.

Simple Views are simple because they hold personal identity over time to be non-analyzable, similar to the case of the Self that enjoys Cartesian self-reference, which is also supposed to be unanalyzable. Hence such views cannot give an informative, or non-circular, criterion of identity, *i.e.*, one not already presupposing that identity.

Simple Views are typically immaterialist. They tend to identify persons with immaterial minds or souls.

Complex Views do specify necessary and sufficient conditions of personal identity over time. They do not presuppose that RDM at t_1 is the same as RDM at t_2, but give necessary and sufficient conditions for that identity to hold, such as, for instance, persistence of body and brain, psychological continuity, or continuity of mental states[2].

Complex Views are typically reductionist about persons. That is the price they pay for being informative.

Baker's Not So Simple Simple View rejects the analysis of personal persistence in terms of subpersonal properties and relations, thereby sharing the attitude of familiar versions of the Simple View, but only up to a point, since it also rejects immaterialism. If I follow Baker correctly, it is not because a person is something simple and unanalyzable that non-circular conditions of persistence inevitably fail. It is rather because the capacity for first-personal *self*-reference (of a reflective or robust kind) is a necessary condition for a person to exist. A persisting self is embedded, so to speak, in the very definition of personhood.

For this reason, a person exists when and only when *her* first-person perspective is instanced, or in all possible times (and worlds) in which it is. If one wishes to specify a (numeric) identity condition holding for some person, one will have to make reference to that person in the *explicans*: *Lynne Baker's* perspective, *your* perspective, *my* perspective.

That is why the explanation is circular, as this "Bakerian Identity Condition" (BIC) makes clear:

> (BIC) $\underline{x}$ at t_1 is the same person as y at t_2 iff the state of affairs of $\underline{x\text{'s}}$ exemplifying a first-person perspective is the same as the state of affairs of y's exemplifying a first-person perspective (Baker 2013, p. 150, emphasis added).

I claim that this is a deflationary theory because it accepts circularity not only as inevitable, but as an obvious consequence of an illuminating truth concerning personhood, namely, that the identity of a person across time cannot really be given in non-personal terms, *i.e.*, "from outside" of that person's life. For what else is persistence over time, for a person, if not living her life, making choices, questioning herself and her choices, suffering remorse and regrets, and the like?

While I do wholeheartedly agree with this last point, I doubt that it implies there is *nothing* to discover about the identity of Lynne Baker that would not be given from Lynne's first-personal perspective, or from within her life. I agree that there is nothing to discover *in sub-personal or impersonal terms* – for such discoveries would not tell us what it is to be Lynne Baker. But I claim that third-personal talk is not necessarily non-personal or sub-personal. To understand why, consider the following case, a slight variation of a well-known argument.

[2] Contemporary examples of Complex Views: S. Shoemaker & R. Swinburne (1984); D. Parfit (1971); D. Parfit (1984); D. Lewis (1983).

3. A Borgesian Library

Suppose I am an amnesiac about what happened to me prior to last year – and in fact I now live in another country, with a different passport and another name. Reading in a library, I discover some works by a certain RDM, which I find extremely exciting. After having read all I can find by and about her, I decide to write a biography of RDM. Now, once finished, what I wrote is a biography, *but not an autobiography*. It happens to be about myself – but I ignore that it is, and I write about myself exactly as I would write about anybody else.

This case has some remarkable implications.

First of all, it shows that third-personal speech need not be impersonal or sub-personal. A biography is a perfect example of this claim.

Further, the fact that personal identity cannot be construed in sub-personal or non-personal terms does not mean that it can only be given from *one's own* personal perspective.

I suppose that Baker would agree with that. While written in third-personal language, a biography cannot help referring to its subject *as the subject of a first-personal perspective*, exactly as we do when addressing mutually in conversation or speaking of other people. Understanding others, reporting their deeds and beliefs, investigating the reasons for their choices is only possible under the assumption that they do have robust first-person perspectives. For this is exactly what rational agency presupposes, as Baker (2013, chapter 8) convincingly shows.

So, the future biographer who will reconstruct my life before succumbing to amnesia as well as the amnesiac span of my life will provide all the necessary evidence that I, the amnesiac person in a library at time t, was in fact RDM, the author of some works written before time t. But will he *need* to take up my own perspective on myself to identify me correctly, thereby showing who I was and am? Not necessarily, I would venture.

The second remarkable thing that this case shows is the role a Cartesian "spirit" plays within the account of a robust first-person perspective, by which I mean the amount of "Near-Cartesianism" it tolerates and exploits in the form of essential or irreducible self-reference.

For suppose that at some point I realize that *I am RDM.* This proposition cannot possibly be replaced by the proposition that *RDM is RDM,* without a very significant loss of information. The first one can be a shocking discovery for me, changing my present life. The second is a tautology.

Used in one way, this argument may be good support (equivalent to that of the messy shopper)[3] for Baker's irreducibility thesis (BIT):

> (BIT) 'I am LB', entails that I have a first-person perspective, which is irreducible and ineliminable (from a true description of the world). (Baker 2013, p. xv)

In fact, the case shows that:

a. *There is a way in which a subject is given to herself, a way of self-reference, which is quite independent of any objective or third-personal reference (such as biographically true descriptions), so that the former (first-personal) can be preserved when the latter (third-personal) is excluded, or "bracketed".*
b. *This first-personal self-reference is essential or irreducible to a third-personal one salva veritate.*

This quite peculiar way in which every person is given to herself, and to no other person, is familiar enough from the Cartesian *cogito*, a kind of reflection explicitly devised to "bracket" any other source of reference to oneself than first-personal self-reference. Descartes' case is even stronger: I could suffer not only from amnesia, but, worse yet, be completely wrong in my beliefs about any state of affairs whatsoever in the world (hyperbolic doubt). And yet I cannot doubt that *I* exist. Such evidence is the upshot of what Baker calls a robust first-person perspective.

[3] J. Perry 1979.

Let us call such Cartesian-style self-reference *transparent* and *absolute*. By calling it "transparent" I mean to say that it is immune from the misidentification error, and by "absolute" I mean that it is unqualified, free from any description or conceptual specification.

This is part of what "having a self-concept" amounts to according to Baker: "A self-concept is a 'formal' (non-qualitative) concept. Its role is to self-attribute a first-person reference – in such a way that the user of a self concept *cannot be mistaken* about who she is referring to" (2013, p. 137). So, a capacity for Cartesian self-reference is at least *part* of a robust first-person perspective. In fact, Baker writes, "a self-concept is constitutive of a robust first-person perspective" (2013, p. 137).

Notice that I am not making any claim about the way in which one acquires a self-concept (one probably cannot obtain it without a body and a common, acquired language). I am simply agreeing with Baker that having a capacity for Cartesian self-reference or a reflective *cogito* is at least a necessary condition for personhood, and hence a property which cannot be eliminated from an adequate ontology. If reductive naturalism entails that it can, then that view is false.

So, a phenomenologist could go along with Descartes and Baker up to this point.

But how far down this road can we follow Descartes? Not very far, I contend. For the case of the Borgesian Library can be read the other way round. Granted, it is only because I can enjoy independent Cartesian self-reference that I may discover that *I am* that author. But now suppose that I never figure out that I am that author, RDM.

Well, in this case, knowledge of myself will be severely incomplete – but that is not very noteworthy, since our knowledge of ourselves is already very incomplete, as is our knowledge of anything real. It is a familiar phenomenological tenet that whatever is real is an infinite source of information, and that knowledge of it is forever inadequate, forever partial.

The relevant point is different. If I am RDM and I do not know that I am RDM, then I literally do not know *who I am* or, even worse, I *have a false belief* about my identity. I believe that I am not RDM.

So in this instance, I not only miss a lot of relevant information about myself, but I am actually mistaken about myself. I incur in a misidentification error.

This fact proves that *there is more* to having a first-person perspective than a capacity for Cartesian self-reference. What more might there be? Well, purely Cartesian self-reference does not tell *whose self* the referred-to self is.

In so far as it is transparent and absolute, it picks out a homeless self, so to speak. In so far as it picks out a particular person, this one, which I fail to recognize as being in fact identical to RDM, it is no longer transparent. My demonstrative or indexical reference to myself here, this person suffering from amnesia, unexpectedly does *not* refer to the person to whom I mean it or I believe it to refer. *My self-concept does not refer to myself as the particular person I in fact am, even if it refers to myself as myself.*

This is a puzzling situation. Let us call it *the crucial puzzle of personal identity* – the one previously mentioned. I think that Baker's haecceitism is a very brilliant response – and even solution – to the apparent paradox involved.

It is a solution delivering us from any heritage of Cartesian immaterialism and/or internalism. For that reason, I do not accept the claim of those critics who take the self of the self-concept to be a purely intentional or merely mental object (as Johnston 2010 does). Baker is extremely clear on this point: "I suggest that we dissociate the idea of the first-person perspective from the Cartesian ideas of transparency, infallibility, and logical privacy" (2013, p. 140), and haecceitism, as we shall see in a moment, supports this statement by pinning the referred-to self in each instance to the particular person in the world to which it belongs.

But this solution, as I anticipated already, is not (entirely) true to the sense of the puzzle and, ultimately, to the intuitive, pre-philosophical sense of the problem of personal identity.

Two further steps now remain: 1) explicating Baker's solution in greater detail, and 2) discerning why it is untrue to life and what different view could give life and the basic question as it arises pre-theoretically their due.

4. Modern Haecceitism

Recall Baker's "core problem". How can a third-personal, exhaustive description of the world leave room for the further fact that I am one of the individuals in it? It cannot, according to the main argument. Yet, if the main argument is based on the irreducibility of Cartesian self-reference, then one has to meet the objection that Cartesian self-reference has no individuating power, at least if each of us is an embodied person. Purely Cartesian self-reference is utterly uninformative about just *whose person* it is supposed to pick out.

Hence, answering this objection is crucial to Baker's personal ontology. That is precisely the aim of what she calls the "Haecceitistic implications" of her ontology (2013, p. 179). What follows is Baker's answer, which we shall present by splitting her Identity Condition for personhood (BIC) into a Specific Identity Component (SIC) and a Numeric Identity Component, which is in fact the Individual Identity Condition, specifying the identity condition of a particular person (PIC).
Given that having a first-person perspective is a necessary condition for being a person, we may define a person as the exemplifier of a first-person perspective, which will be designated as 'F':

> (SIC) x is a person if and only if x exemplifies F essentially.

This Specific Identity Condition of a person "opens up room for a distinction between being a person and being me" (2013, p. 179).
We must now specify the Numeric Identity Condition of a person, the one picking out this particular person, *me* for example, or *you*. We have to define the individuating difference that "constrains" a person to be this person, me. And this individuating difference is a very simple property: *being the same as me*.

The intuition here is the same one that revived Scotus' term "haecceity" in contemporary modal-ontological debate, as opened by widely-read essays such as Plantinga (1974), Adams (1979), and others. Consider a relevant passage from Adams:

> A thisness is the property of being identical with a certain particular individual – not the property that we all share, of being identical with some individual or other, but my property of being identical with me, your property of being identical with you, etc. These properties have recently been called 'essences', but that is historically unfortunate; for essences have normally been understood to be constituted by qualitative properties, and we are entertaining the possibility of nonqualitative thisnesses (Adams 1979, p. 6).

Echoing such an understanding of haecceity, Baker explains: "Haecceity, roughly, is 'thisness', a nonqualitative property responsible for individuation. I want [...] to take an haecceity to be the state of affairs of someone exemplifying a property" (2013, p. 180), and "a haecceity does not add to the 'whatness' of a thing but distinguishes it from other things of the same kind" (2013, pp. 180-181).
In fact, on this conception, haecceity is a property that bears reference to an independently given individual. It specifies the identity condition for being a (particular) person (PIC):

> (PIC) A person y is a particular person x iff y has the haecceity of x (*i.e.*, the property of being identical to x).

In fact, all that claim really amounts to is that a person is me iff this person has my haecceity, that is, is identical to me.
As a property defining the condition for being me, the "property of being identical with me" seems circular. But the fact that PIC is "blatantly circular", says Baker, is no objection: "Circularity follows from the nature of the case" (2013, p. 180).

Haecceity provides Baker's decisive solution to the core problem, captured by the question: How can we understand the fact that a particular person in this world *is me*? What exactly is the condition under which, of all persons now living in the world, at least one and only one, call her RDM, is *me*? Well, RDM must share my haecceity.

To make the point more formally, we must recall (SIC) and (PIC). That RDM is a person means that there is an x such that

> a) x exemplifies F essentially.

This is a Specific Identity Condition valid for any x. Of course we must also specify the Numerical Identity Condition of that person we call "RDM" (there is at least and at most one RDM):

> b) (∃x) (y) [(x = RDM) AND IF (y = RDM) THEN (y = x)].

So finally the condition for RDM to be identical to some particular z, say, me rather than you, is the fact that RDM and z share the same Haecceity, z being the same as RDM:

> c) (∃z) (z = RDM).

To sum up, as Baker remarks, "We are now in a position to understand how my being LB is a fact. The key is that personal identity can be understood in terms of haecceity: x=y if and only if x and y have the same haecceity" (2013, p. 181).

5. Criticism

Well, what is wrong with all of that? Nothing is really wrong, as I said. Yet, modern Haecceitism is not true to life, that is, to the sense of the crucial puzzle of personal identity, and to the basic question underpinning it, Who am I?
When I start wondering about that, it is *not* because I run the risk of mistaking myself for you in the way that I might mistake you for your twin. It is because *there is much more* to my being this particular person than my self-concept affords. That is so even if we add to it all the properties I am aware of having. It is because self-*knowledge* infinitely transcends self-*consciousness* and self-*awareness*, or because each of us is to himself and to others an infinite source of information, like anything worthy of being called a real thing.
In this respect, this who-question has the same sense whether one asks it in the first or in the third person. It takes a life to acquire an even partial knowledge of another person, whereas it takes a few minutes to be acquainted with her or to be able to tell her from someone else.
Raised in the first-person, the who-question has one more peculiarity, namely, that "Cartesian" self-reference which can deceive us into the illusion of being self-transparent. The latter point underpins the crucial puzzle that, although I refer to myself as myself, I can be mistaken about whom I am.
First-person research into self-knowledge is – as our entire literary, religious and philosophical tradition testifies – a serious cognitive adventure, an exploration permitting genuine discoveries. The question "Who am I?" expresses in any case a true desire for further substantive knowledge, further information about the whatness – the *individual* nature or essence of myself. This desire could not possibly be satisfied by answering "You are you", "You are RDM", or even "You are this person here, not that one there". Think of Ulysses, think of Oedipus, think of Dante's wayfarer or of Faust, think of

Macbeth, of King Lear...

A metaphysical theory of personal identity, of course, could not aim at yielding the kind of individual knowledge that the basic question "Who am I?" – or "Who is this person?" – is striving for. Yet a metaphysical theory of personal identity that seeks to be true to life should account for the *meaning* of the basic question, of *that* meaning, actually, that implies a desire for further substantive knowledge. How can such a question arise? What is there in the being of a person – *any* person – that motivates such a question?

One might object that there is another way to understand question "Who is Lynne Baker?". Perhaps it means "Which one of the speakers is Lynne?". Perhaps so, but if this were the only reading, there would be no need for a distinct interrogative *personal* pronoun. Asking which one Lynne is would be just like asking which one of these seats is mine. In fact, if all we can ask for is the *distinctive* feature, or the individual *difference*, of a material particular, then no qualitative and intrinsic feature is relevant, no *content of the person*, so to speak. The circumstances of existence (e.g., the space and time in which a thing exists), as typically registered for persons (e.g., in one's passport or ID card), are quite enough. We can also give a distinctive extrinsic mark to any object, similarly to how we assign a number to each seat in a row.

In fact, why should there be a relative or interrogative *personal* pronoun at all, if that understanding were the only possible one? But it is not. There is another reading, for which "Which one of a plurality of persons is Lynne?" is no synonym (nor is "Which *kind* of person is she?"). *It is a conception on which asking "Who is Lynne?" would make sense even if Lynne were the only person left in the world after a catastrophe*. This question would not inquire after which property picks out Lynne Baker instead of some other person, but would inquire into the inexhaustible, *partially quite visible, but mostly neither visible nor evident individual whatness* of Lynne (in her "*ultima solitudo*", as Scotus would say). It would look for Lynne's *individuality* – or *individual essence*. Ordinary language calls it her *personality*.

Many will object: Aha, that is it, *personality* is a psychological, not a metaphysical notion!

I do not think so. Take an instance of personality: Socrateity. There is nothing psychological to the question: Who is Socrates? The Socrates of Plato's dialogues, the one of Xenophon, the one of Aristophanes? We know *which one of the Athenians* of his generation he was, we have all the information that an identity card might contain in terms of the circumstances of his existence. And yet we still debate *who he really was*. Even if we could never know it, is not there a truth of the matter? If you think there is, you *need* a metaphysics of individuality, if only to argue against post-modern narrative theories of personal identity, according to which – as for the naturalists – there is no truth of the matter, but only a socially negotiated narrative.

So, we need a theory of the individual whatness of a person – of its individual *nature* or essence. A theory telling us what individualizes Socrates' animality and rationality, *i.e.* the common nature he shares with Plato and with Lynne Baker. Here we are indeed looking for something informative, "adding" to the otherwise common whatness of Socrates.

Is Modern Haecceitism such a theory? I think it is not. For Socrates' thisness – the property of being identical to Socrates – "does not add to the whatness of a thing". It is a non-qualitatively differentiating property.

But is it a reasonable request to ask for such a theory? What has metaphysics to do with a person's personality? Is not that a matter of empirical research?

Of course, the question about Socrates is a matter of historical research. But what makes such research possible is that persons do have an *individual* whatness, an individual *nature* – a *personality. That persons have personality seems to me to be as essential to their personhood as is their having a capacity for a robust first-person perspective.*

This would lead us to "add" something to the Specific Condition of Personhood (BIC):

(R1) Anything having personhood has personality (an *individual* nature).

So, what (R1) says is that the individuality of a person is not merely due to that person's instantiating some property. True enough, material particulars are individuals just in that sense. But there is something more to the individuality of a person. Let us call it *personality*.

Surely having personality *is not* the property of being identical to me, or to you, or to some other person, as modern haecceity. For personality *does* "add to the common nature" or the whatness. How? Is such a metaphysical notion – call it an *individual nature* – not empty or vain?

I do not think so, and explicating what essential individuality must contain will yield an outline of an *alternative theory of individuality*, or of an *alternative principle of individuation*.

6. Material Adequacy Conditions of an Alternative Theory

We can identify three sets of contents making up personal individuality.

First of all, such an individuality must surely include all the *circumstances of the existence* of a given person, such as the origin, time, and places of her existence, and thus all the *contingencies of her being*. For there is an inescapable and dramatic link between individuality and contingency. This link is not the whole story, but certainly part of it. Lynne Baker's personality is not really separable from her origin, the circumstances of her birth (*those* parents, and so on), the time and place of her life (including of course nationality), language, education, etc. And these facts are definitely contingent, as contingent as the accident of birth on which they all depend.

Secondly, we must include all a person's *modes of appearance*, chiefly, one's personal physiognomy (in the broad sense including *bodily and dynamic personality*). I take Lynne's visage, way of speaking, and even of walking to be features essentially belonging to Lynneity, along with her intellectual and moral physiognomy, her style of behaviour, her way of thinking, and the like.

Of course, this third class of contents – intellectual and moral personality, a part of which may be manifested in books or personal choices, while other parts may not (or not yet) be – is the first one we tend to think of as being constitutive of Lynne's whatness.

Now, take the first class and the third class of features. The former are on Lynne's passport. Let us call them Lynne's extrinsic properties (accidents). The latter would comprise the bulk of an ideal portrait of Lynne, like a monograph on her as an author, setting aside the biographical data. Let us call them Lynne's intrinsic properties (like her beliefs, character traits, etc.).

In a way, these two sorts of information are linked by the photograph on the passport. Clearly, they are logically independent. *That* person with *that* physiognomy could conceivably have a completely different moral and intellectual personality.

And yet we feel that they must be somehow connected in the thing itself. How?

There is a relation of *ontological dependence* between circumstances of existence (non-qualitative properties) and the whatness of a thing, the set of its qualitative properties.

Intuitively, this relation is obvious in our paradigm case of essential individuality, that of persons. Of course, a human person does *not* merely exemplify humanity, without any further qualification, as the tiles of a roof exemplify the colour red. Each person literally *personalizes humanity*. Each one *enacts* this common nature differently. Each one not only "instantiates" it, but also "substantiates" it. By substantiating it, she individualizes it in all aspects, from her way of walking to her way of loving. She enacts human nature by all her acts, in such a way that her individual physiognomy is easily discerned.

Doubtless, me and this cup in front of me are alike in so far as our existence is contingent. We are both contingent instances of our specific natures. But while the circumstances of the existence of this cup remain accidental to it, mine become part of my whatness, and hence *essential* to me, to my *nature*. They add to it or further qualify it. The accident of birth stops being accidental to a person. This is what living as a person *is*. *This is the individualizing nature of a person*.

What we need in our ontology to do justice to this intuition is therefore *a being capable of transforming contingency* (its accidental circumstances) *into individual essence* (its whatness or nature) – to internalize contingency, so to speak. Once they become *part of* such an individual, accidental circumstances are absorbed within the foundations of that individual's possible futures. This yields some more or less equivalent definitions of personhood, corollaries of (R1):

> (R2) A person is a producer of essence out of accidents;
> (R3) A person is a machine that incorporates existence into its individual essence;
> (R4) A person is a transformer of the accident of birth into a destiny.

7. Towards an Alternative Theory of Personal Identity (and Haecceity)

Suppose our informal, phenomenological intuitions are plausible. Now, what would my formal substitute for Baker's theory of personal identity in terms of Modern Haecceitism be?
Conceiving of haecceity as the property of being identical to one particular person may be perfectly compatible with all these intuitions. Nevertheless, that falls short of a formal expression of the difference between having and not having what I have called an individual nature or personality. So, informally, if somebody were to ask me, "Is Kate's individuality specified by the property of being the same thing as Kate?" (*i.e.*, claim of Modern Haecceitism), I would reply in the negative. For this condition is not a plausible desideratum intimated in the basic question, Who is Kate? Crucial to responding adequately to this basic question is discerning what Kate has done with the circumstances of her existence and how she has become the person she is.
To advance towards a formal rendering of this intuition, we have to embark on a general *conceptual clarification of what an individual nature is*. An individual nature is that by which something is an individual. What, then, is that? It is *a kind of unity*. Let us call it the *unity of containment*. Scotus famously uses the phrase "less than numerical unity"[4]. This is the unity of a common nature, for example, that of personhood. Now what makes personality out of personhood, numerical unity out of a "less than numerical" unity, is the unity of containment. Scotus also calls the latter *substantial* unity, as opposed to *accidental* unity. The intuitive idea is that there is something "keeping together" all the different aspects of Lynne's existence, and this unity becomes apparent (a phenomenon) in so far as a person appears to us as a structured whole (as opposed to a mere sum of "parts").
Now, this containing unity or *ultimate unity of containment* is what I take to be the individuating principle of persons. Moreover, I take this to be their haecceity. So, how might we represent it more formally? On what condition does one have such unity?
Another great metaphysician of individuality – Leibniz, probably after reading Scotus – came upon this same conception of haecceity. The idea, expressed in a more Leibnizian way, is that a genuine individual is such that it possesses all of its properties, whether necessary or contingent, essentially. I could not be anywhere else than here, now – without being a different person. This doctrine is often called "superessentialism".
One may be inclined to object that if this were true, I could not survive a haircut.
One will be so inclined if one thinks of essential properties in terms of logical necessity *de re*, that is, a truth that holds in all possible worlds in which the thing exists. But, in fact, the only essential property that is logically necessary *de re* for any individual whatsoever is *being the same as that individual*, or, in short, Modern Haecceity.
If such were the case with any of my properties, I could not, indeed, survive a haircut. Thank God, that is not the case. Superessentialism means that I can survive a wide range of hairstyles, *but within the range* of what *my* hair can sustain.

[4] "Minus quam numerica unitas", cf. Duns Scotus (1973, *passim*, pp. 391-410).

This gives us a clue for how to define haecceity in more formal terms. Haecceity is not a simple non-qualitative property, as Modern Haecceitism would have it. Haecceity is an essence, namely an individual essence. A given essence (e.g, personhood) is a *constraint on possible (co)variations of properties*. If an entity is not so constrained, it fails to exemplify that essence (e.g., to be a human person).
An *individual* essence or unity of containment (like personality) is a constraint on the possible (co) variations of individualized properties a person may possess while remaining that same individual. We can formally represent this in modal terms. The question is, Within what limits can this person's intrinsic and extrinsic properties co-vary such that she survives those changes in her properties? In fact, I have ruled out both logical equivalence and mere factual conjunction of extrinsic and intrinsic properties. I intimated above that there is a relation of *ontological dependence* between circumstances of existence (non-qualitative properties) and the whatness of a thing, the set of its qualitative properties. Ontological dependence is a relation of "necessity" that is less than logical but more than accidental.
So, for example, suppose that it is true that Lynne could have been born and brought up in Japan instead of in the States. Let us suppose such an alternative course of events is conceivable. Nevertheless, for Lynne as she is *now*, for all the contents of her actual being, it is essential for her to have been born and educated in the States and not in Japan. To be specific, her unity of containment could not possibly hold together being such a distinct American philosopher and speaking Japanese as her only language.
This is how superessentialism works for persons, that is, for producers of essence out of accidental circumstances (recall corollaries R2-R4). The accident of birth, and all contingencies bound up with it, are in a way "swallowed up" by the person's being – they become essential to it.
But if superessentialism holds for things having an individual nature, ultimate unity of containment, or haecceity, that yields our desired formal characterization. Such entities are unique, that is, they satisfy Leibniz's principle of the identity of indiscernibles. Consider the following proposal, let us call it SH (short for "Scotistic Haecceity"), which is, I propose, the genuinely "Scotistic" notion of haecceity:

> (SH) An individual x has a substantial or ultimate unity of containment
>
> IFF:
>
> a) For all F, x, y : $[(Fx \leftrightarrow Fy) \rightarrow x = y]$.

How is this uniqueness to be understood? For Leibniz, it is a metaphysical principle, defining true individuality. It is therefore a necessary truth. But "necessary" in what sense? Can this uniqueness or property of not having indiscernible copies be enjoyed by an individual thing *in all* possible worlds in which it exists? Hardly so. We can always imagine counterexamples along the lines of P.F. Strawson's chessboard-like world (1964, p. 125) where two symmetrical cases are indiscernible and yet remain numerically distinct. On the other hand, including the property of being the same as x (Modern Haecceity) among the properties F would trivialize the principle. Hence, we need another clause preventing SH from collapsing into Modern Haecceity.
Such a collapse would result, quite clearly, in uniqueness being reduced to mere numerical unity. In fact, numerical unity is to uniqueness what "accidental unity" is to "substantial unity" (ultimate unity of containment or Scotistic Haecceity).
Many things have only accidental unity. In fact, the accident of birth and its circumstances make whatever is born (in the broad sense of having a temporal origin) unique in some sense. But this

uniqueness is *accidental* for most kinds of entities. Two bacteria can (in principle) be perfect duplicates, even if each of them is unique in the weak sense of originating at a different point in space-time.

On the other hand, no creature in time can be necessarily unique, if this means *logical* necessity. As a medieval thinker would say: *omne ens est unum*, but only a necessary being is necessarily one in number, or unique. Only God, if God exists, exists necessarily, and only God, if God exists, is necessarily unique. But we exist contingently, if we exist.

We human persons lie somewhere between the bacteria and God (if God exists). How then should we characterize Scotistic Haecceity in more adequate details?

Let us call *conditionally necessary uniqueness* the limitation that makes us different from God. Here "conditionally necessary" refers to the sort of uniqueness that is compatible with and even conditioned by an entity's contingent existence.

So the required enrichment of our characterization of Scotistic Haecceity must capture this idea of conditional uniqueness. We are necessarily unique under the condition of a fatal accident – the accident of our birth.

This amounts to positing a restriction on the modal truth of clause a) of my previous formulation of SH, which as it stands says no more than the principle of the identity of indiscernibles. Clause a) must be true not just in the actual world, not in all the possible worlds either, but only in those worlds which are temporally accessible from the actual world. That is, only in the present and in its future worlds.

Let it be granted that L and M are modal operators for necessity and possibility, respectively, and that most ordinary modal principles hold. We shall obtain the required restriction by *negating* a principle valid within S5, the modal system in which the accessibility relation is an equivalence relation. That excludes from consideration *any* world's being accessible from any other as well as the standpoint of an unconditional, necessary being not bound to space and time. Our second clause will thus restrict the validity of clause a) in such a way that a) is necessarily true only in the sense of conditional necessity:

b) NOT $Lp \rightarrow LLp$ (whatever is necessary, is only conditionally necessary)

NOT $Pp \rightarrow LPp$ (whatever is possible, is only conditionally possible)[5].

This restriction is intended to capture the temporal character of that ultimate unity of containment which is the individual essence of persons, *i.e.*, their personality.

Clauses a) and b) indicate the pertinent elements of a formal presentation (which cannot be carried out within this paper) of my conception of haecceity, a more genuinely Scotistic one, I have suggested, and an alternative to Modern Haecceity.

Suppose that a) and b) help to clarify the basic intuition concerning the individual essence of persons, or personality. That would fulfil the "more ambitious task" of addressing *first* the very nature of personal individuality in order *then* to solve, on its basis, the narrower problem of *personal identity across time*, thereby yielding a non-circular condition for the temporal persistence for persons.

Here, perhaps unsurprisingly, is my suggestion:

A person x at time t_1 is the same person as a person y at time t_2

IFF:

$(\exists x)\ (y)\ [SH(x)\ \text{AND}\ (\text{IF}\ SH(y)\ \text{THEN}\ (y = x))]$.

[5] This double clause b) is true in the Modal System S4, where the accessibility relations among possible worlds is reflexive and transitive, but not symmetric (Hughes & Cresswell 1996).

The idea here is that personal identity across time consists in sharing Scotistic Haecceity or substantial unity. This does not prevent a person from changing, but allows for just those changes that preserve a person's non-accidental unity. Temporal identity is, to put it phenomenologically, change constrained by a consistent global style. Max Scheler had an apt expression for what it is to be identical across time: *Anderswerden*.

8. Informal Conclusions

So what about my haircut? Of course I can survive it. And yet, that is so only because the haircut is within that *bond of possible variations of each one of the properties* admitted by my haecceity. And this is exactly what it means for each property to be essential to me. No property can vary independently of the changing whole which the property is a part of. Possible (co)variations *are different for each individual*. They depend on one's accidental circumstances, as well as on one's freedom. The sum of those constraints constitute one's personality. Or, better yet, in holding to those constraints, the very you-ness of you becomes manifest.

I think that the Latin word *haecceitas* expresses, in Scotus' use of it, the idea of a relation between the *specific* nature of one's personhood (*i.e., the necessary property of a person* qua *person,* her primary kind), and the accidents of one's birth and life (*i.e., contingent circumstances of a person's existence*). The latter are not essential to a possible person, but become essential to the actual person once she is born and has carried on in just the way she has. They are, as it were, swallowed up by the being of that person, becoming "one thing" with her. This word, "*haecceitas*", calls to mind a most dramatic indexical scene of Christianity and that simple utterance, "*Ecce homo*".

"*Ecce*". Here you are. Here and now, with your unique visage and body, with your own singular and novel human destiny, the kind that every person, every individualizer of humanity brings to existence.

REFERENCES

Adams, R.M. (1979), "Primitive Thisness and Primitive Identity", *The Journal of Philosophy,* 76(1), pp. 5-26;

Baker, L.R. (2007), *The Metaphysics of Everyday Life*, Oxford University Press, New York;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, New York;

Duns Scotus, (1973), *Ordinatio*, II, dist. 3, pars 1, in *Iohanni Duns Scoti opera omnia* tom. VII, studio et cura Commisionis Scotisticae (ad fidem codicum edita), praeside P. Carolo Bali´c, Civitatis Vaticana: Typis Polyglottis Vaticanis, 1973, pp. 391-410;

Hughes, G.E. & Cresswell, J. (1996), *A New Introduction to Modal Logic*, Routledge, London;

Johnston, M. (2010), *Surviving Death*, Princeton University Press, Princeton (N.J.);

Lewis, D. (1983), "An Argument for the Identity Theory", in *Philosophical Papers*, vol. I, pp. 55-77, Oxford University Press, New York;

Noonan, H. (2003), *Personal Identity*, Second Edition, Routledge, London;

Parfit, D. (1971), "Personal Identity", *Philosophical Review*, 80, pp. 3-27;

Parfit, D. (1984), *Reasons and Persons*, Clarendon Press, Oxford;

Perry, J. (1979), "The Problem of the Essential Indexical", *Nous,* 13, pp. 3-21;

Plantinga, A. (1974), *The Nature of Necessity*, Oxford University Press, Oxford;

Ricoeur, P. (1990), *Soi meme comme un autre*, Editions du Seuil, Paris;

Shoemaker, S. & Swinburne, R. (1984), *Personal Identity*, Blackwell, Oxford;

Strawson, P.F. (1964), *Individuals: an Essay in Descriptive Metaphysics*, Methuen, London;

Zahavi, D. (2005), *Subjectivity and Selfhood – Investigating the First-Person Perspective*, MIT Press, Cambridge Massachusetts.

MICHELE DI FRANCESCO
Istituto Universitario di Studi Superiori, Pavia
michele.difrancesco@iusspavia.it

MASSIMO MARRAFFA
Università Roma Tre
massimo.marraffa@uniroma3.it

ALFREDO PATERNOSTER
Università di Bergamo
alfredo.paternoster@unibg.it

REAL SELVES? SUBJECTIVITY AND THE SUBPERSONAL MIND

abstract

The current philosophical discussion on the self and consciousness is characterized by a contrast or dilemma between the no-self (eliminativist) perspective, on the one hand, and the arguably naïve account that takes the self as a robust entity, on the other. In order to solve the dilemma, in this paper we suggest restoring a robust theory of the subject based on a bottom-up approach (fully consonant with contemporary neurocognitive science) together with a pluralistic reading of the nature of the science of the mental.

keywords

Self, bottom-up approach, explanatory pluralism, Dostoevskian Machine

1. Introduction and Overview

This paper was originally presented in a workshop addressing what was described as "Lynne Baker's Challenge", that is the thesis that human persons are entities essentially characterized by the possession of a *robust* first-person perspective (a thesis fully articulated by Baker in her recent book, *Naturalism and the First Person Perspective*, 2013). Differently from Baker's and many other talks presented in the workshop, the present contribution does not deal *directly* with the first-person perspective and its metaphysical implications. Rather it stems from the philosophical reflection on neurocognitive studies of subjectivity, and is more interested in *epistemological* and *explanatory* issues than in metaphysical conundrums. Yet it is fully congruent, we think, with Baker's appreciation of the importance of the relation between personal and subpersonal levels of explanation, as expressed, for example, in the following passage:

> Our ability to conceive ourselves as ourselves*[1] is a personal-level capacity. Why does it resist being reduced to or replaced by subpersonal phenomena? If I am right about the robust first-person perspective, then we have an answer to this methodological question: the personal level of reality – the level on which we live and love – is neither eliminable nor reducible to subpersonal levels that supply the mechanisms that make it possible for us to live and love (Baker 2014, p. 333).

We agree with Baker that the relation between personal and subpersonal "levels of reality" raises fundamental philosophical questions, and, among these, the problem of developing a theory of the nature of the self-conscious rational agent congruent with contemporary scientific research is one of the most prominent. We also take very seriously the "methodological" question addressed by Baker in the passage quoted above: why does our ability to conceive ourselves as ourselves* resist being reduced to or replaced by subpersonal phenomena? Indeed, in this paper we try to offer an answer to it; yet, differently from Baker's, our answer is based on a pluralistic reading of the nature of the science of the mental (which, as we shall see, involves a form of explanatory pluralism), rather than on a specific thesis about the metaphysical underpinnings of the first-person perspective.
In particular in our paper we argue that a *robust* account of the self – *i.e.*, of the subject of experience

[1] The star following the second token of "ourselves" indicates a reference to the first person as a first person subject. You cannot substitute it salva veritate with a co-referential expression, such as the name of that person.

– is not only possible, *contra* the eliminativist-style arguments, but also fully consonant with contemporary (neuro)cognitive science. The paper is organized as follows.

In the first section we show that the contemporary science of the mind privileges a bottom-up approach to self-consciousness, based on the notion of cognitive, or computational, unconscious. In the second section we note that, in this context, the self-conscious rational agent is often presented as an illusion. A virtual space of presence, or a center of narrative gravity, is reconstructed as the owner of the stream of consciousness, but is in fact causally inert. In the third section we argue that a robust self is needed to explain the kind of intentional action and self-understanding presupposed by both commonsense psychology and social science. The problem we are faced to can then be presented in the form of a dilemma between the *no-self* (eliminativist) perspective, on the one hand, and the arguably *naïve* account that takes the self as a robust entity, on the other. In order to solve the dilemma, we suggest restoring a robust theory of the subject based on a bottom-up approach together with a pluralistic reading of the nature of the science of the mental. Also, we give some reasons to believe that this robust theory of the self is fully consonant with contemporary (neuro)cognitive science. Finally, in the fourth section, we compare our strategy with Baker's anti-eliminativist approach.

Before going on, we have to introduce a terminological *caveat*. For simplicity's sake, we use the word "self" in a loose way, to refer both to the subject of experience and to the self-representation of oneself that makes an individual a subject of experience. In other words, we do not use explicitly the distinction between "being a self" and "having a self". In fact we share Baker's doubts (or at least prudence) about the concept of self, and we consider this notion not as a primitive, but as a part of a theory of self-consciousness – which is the focus of our research.

2. The Cognitive Unconscious

In the last fifty years, the sciences of the mind have been mostly concerned with unconscious functions. Indeed, cognitive processes studied by cognitive science, such as perception, reasoning or language understanding, are not accessible to consciousness. Only their inputs and outputs (and perhaps some of their fragmentary parts) can be accessed. We are aware of the final results of the processes, not of their internal dynamics. In this perspective, the unconscious is in a way much more important than the conscious, insofar as it is the unconscious that explains the abilities manifested in our behavior.

Let us consider, for example, the case of language. Our understanding of a sentence is immediate. We instantly know whether or not we have grasped (as usually happens) what our interlocutor is telling us. Notwithstanding, a lot of machinery is needed to understand a sentence: a nearly continuous sequence of sounds must be segmented into words, *i.e.*, into meaningful units; a grammatical structure must be associated to the sentence, and not always this structure is the only possible one (in which case one needs to choose the right one); ambiguous or polysemous words are to be interpreted in a manner appropriate to the context, etc. We have no awareness of all these complicated processes, just as we have no awareness of the structures of information – the representations – that must be built up to successfully perform these tasks. We are not conscious of having grammar rules inside our heads and of systematically applying them during the processes of understanding.

The *cognitive* or *computational* unconscious, then, is a level of analysis that is fundamentally subpersonal: the information-processing level, wedged between the personal sphere of first-person phenomenology and the nonpersonal domain of neurobiological events. Such level no longer takes consciousness as something that explains, but rather as something that needs to be explained, analyzed, sometimes even dismantled.

In asking how consciousness, rather than the unconscious, is possible, the cognitive scientist fully endorses Darwin's methodological approach, which, assuming the continuity between animal and human minds, pursues the study of consciousness by virtue of a bottom-up strategy. One begins with what is more simple, primitive, less structured, to reach what is complex, evolutionarily late, structured, without idealistically taking for granted the existence of a self-conscious self grounding the entire mental life. This self is rather the result of a process of construction that starts with subpersonal unconscious processes.

3. Dennett's Eliminativist Account of the Self

From the premise that the nature of the self is non-primary and derivative, many philosophers infer the conclusion that the self is an illusory by-product of the real neurobiological events, and is devoid of any explanatory role (think, for instance, of Dennett, Metzinger, analytical Buddhism). Let us focus on the case of Dennett, arguably the most influential one.

In light of a large amount of data from neurocognitive sciences, Dennett (1991) famously rejects the hypothesis that there is, in some area of the brain, a place where "it all comes together" (Dennett 1991, p. 107) – some sort of central executive system that coordinates all the cognitive operations – and stigmatizes it as "the myth of Cartesian Theater" (Dennett 1991, ch. 5). To this myth Dennett opposes the Multiple Drafts model of consciousness, according to which, at any instant, in any part of the brain, a multitude of "fixations of content" occur. The conscious character of these contents cannot be explained by their occurring in a *special* spatial or functional place (*i.e.*, the "Cartesian Theater"), nor by their having a special format. Rather, it depends on what Dennett (2005) calls "fame in the brain" or "cerebral celebrity" (Dennett 2005, p. 136). Like fame, consciousness is not an intrinsic property of the cerebral processes, but is more similar to "political clout", a kind of influence that determines the extent to which a content affects the future development of other contents distributed all over the brain.

On this eliminative view, a neuroscientific theory of consciousness must be a theory of how the illusion of the subject of consciousness arises (Dennett 1991, 2001, 2005). According to Dennett, an amazing property of *Homo Sapiens* is, precisely, the capacity to create a self: "out of its brain it spins a web of words and deeds" (1991, p. 416). By means of this activity, the biological organism produces a narrative, it posits a "center of narrative gravity". The narrative is the result of the working of a *Joycean Machine*: "In our brains there is a cobbled-together collection of specialist brain circuits, which, thanks to a family of habits inculcated partly by culture and partly by individual self-exploration, conspire together to produce a more or less orderly, more or less effective, more or less well-designed virtual machine" (1991, p. 228). The Joycean Machine is a *software in the brain* which creates the self, a *virtual captain*, a character described in internal and external discourse as the owner of the organism's mental states and as the actor of its actions and decisions, but who is in fact just a represented entity, not the real player in the game of human behavior.

Although Dennett's theory was developed in the early 1990s, recent empirical research is consonant with it. A neurocomputational architecture largely compatible with Dennett's Multiple Drafts Model is that of the Global Workspace Theory (GWT) of consciousness by Bernard Baars (1997). Recently, GWT has been developed in cognitive neuroscience, mainly thanks to Stanislas Dehaene and his collaborators' efforts (see, e.g., Dehaene & Naccache 2001; Dehaene *et al.* 2001). According to these researchers, there are two computational spaces within the brain, each characterized by a distinct pattern of connectivity.

The first space is a set of parallel, distributed, and functionally specialized processors or modular subsystems. These modular subsystems exploit highly specific local or medium-range connections that encapsulate information relevant to its function. The second space is a neuronal global workspace (and hence the theory is now termed "Global Neuronal Workspace Theory", GNWT) consisting of a distributed set of cortical neurons with long-distance connections, particularly dense in prefrontal, cingulate, and parietal regions, which are capable of interconnecting the multiple specialized processors and can broadcast signals at the brain scale in a spontaneous and sudden manner. This global neuronal workspace breaks the modularity of the nervous system and allows the broadcasting of information to multiple neural targets. This broadcasting creates a global availability that is experienced as consciousness and results in reportability.

At least three features of the GNWT are significant for Dennett (see Schneider 2007, p. 318). First, it assumes that the neurocognitive architecture underlying the unity of consciousness is a distributed computational system with no central controller. Second, it makes massive use of recursive functional decomposition, an indispensable requirement to get rid of any homunculus who, nestled in a sort of incarnation of the pineal gland, scans the stream of consciousness. Third, it allows Dennett to hypothesize that the aforementioned "political clout" is achieved by "reverberation" in a "sustained amplification loop" of the winning contents (Dennett 2005, pp. 135-136).

4. The Dilemma: to Self or not to Self?

This eliminativist conclusion about the self is not necessitated by contemporary cognitive science; but, apparently, it is fully consonant with it. Cognitive (and neurocognitive) science starts from the idea of the fruitfulness of a bottom-up approach. This approach does not appeal to our introspective self-knowledge, but to the results of investigations into the gradual construction of human self-awareness: from the automatic and pre-reflexive construction of representations of the external world, through the bodily self-monitoring, to self-consciousness as introspective recognition of the presence of an *inner*, experiential space. The outcome is a criticism of the primacy of self-conscious subjectivity, where the latter, far from being a primary datum, becomes an articulate construction out of several neurocognitive and psychosocial components.

Another result of this approach is the acknowledgement of the mixed and multi-faceted nature of the self: minimal, autobiographical, narrative and social selves appear to reflect different aspects and different stages of the interaction between the neurocognitive and psychosocial components. Despite our appearance of unity, we are, in a literal sense, the product of the fusion of a wide range of composite processes.

We are then faced with a dilemma. Either we give up the classical notion of rational self-conscious agent; or we reject the eliminativist doubts about the self, and find a way to restore a robust theory of the self. We opt for the second horn of the dilemma, and in what follows we show how is possible to maintain a robust theory of the self, without sacrificing, at the same time, the merits of the bottom-up strategy. In other words, we stay with Dennett and neuroscientists in endorsing the bottom-up approach, but we stand apart from them in defending a robust account of the self.

4.1. A Strategy to Deal with the Dilemma

First of all, let us note that, in sketching a robust theory of the self, one should avoid the risk of falling again in anti-naturalist positions. Authors such as Alasdair MacIntyre, Charles Taylor, and Paul Ricoeur see the self as a self-interpreting being in a sense inspired by the hermeneutic tradition (Schechtman 2011). However, hermeneutic tradition is hardly compatible with the bottom-up approach, which involves a commitment to naturalism. A hermeneutical notion of self-interpretation, with its emphasis on meaning at the expense of the psychobiological theme of the unconscious, runs the risk of surreptitiously reintroducing the idealistic conception of the conscious subject as primary subject, since the subjectivity suggested by the hermeneuticists is inevitably intentionalizing – rather than intentionalized by – the unconscious. By contrast, we suggest that the self-interpretation is a theory-driven activity of narrative re-appropriation of the products of the neurocognitive unconscious, quite similar, for instance, to the notion developed by Peter Carruthers in his Interpretive Sensory-Access model of self-knowledge (see Carruthers 2011)[2].

To put it briefly, the robust theory of the self must not be a restatement of a top-down view of self-conscious subjectivity as a datum (the view of the subject as an *a priori*).

[2] This does not mean that we buy Carruthers's theory of consciousness (and self-consciousness) across the board. The similarity concerns [just] the description of the activity of self-interpretation.

Somewhat surprisingly, Dennett sees narrativism and eliminativism about the self as the two sides of the same coin. In his view, the "I" is the useful fiction of a central controller, and its autobiography is a confabulatory by-product of the decentralized activity in the brain, which is actually responsible for the behavior. In other words, the Joycean Machine is anything but an *idle wheel* in the dynamical economy of the body (see Ismael 2006, p. 351). However, we dissent from the eliminativist argument that infers from the non-primary, derivative nature of the self a view of it as an epiphenomenal by-product of neurobiological events – or alternatively, of social (or socio-linguistic) practices – for at least two reasons. First, Dennett's self is said to be an abstraction devoid of real causal powers; yet, at the same time, this illusory character is a useful and even essential device: the complex social organisms that we are need a virtual self for their very survival, their social interactions, their decision making, and so on. Thus, one wonders why Dennett is not disposed to accord to the self a genuine causal role.

Second, let us concede that there is no central place in the brain where all information is gathered together, and no unifying superior function able to coordinate and organize what is processed by many different cognitive modules. This means that integration is not produced by top-down functions. However, this is not to say that the process of *ego* production leads to a pure nothing. We may argue on the contrary that the ability to represent herself as an enduring self does affect the very nature of the agent – making her an intentional subject of reason and action. To put it briefly, the inference from the existence of the *Multiple Drafts* to the *no-self* view is not justified.

Instead of Dennett's deflationist conception of the self, according to which the self is a mere abstraction, analogous to a non-existent, but useful, physical center of gravity, we suggest that there is an open alternative, a realist or somewhat *inflationist* position compatible with everything Dennett says about the architecture of human cognitive systems. According to such an inflationist option, the Joycean Machine is not a deceitful device but a cognitive mechanism that produces a reasonably stable and integrated (autobiographical) self, something that is best understood as the ongoing result of a narrative self-constructing process. Since the expression "Joycean Machine" may be perceived as intrinsically *eliminativistic*[3], we may substitute it with a new theoretical entity: the *Dostoevskian Machine*. The latter can be conceived as an integrated system of internal bottom-up mechanisms[4], which cooperate with external (social and environmental) factors to the process of self-building. A proper understanding of the working of the Dostoevskian Machine would reveal important aspects of human psychical dynamics, and would explain the processes that bring about the emergence of the kind of self-conscious experience that constitutes our autobiographical inner life and which shows itself, *inter alia*, in the use of self-referring linguistic expressions.

The reference to the Dostoevskian Machine allows us to save an important result of the eliminativist approach, namely, the acknowledgement of the mixed and multi-faceted nature of the self: minimal, autobiographical, narrative and social selves appear to reflect different aspects and different stages of the interaction between the neurocognitive and psychosocial components. Despite our appearance of unity, we are, in a literal sense, the product of the fusion of a wide range of composite processes.

To sum up, the bottom-up approach does not force us to endorse the eliminativist conclusions. In contemporary cognitive science there are theoretical tools which allow explaining conscious functions without assuming introspective self-knowledge as a datum, on the one hand, and maintaining a robust notion of the subject, on the other. We propose that self-conscious subjectivity, far from being a primary datum, is an articulate construction out of several neurocognitive and psychosocial components.

In other words, our central point is that the outcome of the Dostoevskian Machine, the product of the machinery in the head that composes the autobiography and controls verbal reports in the first person, *is responsible for stable, integrated and enduring aspects of human behavior.*

[3] Thanks to Michael Pauen for this comment.

[4] Here "bottom-up" means that these mechanisms are not based on any high-level, or full-blooded, representation of the self (this is in fact the *output* of the mechanism taken as a whole). Yet, some previous, relatively precocious, mental structures, such as bodily representations, feed the mechanism.

4.2. Two Reasons for a (Naturalistic) Robust Theory of the Self

There are at least two considerations that can be invoked in favor of our robust-*cum*-naturalistic view of the self. Both have to do with (or partly involve) *hot* aspects of the mind. Let us explore them in turn.

(*a*) The eliminativists disregard the fact that the process of narrative self-construction *includes an essential psychodynamic component*.

Breaking with a long philosophical tradition that has viewed self-consciousness as a purely cognitive phenomenon[5], the most important currents of dynamic psychology show that the construction of affectional bonds and the construction of identity cannot be separated. The description of the self that from 2-3 years of age the child feverishly pursues is an "accepting description", *i.e.*, a description that is indissolubly cognitive (as a *definition* of self) and emotional-affectional (as an *acceptance* of self). Briefly, the child needs a capacity to describe herself in a clear and consistent way, fully legitimized by the caregiver and socially valid. Also, this will continue to be the case during the entire cycle of life: the construction of an affectional life will always be intimately connected to the construction of a well-defined and interpersonally valid identity.

Accordingly, one cannot ascribe concreteness and solidity to one's own self-consciousness if the latter does not possess as a center a description of identity that must be clear and, indissolubly, "good" as worthy of being loved. Our mental balance rests on this feeling of solidly existing as an "I". If the self-description becomes uncertain (*i.e.*, inconsistent), the subject soon feels that her feeling of existing vanishes. This can be the result of a psychopathological process.

In patients with schizophrenia, for example, we can observe that the coherence of the representation of self is compromised or invalidated (see, e.g., Raffard *et al.* 2010), with a consequent loss of the capacity to clearly discriminate the borders between the inner space of the mind and the corporeal and extra-corporeal experiential spaces. The patient, then, develops abnormal defensive measures, aimed to head off the experiential chaos originating from the disintegration of the primary feeling of self.

Or let us consider the case of those patients whose main problem is a chronic feeling of insecurity (or lack of self-esteem, confidence in oneself, solidity of the *ego*, cohesion of the self – terms that we take to be essentially synonymous). According to a tradition in developmental psychopathology that begins with Michael Balint, Donald Winnicott and John Bowlby, the origin of this "basic fault" (Balint 1992) – or "primary ontological insecurity" (Laing 1960) – is to be traced back mainly to early deficiencies in the relationship between the child and the primary attachment figures (see, e.g., Fonagy, Gergely, Jurist & Target 2002). The child's attempts to rationalize the abusive or seriously neglective behaviors of the attachment figures may give rise to dysfunctional self-attributions, *i.e.*, to that *deficiency of identity* that can be found, for example, in patients suffering from narcissistic personality disorder. In some of these patients the feeling of identity is so precarious (the self is so little *cohesive*) that they find it difficult to feel existent and are afraid of completely losing contact with themselves if deprived of the link with situations, things or persons which serve as symbols that help to reassure them about their identity (Kohut 1977).

The waning of the existential feeling of presence may also occur in cases of sudden breakdown of self-esteem, or unexpected emotional upheavals, or when the continuity of the tissue of our sociality is broken, as can happen when one is suddenly thrown in some dehumanizing total institution (see, *e.g.*, the classic Goffman 1961). In such circumstances, the subject strives to cling to her memories, or to the sense of a projectual dignity, or to the secret security of an affiliation: "but if all these fail us, then we realize that our mind becomes empty, and not only we no longer know who we are, but also we literally lose the feeling of being present" (Jervis 2011, pp. 131-132).

[5] See, e.g., Bermúdez: "Self-consciousness is primarily a cognitive, rather than an affective state" (2007, p. 456).

To recapitulate. The conception of self-consciousness that emerges from the bottom-up exploration of the mind – including a dynamic psychology driven by cognitive sciences – is that of an interminable process of self-objectification by the human organism. This consciousness of the self is a description of the self, namely, identity. In its most advanced form, this is finding oneself at the center of one's own orderly and meaningful subjective world, and hence at the center of a historical and cultural environment to which one feels to belong. However, this full-blown self-consciousness is a construction without metaphysical guarantee and thus it is not a faculty guaranteed once for all, being rather a precarious acquisition, continuously constructed by the human organism and constantly exposed to the risk of dissolution (see Marraffa 2013, p. 109).

This precariousness is the key to grasp the defensive nature of the Dostoevskian self-narrative. The construction and protection of an identity that is *valid* as far as possible is something rooted in the organism's primary need to subjectively subsist, and thus to solidly exist as "I". Thus, far from being an epiphenomenal, transient phenomenon, a character in a fiction invented to facilitate the prediction of behavior without any real correlate (a short-lasting *virtual captain*), the incessant construction and reconstruction of a cohesive self – *i.e.*, of an acceptable and adaptively functioning identity – is the process through which our intra- and inter-personal balances are produced, hence the foundation of our mental health. So, in contrast to Dennett's Joycean monologue, the Dostoevskian self-narrative is not empty chatter at all: it is a *causal* center of gravity.

On this view, the onset of self-consciousness is the establishment of a process of self-description, *i.e.*, the self-representing of a system encompassing mechanisms that interact across social, individual/personal, and subpersonal levels of organization (see Synofzik, Vosgerau & Newen 2008; Herschbach 2012; Thagard 2014). The description of identity imposes a teleology (focused on self-defense) on the system.

(*b*) The second consideration that can be invoked in favor of our robust-*cum*-naturalistic view of the self is grounded on the fact that the self-narration produced by the Dostoevskian Machine is not at all contingent and evanescent, since it is firmly anchored on *personality structures*.

Here we have in mind recent theoretical systematizations in personality psychology, where we find that the ability to perceive one's own identity in terms of *narrative identity* stems at least from two cognitive layers: (*i*) traits of personality, largely determined by genetic factors and substantially stable through the life cycle; (*ii*) goals, plans, projects, values and other constructs – *i.e.*, *motivational and strategic* roles and contexts – that define the life of an individual. Narrative identity is then an internalized and evolving story of the self – layered over the person's dispositional traits and characteristic goals and motives – which can provide the jumble of autobiographical memories "with some semblance of unity, purpose, and meaning" (McAdams & Olson 2011, p. 527).

Thus the experimental investigations on the mechanisms underlying the construction of identity can be seen as psychological hypotheses about the functioning of the extended or robust Dostoevskian Machine (rather than an evanescent and transient Joycean Machine). And here the reference to the necessity of a multi-level explanation of the robust Dostoevskian Machine comes in. These explanations, indeed, are located at the intersection of several psychological disciplines: personality psychology, social psychology, developmental psychology, dynamical psychology – all potentially interacting with neurocognitive research.

It is worth to point out that this involvement of a collection of different disciplines suggests, or even implies, a view of the explanatory practices in the sciences of the mind that can be dubbed as "explanatory pluralism". Let us say a few words on this.

Explanatory pluralism is a position in the philosophy of science holding that "theories at different levels of description, like psychology and neuroscience, can co-evolve, and mutually influence each other, without the higher-level theory being replaced by, or reduced to, the lower-level one"

(Looren de Jong 2001, p. 731). The need of increasing the available explanatory resources is the main concern of the pluralist, who distances himself both from the reductionist obsession for ontological parsimony and unification of science, and from the claim for strong autonomy of the special sciences theorists. In particular, against the reductionist claim that when lower-level explanations are completed, the higher-level explanations stop being causally explanatory, explanatory pluralists deny the existence of a *fundamental* explanatory level, and argue that higher-level entities continue to play a causal and explanatory role even when lower-level explanations are complete (see Marraffa & Paternoster 2013).

This is not the place for a detailed analysis of explanatory pluralism and its relevance to certain crucial, foundational issues in cognitive science (see, e.g., McCauley & Bechtel 2001; Craver 2007; Marraffa & Paternoster 2013). It is enough to point out that, as the considerations made in this section should have shown, the problem of giving a comprehensive explanation of self-consciousness, covering all its different levels, can only be addressed by means of a multiplicity of theoretical resources, stemming from different disciplines. In this sense, we take the issue of the self as a case for explanatory pluralism.

5. Concluding Remarks

We started our analysis referring to Lynne Baker's theory of the first-person perspective and to the connected claim that the personal level of reality is neither eliminable nor reducible to the subpersonal level (Baker 2013, 2014). So it could be useful ending with a comparison between Baker's defence of the irreducibility of the first-person perspective and our approach to self-consciousness.

Firstly, we may note that there is a significant agreement on many issues. In particular:

- *We share a critical attitude towards reductive and eliminative accounts of mental phenomena, and in particular of the self.*
- *We share the idea that the personal level of description of human behavior, which characterizes the subject's mental life in terms of commonsense psychology, is essential and cannot be eliminated by direct reductive or eliminative moves.*

Besides, we believe that even if we grant a form of explicative supremacy of subpersonal psychology (an assumption that differentiates our position from Baker's), this does not entail the uselessness or the futility of personal psychology.

So there are convergences between the two theoretical projects, in particular if we consider our representation of the process of self-building as the product of what we called the "Dostoevskian Machine" (as opposed by Dennett's Joycean Machine).

We take the self as a system of subpersonal mechanisms that produces a real, causally efficacious, agent of psychical dynamics (and not a mere *virtual captain*), and this makes Baker's ontological view compatible with the kind of current empirical research we put at the basis of our proposal. This should not conceal the fact that Baker's overall metaphysics of the person is not the most favorable environment for a bottom-up approach to the subject, since it takes a person as a conscious substance (not an immaterial substance, but a substance which cannot be ontologically reduced to something else).

However, nothing in what we say forces *per se* a specific ontological conclusion. Non-reductive physicalism is compatible with the kind of explicative pluralism we endorse (in fact it is the standard view associated at the very beginning of cognitive science with explanatory pluralism). Yet, even some forms of metaphysical reductionism may be compatible: all depends on further and subtle issues concerning the metaphysics/epistemology divide.

If we were forced to express one ontological position that we find in accordance with our epistemological defence of the critical and fundamental role played by subpersonal explanation in psychology, we might quote David Lewis's seminal paper, "Attitudes *De Dicto* and *De Se*":

> I admit that knowledge *de dicto* is incomplete; but not that it is in any way misleading or distorted by its incompleteness. A map that is incomplete because the railways are left off is faulty indeed. By a misleading omission, it gives a distorted representation of the countryside. But if a map is made suitable for portable use by leaving off the "location of this map" dot, its incompleteness is not at all misleading. It cannot be said to misrepresent or distort the countryside at all, though indeed there is something that cannot be found out from it [...] An encyclopaedia that tells you where in logical space you are is none the worse for being neither signpost nor clock. Knowledge *de dicto* is not the whole of knowledge *de se*. But there is no contradiction, or conflict, or unbridgeable gap, or even tension, between knowledge *de dicto* and the rest. They fit together as nicely as you please (1979, p. 528; 1983, p. 144).

Adapting the quote to our analysis, we may say that the subpersonal description of human mind may be incomplete, but this does not mean that it is mistaken. We do not address this issue further, however.

We content ourselves with our attempt to show (1) that a more dialectical relationship between personal and subpersonal levels of psychological explanation is both possible and necessary to develop a theory of self-consciousness; (2) that a realist theory of the self offers an explanatory framework that is more useful to the understanding of self-consciousness than its eliminativistic anti-realist alternative; (3) that in the process of the construction of a theory of self-consciousness we need to wide our psychological horizon to take into consideration motivational and affective components that have been neglected by orthodox cognitive science, and (4) that this requires to widen our conceptual tools and suggests the adoption of epistemological pluralism[6].

[6] We are grateful to the participants of the Workshop "Naturalism, the First Person Perspective and the Embodied Mind. Lynne Baker's Challenge: Metaphysical and Practical Approaches" (San Raffaele University, June 3rd-5th 2014), and in particular to Lynne Rudder Baker, Mario De Caro, Michael Pauen and Alfredo Tomasetta for their valuable comments.

REFERENCES

Baars, B. (1997), *In the Theater of Consciousness*, Oxford University Press, Oxford;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;

Baker, L.R. (2014), "The First Person Perspective and its Relation to Natural Science", in M. C. Haug (ed.), *Philosophical Methodology: The Armchair or the Laboratory?*, Routledge, London, pp. 318-333;

Balint, M. (1968/1992), *The Basic Fault: Therapeutic Aspects of Regression*, Northwestern University Press, Evanston (IL);

Bermúdez, J.L. (2007), "Self-Consciousness", in M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Blackwell, Oxford, pp. 456-467;

Carruthers, P. (2011), *The Opacity of Mind: The Cognitive Science of Self-Knowledge*, Oxford University Press, Oxford;

Craver, C.F. (2007), *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Clarendon Press, Oxford;

Dehaene, S. & Naccache, L. (2001), "Toward a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework", *Cognition*, 79, pp. 1-37;

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.F., Poline, J.B. & Rivière, D. (2001), "Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming", *Nature Neuroscience*, 4(7), pp. 752-758;

Dennett, D. (1991), *Consciousness Explained*, Little Brown, Boston;

Dennett, D. (2001), "Are We Explaining Consciousness Yet?", *Cognition*, 79, pp. 221-237;

Dennett, D. (2005), *Sweet Dreams*, MIT Press, Cambridge (MA);

Fonagy, P., Gergely, G., Jurist, E.L. & Target, M. (2002), *Affect Regulation, Mentalization and the Development of the Self*, Other Press, New York;

Goffman, E. (1961), *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*, Doubleday, New York;

Herschbach, M. (2012), "On the Role of Social Interaction in Social Cognition: A Mechanistic Alternative to Enactivism", *Phenomenology and Cognitive Sciences*, 11, pp. 467-486;

Ismael, J. (2006), "Saving the Baby: Dennett on Autobiography, Agency, and the Self", *Philosophical Psychology*, 19(3), pp. 345-360;

Jervis, G. (2011), *Il mito dell'interiorità*, Bollati Boringhieri, Turin;

Kohut, H. (1977), *The Restoration of the Self*, International Universities Press, New York;

Laing, R. D. (1960), *The Divided Self: An Existential Study in Sanity and Madness*, Tavistock, London;

Lewis, D. (1979), "Attitudes *De Dicto* and *De Se*", *The Philosophical Review*, 88, pp. 513-543 (reprinted in Id. [1983], *Philosophical Papers*, Vol. I, Oxford University Press, Oxford, pp. 133-159);

Looren de Jong, H. (2001), "A Symposium on Explanatory Pluralism", *Theory & Psychology,* 11, pp. 731-735;

Marraffa, M. (2013), "De Martino, Jervis, and the Self-Defensive Nature of Self-Consciousness", *Paradigmi*, 31, pp. 109-124;

Marraffa, M. & Paternoster, A. (2013), "Functions, Levels and Mechanisms. Explanation in Cognitive Science and its Problems", *Theory & Psychology*, 1, pp. 22-45;

McAdams, D.P. & Olson, B.D. (2010), "Personality Development: Continuity and Change Over the Life Course", *Annual Review of Psychology*, 61, pp. 517-542;

McCauley, R.N. & Bechtel, W. (2001), "Explanatory Pluralism and the Heuristic Identity Theory", *Theory & Psychology,* 11, pp. 736-760;

Raffard, S., D'Argembeau, A., Lardi, C., Bayard, S., Boulenger, J.P. & Van der Linden, M. (2010), "Narrative Identity in Schizophrenia", *Consciousness and Cognition*, 19, pp. 328-340;

Schechtman, M. (2011), "The Narrative Self", in S. Gallagher (ed.), *The Oxford Handbook of the Self*, Oxford University Press, Oxford, pp. 394-416;

Schneider, S. (2007), "Daniel Dennett on the Nature of Consciousness", in M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Blackwell, Oxford, pp. 313-324;

Synofzik, M., Vosgerau, G. & Newen, A. (2008), "Beyond the Comparator Model: A Multifactorial Two-Steps Account of Agency", *Consciousness and Cognition*, 17(1), pp. 219-239;

Thagard, P. (2014), "The Self as a System of Multilevel Interacting Mechanisms", *Philosophical Psychology*, 27(2), pp. 145-163.

MASSIMO REICHLIN
Università Vita-Salute San Raffaele
reichlin.massimo@unisr.it

FIRST-PERSON MORALITY AND THE ROLE OF CONSCIENCE

abstract

I build on Baker's insight concerning the relationship between having a robust first-person perspective and being a moral agent in order to show the defects of some recent projects of naturalisation of morality. I argue that morality depends upon having conscience, and is an inherently first-personal experience. I then move on to criticise Baker's too neat distinction between a rudimentary and a robust first-person perspective, and suggest that Baker excessively downplays the role of embodiment in her account of what it is for the same first-person perspective to be instantiated across time.

keywords

First-person perspective, conscience, embodiment, natural kind

1. In *Naturalism and the First-Person Perspective* (2013) Lynne Baker convincingly argues that the first-person perspective (FPP) is an irreducible element of a complete ontology, and one that cannot be accounted for in terms of neural processes. I agree with this important conclusion, and with the idea that the possession of a FPP accounts for many relevant aspects that distinguish the lives of persons and make them peculiarly valuable and important.
In the first part of this article, building on Baker's insight, I will stress the peculiar link existing between moral experience and the FPP: I will suggest that we cannot adequately explain morality unless we acknowledge the peculiarity of the FPP. In the second part, focusing on human infants' early capacities for social relations, I will suggest that Baker excessively downplays the role of embodiment in her characterisation of the FPP, and that her neat distinction between the rudimentary and the robust FPP might be revised. I will conclude suggesting that Baker's theory suffers from a difficulty in the definition of what counts as an instantiation of the same FPP across time, and that this depends on her view of the relationship between persons and their bodies.

2. One central contention in chapter 9 of *Naturalism and the First-Person Perspective* is that "nothing can be a moral agent without a robust first-person perspective. Since only persons can have robust first-person perspectives, only persons can be rational or moral agents" (Baker 2013, pp. 192-193). This point, I think, is well taken. At the same time, it runs against some recent views coming from empirical studies of moral judgment. In these studies, it is suggested that morality is a much less complex phenomenon than we usually think; particularly, that moral judgment is inherently based on our emotions (Dalgleish 2004), and that all beings showing some form of empathy or possessing a system of unreflective, automatic emotions (the so called system 1) can roughly be considered moral agents. Research on the neuroscientific bases of moral judgments, as well as on the psychological and neuroscientific dimensions of empathy (Moll, de Oliveira-Souza, Zahn & Grafman 2008), have contributed much to these developments, and to the considerable revival of Humean views on morality (Nichols 2004 and Prinz 2007). According to these views, moral judgments are nothing more than gut feelings, or emotional, automatic reactions mediated by our 'sentimental' brain (Haidt 2001 and 2007); alternatively, they can be reconstructed as the outcome of a double level of processing by different areas of our brain, that is, the automatic reactions by system 1, as corrected and integrated

by the reflective, computational processes of system 2 (Greene 2008 and 2009).

While it would be wrong to neglect the importance of the neuroscientific findings for an adequate understanding of morals, it must be stressed that reconstructions such as these account in a very partial way for the whole of moral experience. The lesson to be learned, no doubt, is that morality has its roots in the deep structures of our 'sentimental' brain; however, the recent focus on emotive, automatic processes, leaves much out of consideration. The concept of morality with which these studies are working is in fact centred on swift judgments in very specific, often dramatic and sometimes bizarre quandary situations, such as those dealing with 'trolley cases', crying babies and the like (Greene, Sommerville, Nystrom, Darley & Cohen 2001). But this focus fails to account for the perhaps most important part of morality, *i.e.* the one that has to do with the development of more general patterns of reaction and disposition to act, patterns which crucially involve an idea of oneself. To put the point somehow more bluntly: an account of moral experience centring exclusively on the generation of outputs, in the form of third-person judgments on the rightness or wrongness of some considered course of action, misses most of what morality is all about. And, following Baker's point, I would like to say that most of what such an account misses has to do with the FPP.

The basic fact about the moral point of view, I would say, is that it is not third-personal, or at least, that it cannot be wholly accounted for in third-personal terms; it is an essential element in moral evaluation that rightness and wrongness cannot be defined from a spectator's viewpoint on the facts of the world (not even on 'moral', or 'axiological' facts there to be valued), but involve the adoption of a FPP. When I ask myself what is it right to do, I am not contemplating this question from an external viewpoint, I am not asking what would be the best thing to do 'from the point of view of the Universe', as it were: what I am actually asking is what *I* should do, what are the best reasons on which *I* may act, and therefore on which *I* ought to act. Of course, my gut feelings, or emotional automatic reactions, may mark the beginnings of a moral response, and may suggest the adoption of an attitude: what is crucial, however, and distinctive of an authentic moral judgment, is the fact that I decide to endorse that reaction. I stop for a while, reflecting on what are my best reasons to act on in the circumstances, and end up with the decision that that reaction is the most appropriate way to face the situation. This decision involves the consideration of my practical identity as a moral subject, that is, the consideration of my personal values, commitments and ideas of the good life (Reichlin 2014).

Baker aptly suggests that the possession of a rudimentary FPP does not enable the adoption of this specifically normative perspective, since moral reasons are always reasons bearing on the relationship between myself and others: therefore, in order to make moral judgments, I must have some image of myself as myself, and of myself in the web of my relationships with others. To eliminate or reduce the FPP, therefore, is also to reduce moral decision-making to the image of an evolutionary mechanism wired into our brains in order to produce useful behavioural outputs, as judged from the evolutionary viewpoint of the reproductive fitness of the individual and of the social group. Alternatively, it is to reduce it to the consequentialist image of a tool for correcting our biased automatic reactions with a view to maximising utility, thanks to our reflective, computational inputs from system 2. I am not suggesting that these processes play no part in the complex phenomena of our moral experience: rather, that they do not capture the whole of it, nor its main core.

A robust FPP is needed to account for a plausible view of morality. Living a moral life primarily has to do with constructing one's character or moral personality, that is, with developing habits and dispositions to feel, judge and choose according to an ideal image of oneself and an ideal of a good life. Morality, therefore, is first-personal in its essence: it does not have to do only (nor, I would say, mainly) with producing consequences (even though, of course, consequences too matter in a moral decision); its main core concerns establishing relationships with other people, and adopting principles to shape our treating one another with respect. Morality has to do with the contribution that I* – a symbol that Baker uses to refer to our capacity to think of ourselves as ourselves – give to the state of the world, through my choosing the kind of person that I* want to be, in my relationship with others. As Thomas Nagel famously put it, morality cannot be entirely accounted for in terms of agent-neutral reasons, that is, of reasons that are valid and applicable whoever is the agent (Nagel 1986); there is a fundamental aspect of morality in which agent-relative reasons are involved, that is, moral rightness and wrongness also depend on the fact that it is *me* who will be doing *x*, or producing consequence *y*. And the fact that I will be doing the action bears on my self-image, it is mirrored in it and will accompany my self-conception from this moment onwards. If the 'naturalisation' of morality is the project of translating this basic first-person experience of the commitment to moral principles and to an idea of the good life in a wholly third-personal language, then it means to miss the main point of morality, or to explain it away.

The Western philosophical and theological tradition has an apt word to capture this specifically first-personal character of moral experience: this word is 'conscience'. Conscience is perhaps no longer a fashionable word in philosophical language; nonetheless, we can say that morality depends on having conscience, in the precise sense of having a first-personal view on our agency and relationship with others. The etymology of the word tells us this very clearly, for *conscientia* refers to a peculiar kind of knowledge. It is knowledge (*scientia*) of something that is not entirely separated from the knower, since it is knowledge that, at the same time, involves an idea of oneself (*cum-scire*): knowledge that binds the knower, that directly involves the one who has it. Conscience, therefore, is not equivalent to moral sense, if this is conceived in the standard Humean meaning, according to which all that there is to morality is feeling some pleasurable or unpleasurable sentiments with reference to actions and characters. Conscience, in fact, has to do not with simple 'moral perceptions', but with the progressive structuring of one's framework of attitudes and patterns of reaction, in the light of our experience of interhuman relationships and aided by the internalisation of other people's looks and judgments on ourselves.

Friedrich Nietzsche once famously wrote that conscience depends on authority and therefore "it is not God's voice in man's breast, rather the voice of some men in man" (Nietzsche 1996, II, 2, § 52). He was partly right, because the authority of conscience depends in part on internalising others' reactions and judgments, so that their voice resonates in us and their gaze on us shapes the way in which we look to ourselves. But of course, he was also partly wrong, because our conscience is the form of our practical identity, that is, of that conception of virtue and of the good life that we define for ourselves and we aim to instantiate in our choices and actions. In Baker's terms, we can say that conscience is also a product of our FPP. This is therefore my conclusion for this first section: there is a relationship between having conscience and having the capacity to conceive of oneself as oneself; the two are generated in a common process and mutually support each other.

3. I have suggested that there is an important and reciprocal relationship between conscience and the robust FPP, and that this is why some recent projects of naturalisation fail to grasp what really is at stake in morality. I now want to show that this is no reason to forget the close relationship pointed out by empirical research between the emotional experience of human relationships and morality; moreover, that we should also take into account the role played by very early experiences of sociality in establishing both conscience and the robust FPP.

One important conclusion of contemporary studies in developmental psychology is that infants show very early signs of the distinction between themselves and the others, and very rapidly set out for the development of a sense of agency. For example, infants in the first hours of their life already distinguish their body from others', as shown by their reacting more vivaciously to the experimenter's stimulation of their cheek, than to a similar self-stimulation (Rochat & Hespos 1997). An analogous observation is provided by the fact that four-month old babies try to reach for objects that are being shown to them only within the sphere of their grasp, and show much hesitation when the object's distance is such that the attempt would endanger their bodily equilibrium. This body-scheme, of course, is subpersonal, but it is nonetheless a very early sign of the beginnings of the consciousness of oneself.

Moreover, infants also show very early signs of agency, as evidenced for example by experiments in which two-month old babies modulate their way of sucking a 'musical' dummy, according to the different sounds that it produces (Rochat & Striano 2000). These findings suggest that there is a very early experience of one's body that is a constitutive element in the concept of oneself as a distinct and original source of action. Moreover, it is well known, as mentioned by Baker herself, that humans possess a unique capacity for social interaction, cooperation and mind-reading (Tomasello 1995 and 2009), as it is shown, for example, by their very early capacity to imitate and to distinguish facial emotions. We can also add that, in a few months from birth, infants start to develop attitudes of reaction to other's actions that are clear forerunners of social and moral behaviour. For example, two-month old babies react negatively to the sudden interruption of an interaction by an adult, as if there was a sort of 'breach' of an implicit rule (Rochat 2001), and seven-month old babies differentiate their social expectations according to the fact that they are interacting with some privileged person, or with an unknown one (Layton & Rochat 2007).

The lesson that can be drawn from these data is that infants learn to identify themselves through the relationships with others: it is through the experience of the others' gaze, and the very implicit realisation of being the object of others' representation and care that the sense of oneself emerges. Self-consciousness, and the future development of the capacity for moral agency are rooted in our very early, and peculiarly human, interest for reputation, that is, for the way in which others see us. We can say that the emergence of the self is tied to the social interactions of the human infant with other people: in fact, there is a sort of co-emergence of the ideas of the self and of the others.

Now Baker is of course aware of these data. Her strategy to mark the peculiarity of persons, as distinguished from human biological organisms, is to establish a distinction between a rudimentary and a robust FPP: human babies are provided with a rudimentary FPP, but what makes us unique and places us above the animal kingdom is the possession of a robust FPP, which does not emerge until the kid possesses language and the capacity for I-thoughts. This distinction is fairly persuasive, but I suggest that it should not be taken as marking a neat boundary. In fact, since the acquisition of a robust FPP is the acquisition of a sense of oneself as oneself, the infants' early capacities to distinguish themselves from others, and to interact with them in significantly complex ways, can be considered as early steps on the way to the development of a robust FPP. As a matter of fact, it would be implausible to suggest that the robust FPP 'magically' emerges with the appearance of language; it is much more plausible to say that the sense of the self is largely acquired through very early experiences of oneself as oneself in a non-linguistic dimension, and that the acquisition of language completes and perfects the process, enabling a much wider experience of one's agency and relationships with others. To put it bluntly: it seems possible to have a very robust sense of oneself without having developed those complex verbal capacities that eventually enable the individual to express one's self-awareness.

This, of course, is not meant to suggest the existence of any kind of proto-moral behaviour in very young infants; as already noted, I accept Baker's view that only individuals with a robust FPP can be full-blown moral agents. However, the preceding observations show that the early social interactions, essentially mediated by the experience of a lived body and characterised by an emotional load, are vital elements in the formation of a robust FPP, which – as we saw – is intertwined with the generation of conscience and moral agency: as Philippe Rochat put it, the process through which children attract the gaze of others, while independently exploring their environment, is a seminal element that leads them to become increasingly self-conscious, and represents "the ontogenetic roots of the human moral sense" (Rochat 2012, p. 390). It can be added that this progressively emerging conception of oneself as oneself is mediated by some specifically moral behaviour, that is, by the adults' behaviour of caring for the infant. In this sense, we can say that the child acquires the conception of herself partly by learning to be the object of a loving relationship for her mother and other privileged persons. The co-emergence of the sense of oneself and of the sense of others is mediated by experiences of care and affection, so that we may say that the *primordia* of conscience – *i.e.* the implicit notion of being in an ethical relationship with others, and of laying claims to others' attention – are one basic factor in the emergence of a robust FPP.

Baker rightly insists on the fact that it is the robust FPP that makes morality possible: but it is also clear that some causal work goes in the opposite direction. Though the definitive acquisition of a robust FPP is the result of acquiring language, the early experiences of one's lived body and of being thrown from the start into the basic forms of ethical relationship contributes much to that result. As noted by Baker, the robust FPP implies acquiring the empirical concepts expressed by a public language, and this implies "to have social and linguistic relations" (Baker 2013, p. 139). What I want to stress is that social and linguistic relations of a specifically ethical kind are present in early infancy, and help the construction of the child's sense of oneself as oneself long before the child acquires the active use of language. This may suggest that the distinction between rudimentary and robust FPP should not be overemphasised, and the continuity between the two should be given proper recognition.

4. I have argued that there are reasons to doubt that the acquisition of cognitive abilities (*i.e.*, the use of a syntactically complex language) is the only means through which a robust FPP becomes effective, for the pre-linguistic, lived experience of embodiment in a biological organism, with the emotional experiences that this allows, is a relevant step towards the establishment of a robust FPP. Before the acquisition of the *concept* of oneself, the fact of being situated in a bodily condition, of experiencing one's body as a causal factor in effecting changes in the world of things and persons, and of experiencing a complex set of perceptions and emotional reactions associated to this embodied situation, are vital contributions to reaching a *sense* of oneself, which in turn is a major contribution to the establishment of a robust FPP. In other words, I have argued for a softening of the distinction between the rudimentary and the robust FPP. Now, I wish to give a look at the consequences of these observations for Baker's definition of the boundaries of personhood.

Baker does not assume that the possession of a robust FPP, with the associated linguistic competence, is a necessary condition of personhood, and of the full moral status that is proper to persons: she subscribes to the view that human children, also at an age at which they certainly lack a robust FPP, are already persons. The reason she gives for this view is that human infants have a rudimentary FPP *essentially*: and the reason why they have it essentially is that they are "of *a kind* that develops robust first-person perspective" (Baker 2013, p. 44, emphasis added). This makes a difference with other mammalians who have rudimentary FPP: these are of kinds that do not develop robust FPP, and therefore have the rudimentary FPP only contingently. Now, this seems to mean that, according to Baker, there is no need to *presently* possess the capacities for a robust FPP in order to be a person; it is enough that you have the *capacity to develop* a robust FPP in the ordinary history of development that is proper of your natural kind. And in fact, in the paper read at the Summer School on "Naturalism, First-Person Perspective and the Embodied Mind", Baker writes that the dividing line between a human infant, who is person, and a nonhuman organism, who is not, is that the first, but not the second, "has a remote capacity to develop a robust first-person perspective" (Baker 2014, p. 22-23); and, as defined by Baker, a remote capacity "is a second-order capacity to develop a capacity" (Baker 2014, p. 23).

I see two possible objections to this. The first is that, in the light of the scientific evidence very sketchily summarised in the preceding section, this capacity can hardly be said to be remote:

infants in the first years of age are in fact actively making their way on the road that brings to the full possession of a robust FPP, and their level of individual agency and social interaction is so high that they can fairly be said to possess a stable sense of themselves, in a way that cannot be said of most nonhuman animals. There is in fact much more continuity between human infants and grown children than Baker seems willing to allow: specifically, it seems that we can justifiably say that pre-linguistic infants enjoy some relatively complex form of inner life, so that Baker's contention according to which "without a robust first-person perspective, there would be no inner life at all" (Baker 2013, p. 140) is highly questionable.

The second, and more important, objection is that, if being *of a kind that naturally develops* a robust FPP is a sufficient condition in order to be recognised as a person, even though the relevant capacities are still in the process of construction, than it is not clear why a sentient human foetus should not be considered a person as well. The two conditions that Baker stipulates for having a rudimentary FPP are the possession of consciousness and of minimal agency: now, there is convincing evidence that a foetus with a sufficiently developed cortical function (starting from about the 24th week of pregnancy) does satisfy the first condition for it (Lagercrantz 2014), and, depending on how minimal agency can be, it might be said to partly satisfy the second as well. Therefore, since a sentient foetus is likely to possess a rudimentary FPP, and is a being *of a kind that naturally develops a robust first-person perspective*, it should be considered a person, according to Baker's criterion.

But, even though we should accept that a sentient human foetus does not satisfy the second condition, the mention of the concept of a natural kind invites the following line of argument: if being of a kind that *naturally develops* a robust FPP is sufficient in order to be considered a person, than being of a kind that *naturally develops* first rudimentary, and then robust FPP should be taken as sufficient ground for personhood as well. To stress the role of being an individual of a certain natural kind, in order to grant the human newborn the status of a person, is to accept – at least implicitly – that there is some ontological significance in being human, that is, in being a living organism "of a kind that typically develops a robust first-person perspective" (Baker 2013, p. 148). Therefore, one might be tempted to say that, at the time that the human individual is present, what you have is a biological individual of a kind that naturally develops a rudimentary FPP, at some time (perhaps late) during pregnancy, and then a robust FPP within the first three years after birth. Where should we stop this regress? Probably at the time when a human organism is present, which can safely be said some two weeks after fertilisation, when the cells' totipotency is lost and twinning can no longer occur (Ford 1991). This move would partly reconcile Baker's view with animalism, in that it would acknowledge the relevance of human biology among the conditions that account for what we essentially are: it would acknowledge that embodiment in a biological organism is a basic condition for the future emergence of the robust FPP, *i.e.*, of the distinctive mark of full-fledged personhood. Baker comes close to recognising the importance of biology for persons when she objects to Descartes' conception of the self as a solitary thinker, with no hands, nor flesh, nor blood (Baker 2013, p. 140). However, her recognition of our necessary embodiment is limited by her insistence that our body might be substituted by a bionic, wholly engineered one, while we continued to exist so long as we had our FPP. Baker's confidence in asserting this view is based on the well-known cases of the implantation of artificial limbs and other prosthetic parts that can be meaningfully integrated into our bodily scheme. I agree that an individual with an artificial arm is still the same person, because – among other things – she exemplifies the same FPP: but I think that there are limits to the alteration of our bodily image, beyond which a bionic body, made up of entirely engineered parts, can no longer support the same FPP.

Baker's view greatly underestimates the connexion between my having *this* body and my having *my* FPP. My body is not a mere biological object, it is a lived body and the very condition of *my* being open to the outside world, that is, of *my* having consciousness and a FPP: what does the fact that my FPP is the condition of my persistence mean, if it does not mean that such perspective emerges from *this particular* body? If the single fact that accounts for your being the same person that you were as an infant is that "there is a single exemplification of the dispositional property of having a first-person perspective both then and now" (Baker 2014, p. 2), is it not the case that such a FPP can be the same because it is a dispositional property of the same body? In particular, it might be said that I can be the same individual and have the same FPP only so long as I have the same brain; but we may also add, so long as I conserve the experience of looking at the world from the same bodily perspective, of acknowledging myself as the bearer of certain expressions and the user of certain gestures that mirror my inner life. It is the importance of our embodiment for our sense of ourselves that makes cases of deep and persistent disfigurement so dramatic, and imaginary tales such as Kafka's *The Metamorphosis* so irretrievably tragic. It is simply the fact that the alteration of our body affects our sense of ourselves, and we cannot exemplify the same FPP while inhabiting an utterly different body. It is true, of course, that my body is ever in the process of changing, and the cells that constituted it years and even months ago, are no longer those that constitute it now; and my brain, that coordinates all living functions and enables consciousness, is constantly changing its neurons as well. But it is also clear that a form or some other principle of unity testifies to the permanence of the same organism across the changes of its material constituents. If it is not this continuing human organism that displays the dispositional property of conceiving oneself as oneself, who or what does? If the 'miness' of the FPP across time is not accounted for by my biological continuity, than it can only be linked to my psychological continuity, as in standard psychological views of identity. But this conclusion would run into the difficulties that Baker herself raises against psychological views (Baker 2000).

5. For all her insistence on the anti-Cartesianism of her position, a slight element of dualism still lingers in Baker's view: once severed from its emergence in *my* specific bodily condition, my FPP seems an analogous either of a Cartesian soul, that might be transferred in a different, non biological body, or of a Parfitian collection of connected mental states that might be teletransported into another planet. Baker surely is not willing to accept either view: and she would also reject the partly 'animalistic' one I suggested. But has she the conceptual resources to avoid this move? And more importantly, does the reconciliation so accomplished of the FPP with the relevance of embodiment in a biological organism necessarily condemns its defender to naturalism? In other words, is the irreducibility of the FPP necessarily tied to the constitution view and to the denial of the role played by our biological bodies in shaping our identities as persons? I do not think so. It is persons – I concur – not brains nor minds who "are subjects of experience, or are rational or moral agents" (Baker 2013, p. 142): however, we persons cannot think of ourselves as ourselves outside this organic body, for a non organic, bionic body would not be *me*, in that it would not support *my* FPP. Perhaps this is only an intuition, but it is a very robust one.

REFERENCES

Baker, L.R. (2000), *Persons and Bodies: A Constitution View*, Cambridge University Press, Cambridge;
Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;
Baker, L.R. (2014), *Cartesianism and the First-Person Perspective*, this issue, pp. 20-29;
Dalgleish, T. (2004), "The Emotional Brain", *Nature Reviews Neuroscience*, 5(7), pp. 583-589;
Ford, N. (1991), *When Did I Begin? Conception of the Human Individual in History, Philosophy and Science*, Cambridge University Press, Cambridge;
Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., & Cohen, J.D. (2001), "An fMRI Investigation of Emotional Engagement in Moral Judgment", *Science*, 293(2105), pp. 2105-2108;
Greene, J.D. (2008), "The Secret Joke of Kant's Soul", in W. Sinnott-Armstrong (ed.), *Moral Psychology. Volume 3: Emotion, Brain Disorders, and Development*, The MIT Press, Cambridge, Mass., pp. 35-79;
Greene, J.D. (2009), "The Cognitive Neuroscience of Moral Judgment", in M.S. Gazzaniga (ed.), *The Cognitive Neurosciences IV*, The MIT Press, Cambridge, Mass., pp. 987-999;
Haidt, J. (2001), "The Emotional Dog and its Emotional Tail: A Social Intuitionist Approach to Moral Judgment", *Psychological Review*, 108(4), pp. 814-834;
Haidt, J. (2007), "The New Synthesis in Moral Psychology", *Science*, 316(5827), pp. 998-1002;
Lagercrantz, H. (2014), "The Emergence of Consciousness: Science and Ethics", *Seminars in Fetal & Neonatal Medicine*, 19(5), pp. 300-305;
Layton, D. & Rochat, P. (2007), "Contribution of Motion Information to Maternal Face Discrimination in Infancy", *Infancy* 12(3), pp. 1-15;
Moll, J., de Oliveira-Souza, R., Zahn, R. & Grafman, J. (2008), "The Cognitive Neuroscience of Moral Emotions", in W. Sinnott-Armstrong (ed.), *Moral Psychology. Volume 3: Emotion, Brain Disorders, and Development*, The MIT Press, Cambridge, Mass., pp. 1-18;
Nagel, T. (1986), *The View From Nowhere*, Oxford University Press, Oxford;
Nichols, S. (2004), *Sentimental Rules. On the Natural Foundations of Moral Judgment*, Oxford University Press, Oxford;
Nietzsche, F. (1996), *Human, All Too Human: A Book for Free Spirits*, R. J. Hollingdale & R. Schacht (eds.), Cambridge University Press, Cambridge;
Prinz, J.J. (2007), *The Emotional Construction of Morals*, Oxford University Press, Oxford;
Reichlin, M. (2014), "Neuroethics and the Rationalism/Sentimentalism Divide", in C. Lumer (ed.), *Morality in Times of Naturalising the Mind*, de Gruyter, Berlin, pp. 127-143;
Rochat, P. (2001), *The Infant's World*, Harvard University Press, Cambridge, Mass.;
Rochat, P. (2012), "Self-Consciousness and 'Conscientiousness' in Development", *Infancia y Aprendizaje*, 35(4), pp. 387-404;
Rochat, P. & Hespos, S.J. (1997), "Differential Rooting Response by Neonates: Evidence for and Early Sense of Self", *Early Development and Parenting*, 6(150), pp. 1-8;
Rochat, P. & Striano, T. (2000), "Perceived Self in Infancy", *Infant Behavior and Development*, 23(3-4), pp. 513-530;
Tomasello, M. (1995), "Joint Attention As Social Cognition", in C. J. Moore & P. Dunham (eds.), *Joint Attention: Its Origins and Role in Development*, Lawrence Erlbaum Publishers, Hillsdale, NJ, pp. 103-130;
Tomasello, M. (2009), *Origins of Human Communication*, Bradford Books/The MIT Press, Cambridge.

SESSION 2

CONTRIBUTED PAPERS

Alfredo Tomasetta (Istituto Universitario di Studi Superiori, Pavia)
We are Not, Fundamentally, Persons

Marc Andree Weber (Albert-Ludwigs-Universität Freiburg)
Baker's First-Person Perspectives: They Are Not What They Seem

Sofia Bonicalzi (Università di Pavia)
Does Reductivist Event-causal Compatibilism Leave Anything out? Lynne Baker's *Reflective-Endorsement* and the Bounds of the Traditional Analyses of Moral Responsibility

Alan McKay (The Queen's University of Belfast, Northern Ireland)
Constitution, Mechanism, and Downward Causation

Treasa Campbell (New Europe College, Bucharest)
A Humean Insight into the Epistemic Normativity of the Belief in the Self

Bianca Bellini (Università Vita-Salute San Raffaele)
Towards a Faithful Description of the First-Person Perspective Phenomenon: Embodiment in a Body That Happens to Be Mine

Patrick Eldridge (Katholieke Universiteit Leuven)
Observer Memories and Phenomenology

Gaetano Albergo (Università di Catania)
The First-Person Perspective Requirement in Pretense

Giuseppe Lo Dico (Università Cattolica, Milano)
Introspection Illusion and the Methodological Denial of the First-Person Perspective

Valentina Cuccio (Università di Palermo)
The Notion of Representation and the Brain

ALFREDO TOMASETTA
Istituto Universitario di Studi Superiori, Pavia
alfredo.tomasetta@iusspavia.it

WE ARE NOT, FUNDAMENTALLY, PERSONS

abstract

We are fundamentally persons, so Lynne Baker says; I argue that, assuming her metaphysical framework, this cannot be the case.

keywords

Human persons, primary kind, God, angels, souls

In this paper I want to argue that, assuming Lynne Baker's metaphysical framework, one has to conclude that, contrary to what she says, beings like us are not fundamentally persons.

Let us begin by considering some features Baker attributes to beings like us, and let us focus on me, for the sake of simplicity.

As is well known, Baker is a prominent supporter of what can be called the 'metaphysics of constitution'[1], and, she says, I am constituted by, but not identical with, my body[2]. And notice: I am essentially constituted by a body, even though not necessarily a human one; I could, in fact, be constituted by an artificial or a bionic or even a spiritual body, but I could not survive the sudden disappearance of all bodies[3].

So I am constituted by a body and this, according to Baker, is an essential feature of mine. What other properties do I possess? Well, I have many other properties, but the one which characterizes me fundamentally is the property of being a person: *person*, as Baker says, is my 'primary kind'[4]. Let us briefly see what, exactly, a primary kind is.

For any entity x we can ask "What fundamentally is x?" and the answer will be what Baker calls "x's primary kind": everything that exists is of *exactly one* primary kind – e.g. a horse, a tomato, a passport, an apple, a statue, a dog, and so on and so forth[5]. Moreover, an object's primary kind determines what sort of changes it can undergo and still exist, and what sorts of changes would result in its ceasing to exist altogether; put briefly, an object's primary kind determines its persistence conditions, so that if K is a primary kind, and x and y are Ks, then x and y have the same persistence conditions, namely the ones K determines[6].

What I have said so far will be, of course, very familiar to every reader of Baker's books and papers: I have simply given a brief summary of some of the theses Baker most frequently insists on. So one may be surprised to discover that these theses seem to lead quickly to a thorny problem.

[1] Baker (2000, 2007a). See, also, for example, Wasserman (2004) and Olson (2007, pp. 48-59).
[2] Baker (2000, pp. 91-101; 2007a, pp. 67-94).
[3] Baker (2000, p. 214; 2007b).
[4] See, for example, Baker (2000, p. 96; 2007a, p. 38).
[5] For example: Baker (2000, pp. 39-40; 2007a, pp. 67-68).
[6] See, for example, Baker (2000, pp. 39-40; 2007a, p. 33, pp. 219-220; 2013, p. 224).

Let us see what the problem is by considering the following argument, whose first and second premises simply restate two of Baker's main tenets which I have just talked about:

> *Premise 1) Person* is a primary kind.
>
> *Premise 2)* If *Person* is a primary kind, and x and y are persons, then x and y have the same persistence conditions.

Now add to these two premises the following thesis:

> *Premise 3)* God (if He exists), angels (if they exist), immaterial or Cartesian souls (if they exist), and beings like us are all persons.

From the three premises just stated, one can immediately conclude that God, angels, Cartesian souls and beings like us all share the same persistence conditions. But this, of course, is simply absurd (for example: we cannot survive the disappearance of all bodies – we are essentially constituted by a body – while God, angels and Cartesian souls can). So here we have a real predicament: what premises would Baker reject?

Consider the possibility of rejecting premise 3). Perhaps a non-Christian philosopher would be inclined to say that it is a mistake to think of God as a person – and so she would deny the thesis according to which if God exists, then God is a person. Yet, notice that this idea is a non-starter for Baker, who is a committed Christian.

But let us set aside divine – and angelical – topics, and let us focus just on Cartesian souls. These entities have a sophisticated mental life – they reason, desire, hope, feel, and so on: denying that these things have the *status* of persons is indeed very implausible, and so it seems difficult to deny premise 3) entirely.

But supposing premise 3) was concerned just with souls, a friend of Baker could perhaps say that they are not persons exploiting the following idea: according to Baker, if x is a person, then x has a language, and if something has a language, then it belongs to a linguistic community[7]. But, one could say, souls cannot belong to a linguistic community, so souls are not persons – and the 'just souls' version of premise 3) would be refuted.

And yet: is it really true that souls cannot belong to a linguistic community? I do not think so. Suppose that something like Descartes metaphysics is on the right track, and so suppose that there are immaterial souls causally interacting with bodies and, through these bodies, with each other: given this mutual interaction it is quite obvious, it seems to me, that these souls *can* belong to a linguistic community, and so they may well be persons.

Thus the prospects for denying premise 3), even in its 'just souls' version, are, I think, rather dim. Let us focus, then, on the second premise, and let us consider three different ways of denying it.

1st way – A denier of premise 2) could say: "It is true that *person* is a primary kind, and it is true that, God, angels, souls and beings like us are all persons; but it is not true that God, angels and souls share with us their persistence conditions. This is because God, angels and souls simply *cannot have persistence conditions*. Why so? Well, God is the absolute, infinite being, an entity to which one cannot correctly attribute any persistence condition; as for angels and souls, they are immaterial beings and it is not clear what would make them cease to exist".

[7] See, for example, Baker (2013).

To this I offer two answers.

a) The persistence conditions associated with an entity x can be thought of as determining two disjoint sets: the set of what x can survive and the set of what x cannot survive. In the case of God the second set is plausibly empty, but this is not to say that God does not have persistence conditions: rather, He possesses *trivial* persistence conditions, which is quite another thing. As for angels and souls, God certainly could annihilate them: so they do seem to have persistence conditions, and not even trivial ones.

b) But let us concede, for the sake of argument, that God, angels and souls do not have persistence conditions. In this case, and by Baker's own lights, one has to confront a troublesome consequence. Let us see what this consequence is, by first considering the following principle held by Baker: for every possible world w and every time t,

> (PC) If x exists in w at a time t and x is not eternal in w, then x has persistence conditions in w[8].

Now, let us focus on souls, and consider any possible world w in which souls exist. We are assuming that souls cannot have persistence conditions, and so souls do not have persistence conditions in w. So, by PC and *modus tollens*, one has to conclude that

> It is not the case that (souls exist in w at a time t and souls are not eternal in w).

So, either souls do not exist in time in w – that is, in w they exist outside of time – or they are eternal in w[9]. And, given that "eternal" can mean "outside of time" or "existing at each moment in time", the upshot is that, in w, either souls exist outside of time or they exist at each moment in time.

Let us state briefly where we have got to: if one assumes that souls cannot have persistence conditions, then

> For every possible world w in which souls exist, either souls are outside of time in w, or they exist at each moment in time in w.

But, of course, Cartesian souls do not exist outside of time, and so, for every possible world w in which souls exist, they exist at each moment in time in w. And this is the troublesome consequence of assuming that souls cannot have persistence conditions: saying that for every possible world w in which souls exist, they exist at each moment in time in w, means that it is *metaphysically impossible* for a universe inhabited, at a certain time, by souls to exist without souls – a quite implausible thesis by itself, and certainly not a thesis that most committed Christians like Baker would be happy to endorse.

2nd way – A denier of premise 2) could, nonetheless, try another line of argument: "The persistence conditions of beings like us are not determined solely by our being persons, but also by the bodies that constitute us. So it is true that *person* is a primary kind, and it is true that, God, angels, souls, and beings like us are all persons; but it is not true that God, angels and souls share with us their persistence conditions, because our persistence conditions are partly determined by the bodies that constitute us, and these bodies do not constitute God, angels and souls". In conversation Baker herself

[8] Baker (2007a, p. 221). Reference to possible worlds is mine but it can be considered implicit in Baker's original statement.

[9] Let me unravel this line of reasoning a little. It is not the case that (souls exist in w at a time t and souls are not eternal in w) implies that either (1) it is not the case that souls exist in w at a time t or (2) it is not the case that souls are not eternal in w. Let us consider (1). Souls do exist in w, we have assumed, so if (1) is true, then it has to be the case that souls exist in w and they do not exist at a time t. But, of course, time t is a variable standing for any time whatsoever, and so one has to say that souls exist in w and that they do not exist at any time; therefore souls exist in w outside of time. Finally, and obviously, (2) implies that souls are indeed eternal in w.

has suggested a reply along these lines to me but I have to say that I find it quite puzzling, and I am going to briefly explain why.

Certainly God is not constituted by anything – and Baker says so following what most Christian traditions have upheld[10]; moreover postulating a sort of 'spiritual stuff' constituting angels or souls is really quite implausible. So one should say that God, angels and souls are not constituted by anything, and therefore that their persistence conditions are fully determined by the primary kind to which they belong, namely the primary kind *person*. But then – first problem – it seems rather peculiar to say that this is not the case for beings like us.

More importantly – second problem – it is quite difficult to reconcile the idea according to which we are fundamentally persons with the idea that what we fundamentally are does not fully determine our persistence conditions.

And to these one may add a third problem. Baker says that the body that is now constituting me, let us say B, belongs to the primary kind "human body", and so, of course, B cannot survive the disappearance of all biological bodies[11]. But if B contributes to determining my persistence conditions, it seems that I cannot survive the disappearance of all biological bodies, either, and this runs against what Baker says about beings like us, namely that we can have bionic or artificial bodies, and so that we *can* survive the disappearance of all biological bodies.

3rd way – Let us finally consider a third way to deny premise 2) which is somewhat related to the one just examined[12]: "We are fundamentally persons, and *person* is a determinable kind-property which can be determined in different ways – *human person* being one such possible determination. If so, then, arguably, from 'x and y are persons', it does not follow that 'x and y have the same persistence conditions' – *contra* premise 2)". Is this a convincing line of reasoning? Clearly, it does not seem to be. If we are fundamentally persons, then *person* is our primary kind – a primary kind, Baker says, is by definition the kind-property which determines what a thing fundamentally is. So according to the proponent of the 3rd way, *person* is at the same time a primary kind and a determinable kind-property. But how could a *determinable* kind-property determine what a thing fundamentally is? Determinable kind-properties, such as *mammal*, *artifact*, *elementary particle*, or *vegetable*, clearly do not define the fundamental nature of their bearers, as instead kind-properties such as *horse*, *statue*, *electron* or *cabbage* do. So one cannot say, on pain of contradiction, that person is both a primary kind *and* a determinable kind-property.

To conclude: I have considered some arguments through which Baker could deny premise 2) or premise 3), and could block the conclusion that God, angels, Cartesian souls and beings like us all share the same persistence conditions; but these arguments, I have tried to show, fail. Now, perhaps Baker has the resources and the ability to plausibly deny, in different ways, premise 2) or 3), but I cannot see how this could be done. So, I believe, the only choice left is to deny premise 1), but this means that *person* is not a primary kind, and *a fortiori* that it is not *our* primary kind. So, assuming Baker's metaphysical framework, we are not fundamentally persons, which is what I wanted to argue for.

[10] Baker (2007a, p. 79).

[11] Suppose it can: then B, which is fundamentally a human organism, can exist in a world deprived of all biological bodies, which is absurd.

[12] I owe this objection to an anonymous referee.

REFERENCES

Baker, L.R. (2000), *Persons and Bodies: A Constitution View*, Cambridge University Press, Cambridge;
Baker, L.R. (2007a), *The Metaphysics of Everyday Life*, Cambridge University Press, Cambridge;
Baker, L.R. (2007b), "The Metaphysics of Resurrection", *Religious Studies*, 43(3), pp. 333-348;
Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;
Olson, E.T. (2007), *What are We?*, Oxford University Press, Oxford;
Wasserman, R. (2004), "The Constitution Question", *Nous*, 38(4), pp. 693-710.

MARC ANDREE WEBER
Albert-Ludwigs-Universität Freiburg
andree-weber@t-online.de

BAKER'S FIRST-PERSON PERSPECTIVES: THEY ARE NOT WHAT THEY SEEM

abstract

Lynne Baker's concept of a first-person perspective is not as clear and straightforward as it might seem at first glance. There is a discrepancy between her argumentation that we have first-person perspectives and some characteristics she takes first-person perspectives to have, namely, that the instances of this capacity necessarily persist through time and are indivisible and unduplicable. Moreover, these characteristics cause serious problems concerning personal identity.

keywords

Personal identity, first-person perspective, fission, Lynne Baker

1. Introduction

The notion of a first-person perspective (FPP) seems as natural as any that one could hope for. After all, each of us has his or her particular point of view, from which he or she perceives the world. To deny that such a point of view exists is absurd. To doubt that it plays a vital role in many practical respects is barely reasonable. Phenomenologically[1], neither the fact that we have FPPs nor their most essential characteristics are a matter of controversy.

Lynne Baker's concept of an FPP, as she develops it in her latest book, *Naturalism and the First-Person Perspective* (2013a), is not quite as straightforward as our ordinary notion. Though she seems to take them to be identical, and though she justifies our having FPPs (in her sense) by an appeal to phenomenology and common sense, what she calls an FPP is more theory-laden. Or so I will argue.

I start by indicating that something that has a point of view need not thereby persist through time, as Baker takes instances of FPPs to do, and that, for quite similar reasons, such instances need not be indivisible or unduplicable (section 2). I then point out metaphysical (in section 3) and practical (in section 4) complications concerning personal identity to which those unneeded characteristics give rise. These complications, albeit not irresolvable, should be regarded as sufficiently severe to cause us to think twice about FPPs. Indeed, one can give an account of FPPs that preserves Baker's irreducibility thesis – the defence of which comprises by far the most space and effort in her book – without having the disadvantages I criticise (section 5). Such an account, though quite unlike Baker's, is no less common-sensical than hers and is better supported by our intuitions on personal identity. The upshot of all this is that Baker's concept of an FPP, which seems to emerge so naturally, relies heavily on presuppositions in a way that is not made transparent by what she writes.

2. Dubious Characteristics of First-Person Perspectives

Baker defines an FPP as the capacity to make self-attributions of first-person reference (Baker 2013a, pp. 33-35) such as

(1) I am glad that I* am a philosopher now.
(2) I deeply regret that I* once was a fortune-teller.

[1] Here and hereafter, I use "phenomenology" and its derivatives in the broadly analytic sense in which the word simply refers to the qualitative character of experience.

Both statements are self-attributions (since they are about the utterer of the sentence), and they are of first-person reference (since the "I" marked with an asterisk refers back to the same person as the "I" in the main clause). By making such statements, in which the utterer is both subject and part of the object of the thought, one shows one's ability to think about oneself as oneself. As it is certainly true that normal human beings have the capacity to think and utter statements such as (1) or (2), it seems to be beyond doubt that human beings have FPPs in this sense.

Let us note, however, a difference between (1) and (2): In order to be a *self*-attribution, (2) presupposes that I persist through time, whereas (1) does not. I can be glad that I am a philosopher now without having to admit that I existed yesterday, whereas I cannot deeply regret that I once was a fortune-teller without assuming that I already existed in the not-too-recent past. Baker does not distinguish between *synchronic* self-attributions of first-person reference such as (1), which can be literally true without me persisting through time, and *diachronic* self-attributions of first-person reference such as (2), which cannot. However, her discussion of personal identity clearly shows that she assumes that a particular FPP can be exemplified by the same entity for longer than a moment (for instance, there would be no problem at all with fission cases if none of the persons involved lived for more than a short while). Thus, it is safe to assume that her concept of an FPP presupposes that we persist through time – a fact that is controversial, to say the least, from a phenomenological point of view. G. Strawson, for example, doubts our persistence through time for purely phenomenological reasons; all we can perceive, according to him, is a moment of consciousness that purports to have memories of other moments of consciousness (Strawson 2003, pp. 356-359).

In addition, merely having an FPP does not entail that someone who utters a diachronic self-attribution thereby makes a true statement of personal identity: To have the capacity to make a certain kind of self-attribution does not include the truth of this self-attribution. According to Baker's definition of FPPs, conclusive evidence for the fact that each of us has an FPP comes from our use of sentences such as (1) and (2); it is not required by that definition that the use is correct. Take, for example, a theory according to which there are many short-lived instances of FPPs that follow one another and together form what we commonly call a person. Given that we have our FPPs essentially, as Baker claims, I would then exist for only a moment; hence, it would be literally false for me to utter sentences like (2), which presuppose my persistence through time[2]. Such a theory is not ruled out by Baker's definition. In order to preclude it, she has to presuppose that our capacity to utter and understand diachronic self-attributions of first-person reference guarantees the literal truth of these sentences. In doing this, she dismisses from the outset the theories of philosophers such as Hume, Russell, Perry, Parfit, Lewis, Noonan and Strawson, all of whom claim that a person is nothing more than a series of interrelated mental and (perhaps also) physical events[3].

Of course, Baker could claim that having an FPP indeed presupposes persistence through time because it involves having literally true memories, making literally real commitments, and so on. Then, however, we are in need of a further argument for the claim that we have FPPs[4] because we cannot rely on phenomenology or on common sense anymore. The reasons are that persistence through time is phenomenologically doubtful and that common sense is silent when it comes to highly theoretical ontological matters, such as whether we are enduring or perduring entities. Baker thus faces a dilemma: The more interesting the characteristics she takes FPPs to have, the less clear it is whether we have FPPs at all.

[2] Uttering such diachronic self-attributions could still be correct in a less literal way, because we can, when faced with their obvious falseness, reinterpret our personal pronouns by taking them to refer not to what we essentially are but to what we would commonly call a person, namely, an aggregate of instances of FPPs that extends over time.

[3] See Hume (1739, pp. 164-171), Russell (1957, p. 89), Perry (1972), Parfit (1984, pp. 210-217, 261-266), Lewis (1983), Noonan (2003, p. 228) and Strawson (2003).

[4] An argument on that point is defended in Nida-Rümelin (2010, pp. 198-201). According to Nida-Rümelin, self-attributions are conceptually prior to self-identifications and cannot sensibly be regarded as false. Taken together, these two claims establish that our capacity to make them entails our persistence through time.

Similar lines of argument can be put forward against the presumed indivisibility and the presumed unduplicability of instances of FPPs. For Baker, "[a]n exemplification of the first-person perspective is like a haecceity, or individual essence" (Baker 2013a, p. 149 n. 6). However, this haecceitistic nature, in which properties such as indivisibility and unduplicability are grounded, does not follow from her characterisation of an FPP as "the capacity to conceive of oneself as oneself* in the first person" (Baker 2013a, p. 35)[5]. For instance, if we regard diachronic self-attributions such as (2) as true only in the less literal way described in the footnote 2, or if we regard them as understandable only if their subjects and their objects share the same brain or body, then our capacity to make them does not require anything like an individual essence.

Besides not being entailed by the relevant definition, indivisibility and unduplicability are also highly questionable characteristics of FPPs. To see this for indivisibility, take a fission case, in which the brain of a person, say, Angela Merkel, is split into two half-brains, each of which is transplanted into the head of another person sufficiently similar to the original. The resulting person who has received the left half-brain is called Lefty and the resulting person who has received the right half-brain is called Righty. Both Lefty and Righty, it is assumed, are in every relevant respect proper mental successors of Merkel: They remember being her and share her thoughts, desires, beliefs and character traits. Thus, each of them has the impression of experiencing Merkel's FPP, though they now obviously have different FPPs. It appears that Merkel's FPP has been divided.

For unduplicability, take a scenario in which Merkel is scanned, and the screening data is used to generate a perfect physical duplicate of her. It is then plausible to assume that this duplicate remembers being her as well. In other words, the duplicate has the impression of experiencing Merkel's FPP up to the point of time at which the duplication procedure started, though she now obviously has a perspective different from Merkel's. It appears that Merkel's perspective has been duplicated.

The notorious complaint against this kind of reasoning is that it relies heavily on unrealistic, "far-out" thought experiments. For example, it is taken for granted that it is indeed possible that both Lefty and Righty are proper mental successors of Merkel, and that there are perfect physical duplicates of her. But why should that be so? Moreover, even if it were so, why should we build our philosophical theories around scenarios that are far from being realised?

Though I think that these questions can be answered[6], this is not the place to discuss them. The point to make here is that Baker, though she has "little patience" (Baker 2013a, p. 153) with thought experiments such as fission, uses them to bring out certain features of her position more clearly[7], and, more importantly, she does not seem to regard a rejection of far-out thought experiments as a precondition for her theory. So even if she, who considers herself to be a "Practical Realist" for whom it is of considerable interest whether a scenario is a real-life case (Baker 2013b, p. 38), rejects any lesson drawn from highly hypothetical cases, she certainly would not wish her theory to be attractive only for those who share her scepticism. If this is true, it is legitimate to confront her account with critical thought experiments.

In short, our intuitions on personal identity (and hence, if we take qualitative experience to include thought experiment intuitions, our phenomenological evidence) give us no reason to suppose that instances of FPPs are indivisible or unduplicable or persisting through time; quite the contrary. Neither do these characteristics follow from Baker's definition of FPPs. Moreover, to suppose that their instantiation is warranted by common sense would mean misinterpreting the role of ordinary judgment in theoretical discussions, in which the uncritical preservation of an alleged mode of

[5] Strictly speaking, this is her explication of *robust* FPPs but we can safely ignore this difference here.

[6] See, for example, Sorensen (1992, pp. 21-50, 274-289), Gendler (2004) and Williamson (2007, pp. 179-207) for general considerations about the significance of thought experiments, as well as Nagel (1971), Parfit (1984, p. 219, pp. 245-247, p. 255), Kolak (1993) and Eklund (2002) for a defence of bizarre thought experiments on personal identity, including fission.

[7] See in Baker (2013a) e.g. pp. 153f. for fission and p. 149 for a perfect replica.

thinking is not, by default, a requirement of rationality. What Baker's account lacks is either further argumentation or an explicit statement that it rests on highly controversial preconditions.

3. Metaphysical Complications Concerning Personal Identity

Baker dubs her theory of personal identity the "Not-So-Simple Simple View". A simple view of personal identity is one that offers no non-trivial, non-circular, non-identity-involving conditions for personal identity. Baker's theory is simple in this sense because she defines persons as entities that have an FPP essentially (Baker 2013a, p. 149), and because one cannot give, according to her, any informative identity criteria for FPPs (Baker 2013a, pp. 154ff.). As her theory is nevertheless compatible with materialism (Baker 2013a, p. 151), it is not-so-simple.

Since persons are individuated by FPPs, Baker's theory relies heavily on her assumption that the instances of FPPs persist through time and are indivisible and unduplicable. This can best be illustrated by means of a fission case, such as the one in which Angela Merkel's brain is divided. Here, the indivisibility of instances of FPPs entails that at most one of the two fission products can share Merkel's FPP. But which one? According to Baker,

> [t]he answers are either Lefty, Righty, or neither, and the Not-So-Simple Simple View is compatible with all three answers. We may not know which is the correct answer, but the Not-So-Simple Simple View implies that there is a fact of the matter that depends on whether Lefty or Righty or neither has the original person's first-person perspective (Baker 2013a, pp. 153-154.).

Like adherents of theories of immaterial substances, Baker claims here that there is a fact of the matter whether Lefty, Righty or neither has the original person's FPP, although there is no empirical evidence whatsoever concerning who shares her perspective, given that the scenario is perfectly symmetrical. Thus, Baker has to admit that we may not know which person shares the original person's FPP, even though the case seems quite clear because both Lefty and Righty are perfect mental successors of the original person. One cannot help having the impression here that instances of FPPs are indivisible precisely because being the same person should be defined in terms of having the same FPP.

Unlike theorists of immaterial substances, Baker can claim that there is a fact of the matter in fission cases without claiming the existence of philosophically suspect soul-like entities (note that FPPs are properties, not objects). However, we should not be too quick to credit that to her concept of FPP because a simple view need not involve this concept in order to be not-so-simple. Many properties that supervene on, or are emergent from, physical ones can do the work of FPPs with respect to personal identity. For instance, a view according to which personal identity supervenes (for some reason or other) on the identity and intactness of a certain part of the brain would yield the same results: There is a (perhaps unknowable) fact of the matter as to what happens in thought experiments such as fission cases, namely, that the post-fission person who owns the critical part of the brain is identical to the original person, and we need not invoke immaterial substances but only particular supervenience facts. In addition, only FPPs in Baker's sense are sufficient for her theory of personal identity because having an FPP in the ordinary, phenomenologically harmless sense does not imply persisting through time or being indivisible and unduplicable. In short, Baker-style FPPs are not necessary for a not-so-simple simple view of personal identity, and ordinary FPPs are not even sufficient for such a theory.

Things are even worse. Baker's account also has severe metaphysical consequences that she does not discuss. In order to explain them, I will present two thought experiments, one given by Parfit, the other inspired by him. In Parfit's so-called "Combined Spectrum" (Parfit 1984, pp. 236-240), a series of cases is described in which one person is transformed by a molecule per molecule exchange into another, say, Angela Merkel into Vladimir Putin. In the first case, only one molecule of Merkel's body is replaced by the respective molecule of Putin's body (it does not matter which molecule). In the second case, a second molecule of Putin replaces a second molecule of Merkel. And so on. In the last case of the spectrum, all of her molecules are replaced by Putin's. Near the one end of the spectrum, the resulting persons are clearly more similar to Merkel and thus can be said to still have her FPP. Near the other end of the spectrum, the resulting persons are clearly more similar to Putin and thus can be said to have his FPP. However, what can be said about the cases in the middle of the spectrum? As Merkel and Putin obviously do not share the same FPP, and as there can be no persons without an FPP at all, the persons in the middle either share Merkel's perspective or Putin's or have a new one. Whatever option one chooses, there have to be, somewhere in the spectrum, two adjacent cases that differ only by one molecule (and hence are qualitatively more or less identical) but exemplify different FPPs (for example, either Merkel's and Putin's, or Merkel's and a new one). This is highly implausible.

In another series of cases (call them "Reunion Spectrum"), a fission case is followed by the fusion of the fission products. Imagine Merkel's brain split into two half-brains that are kept separate for some time and then unified again, so that the unified brain has memories of Lefty, of Righty and of pre-fission Merkel. In this Reunion Spectrum, the cases vary with respect to the time that passes between fission and fusion. At one end of the Reunion Spectrum, several years lie in between, and during that time obviously two distinct persons existed, exemplifying two distinct FPPs. At the other end of the spectrum, only a very short time – a second, say – has gone by, and only one person was involved, namely, Angela Merkel[8]. Somewhere near the middle of the spectrum, however, there have to be two adjacent cases, differing only in that the reunion takes place a second earlier in one of them, such that in one of them only one FPP is exemplified, whereas in the other one two FPPs are exemplified. This, again, is highly implausible.

So each of the spectra reveals the implausibility of the assumption that personal identity depends on a property that is – in the cases under consideration, in which all persons have robust FPPs – either definitely exemplified or definitely not exemplified. The easiest way out seems to be to deny outright the validity of these spectra. However, to say it again, denying the validity of far-out thought experiments is only an option for Baker if she is prepared to view scepticism concerning such thought experiments as a necessary condition for her account, and thereby to limit its scope and persuasive power.

4. Practical Complications Concerning Personal Identity

In addition to these metaphysical complications, there are practical ones. One of the most prominent reasons why we are interested in personal identity at all is that we hope to find answers to questions such as "What matters in survival?" and "Under what conditions is someone responsible for particular actions?". It is hard to believe, however, that virtually imperceptible deviations (such as in adjacent cases of the spectra) can make all the difference when it comes to survival or moral responsibility. Take the fission case: If the original person committed a crime *before* the fission took place, who is morally responsible for that crime, Lefty or Righty? It is implausible that it is whoever shares the original person's perspective, for two reasons. For a start, we are simply not able to determine if this is Lefty or Righty or neither. More importantly, sameness of memories, beliefs, desires and character traits seems far more important for moral responsibility than some non-empirical and unknowable fact. If this is right, then being morally responsible does not consist in having the same FPP.

Similarly, what matters for me in terms of my survival? That there is some future person who is psychologically continuous to me in the sense that this person shares my thoughts, feelings, memories and so on, or that there is some future person who shares my FPP? Normally, psychological

[8] In Baker (2000, pp. 162ff.), Baker states that there is only one FPP exemplified in reunion cases, though there are two streams of consciousness. This explanation makes it even harder to individuate FPPs.

continuity and having the same FPP go hand in hand. In thought experiments in which they come apart, however, it becomes obvious that psychological continuity is of primary importance for us and that we do not care about things we may not even know, namely, whether our FPP is still the same. Therefore, Baker-style FPPs are inappropriate to capture what lies at the heart of the debate on personal identity: Questions concerning rationality and morality.

5. An Alternative Account of Irreducible First-Person Perspectives

If it were clear that we have FPPs in Baker's sense, then we would have to live with these consequences. We can, however, save the most central feature of Baker's account, namely, the claim that FPPs are irreducible to impersonal properties, without being committed to an implausible view of personal identity. As I have explained, instances of FPPs need not persist through time. We could, alternatively, take them to be, for example, moments of consciousness (MOCs). Consider an aggregate, a mereological sum, of MOCs that are psychologically interrelated, so that there is continuity of memories, desires and beliefs between earlier and later MOCs. There is no entity (or, if we take the mereological sum itself to be an entity, only a trivial entity) that unifies these MOCs. Nevertheless, we could, as a matter of convention, regard all these MOCs as sharing the same FPP.

If we do so, there are two FPPs involved in a fission case: One that is the exemplification of the MOCs that together made up the original person plus the MOCs that together make up Lefty; and one that is the exemplification of the MOCs of the original person plus those of Righty. Thus, the original person is divided into two persons. Similarly, one could show that instances of FPPs are duplicable. Furthermore, we could understand FPPs in this way when we read Baker's explications of agency and moral responsibility (Baker 2013a, pp. 183-206); though agency and moral responsibility presuppose, according to her, an FPP, there is no reason to think that a phenomenologically harmless one whose instances need not persist through time does not suffice. Of course, Baker would not accept this account; and neither would I like to endorse it; I just state it to show that there is a theoretical option that Baker does not consider.

Why are FPPs, as they appear in this view, irreducible? Compare what Baker writes about the uninformativeness of simple views of personal identity:

> [I]t is impossible to have informative necessary and sufficient conditions for transtemporal personal identity: Persons are basic entities; being a person does not consist in satisfying nonpersonal conditions. So, any correct account of personal identity must be uninformative; otherwise, it would be reductive (Baker 2013a, p. 154; footnote omitted).

If we take the instances of FPPs to be MOCs rather than persisting entities, we reduce persons to specific aggregates of MOCs. We give a reductive account of personal identity. That is what Baker criticises in the quotation: The reductiveness of accounts that deny that the owners of FPPs persist through time. She overlooks, however, that we have to distinguish two kinds of reductionism: reductionism that reduces persons to, for example, MOCs, and reductionism that reduces intentional states to neural states. It is this last kind of reductionism against which Baker's arguments for the irreducibility of FPPs are directed: A reductionism of the mental in terms of the physical. Logically independent from it is the first kind of reductionism, the reductionism of the persistent in terms of the momentary.

To be sure, there are philosophers who argue for a close tie between these two kinds of reductionism, most notably Sydney Shoemaker[9]. Whether there really is such a close tie, however, is a matter of controversy. Therefore, it is best not to take it for granted that the irreducibility of intentionality entails the irreducibility of persons as persisting entities. Arguably, it seems possible to combine an account of irreducible FPPs both with a more conclusive phenomenology and with a complex, and hence more informative, theory of personal identity.

[9] See, e.g., Shoemaker (1984) and Shoemaker (1997).

REFERENCES

Baker, L.R. (2000), *Persons and Bodies*, Cambridge University Press, Cambridge;

Baker, L.R. (2013a), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;

Baker, L.R. (2013b), "Technology and the Future of Persons", *The Monist*, 96, pp. 37-53;

Eklund, M. (2002), "Personal Identity and Conceptual Incoherence", *Noûs*, 36, pp. 465-485;

Gendler, T. (2004), "Thought Experiments Rethought – and Reperceived", *Philosophy of Science*, 71, pp. 1152-1163;

Hume, D. (2007/1739), *A Treatise of Human Nature*, Oxford University Press, Oxford;

Kolak, D. (1993), "The Metaphysics and Metapsychology of Personal Identity: Why Thought Experiments Matter in Deciding Who We Are", *American Philosophical Quarterly*, pp. 39-50;

Lewis, D. (1983), "Survival and Identity", in *Philosophical Papers*, Vol. 1, Oxford University Press, Oxford, pp. 55-77;

Nagel, T. (1971), "Brain Bisection and the Unity of Consciousness", *Synthese*, 22, pp. 396-413;

Nida-Rümelin, M. (2010), "An Argument from Transtemporal Identity for Subject Body Dualism", in G. Bealer & R. Koons (eds.), *The Waning of Materialism*, Oxford University Press, Oxford, pp. 191-212;

Noonan, H. (2003), *Personal Identity* (2nd ed.), Routledge, London;

Parfit, D. (1984), *Reasons and Persons*, Oxford University Press, Oxford;

Perry, J. (1972), "Can the Self Divide?", *The Journal of Philosophy*, 69, pp. 463-488;

Russell, B. (1957), "Do We Survive Death?", in *Why I am Not a Christian and Other Essays on Religion and Related Subjects*, Unwin Books, New York, pp. 88-93;

Shoemaker, S. (1984), "A Materialist's Account", in S. Shoemaker & R. Swinburne, *Personal Identity*, Blackwell, Oxford, pp. 67-132;

Shoemaker, S. (1997), "Parfit on Identity", in J. Dancy (ed.), *Reading Parfit*, Blackwell, Oxford, pp. 135-148;

Sorensen, R. (1992), *Thought Experiments*, Oxford University Press, Oxford;

Strawson, G. (2003), "The Self", in R. Martin & J. Barresi (eds.), *Personal Identity*, Blackwell, Oxford, pp. 335-377;

Williamson, T. (2007), *The Philosophy of Philosophy*, Blackwell, Oxford.

SOFIA BONICALZI
Università di Pavia
sofia.bonicalzi@gmail.com

DOES REDUCTIVIST EVENT-CAUSAL COMPATIBILISM LEAVE ANYTHING OUT? LYNNE BAKER'S *REFLECTIVE-ENDORSEMENT* AND THE BOUNDS OF THE TRADITIONAL ANALYSES OF MORAL RESPONSIBILITY*

abstract

Promising to be the best companion for scientific naturalism, compatibilism usually espouses a reductivist event-causal background. Lynne Baker challenges this view, arguing that compatibilist moral responsibility also requires an irreducible "first-person perspective". In this paper I will provide some arguments for claiming that (Frankfurt-type) event-causal accounts cannot avoid making reference to some sort of agential properties. In the second part, I will present the proposals formulated by Nelkin and Markosian for defending agent-causation, before returning to the theme with which I began, this time considering Frankfurt's view in the light of Baker's reading.

keywords

Compatibilism, event-causal views, agent-causal views, Reflective-Endorsement

1. Preliminary Notes

Compatibilism, as I understand it, is a label that characterizes several different views, which share the idea that the truth of determinism is compatible with the existence of free will and/or with the plausibility of moral responsibility attributions. Promising to be the best companion for scientific naturalism, compatibilism – with some notable exceptions (Nelkin 2011; Markosian 1999, 2012; Horgan 2007) – usually espouses an event- (or state-) causal background, in which intentional action is explained in terms of the interaction between different mental states that causally determine one's choices. In her latest book (2013), Lynne Baker claims that moral responsibility (like agency) requires something more than this and, in particular, it requires an irreducible first-person perspective, something that is usually not so welcome in scientific naturalistic views. Since Baker does not want to give up either ("near") naturalism, or some event-causal background, or compatibilism, her proposal sounds especially challenging.

The paper is organized as follows. First, in order to carve out a more specific battlefield, I will focus on determinism-friendly accounts originating from Harry Frankfurt's seminal work. My working hypothesis is that Frankfurt-type compatibilism faces some difficulties in explaining how we are in control of our actions and, in particular, what happens when one experiences a clash among different motivational streams. I will try to show that these accounts cannot avoid making reference to some sort of agential properties, and I will mention the proposals formulated by Dana Nelkin and Ned Markosian to defend compatibilist agent-causation. Then – given the dubious implications of these approaches and with some new concepts in place – I will return to the theme with which I began, this time considering Frankfurt's position in the light of Baker's reading.

The paper has greatly benefitted from the discussion with Lynne Rudder Baker and the audience of San Raffaele Spring School of Philosophy (2014). I am also grateful to the referees for providing constructive comments. I am deeply indebted to Derk Pereboom, who read previous versions of the present work, suggesting many determining improvements. My special thank goes to Mario De Caro and Luca Fonnesu for their invaluable intellectual support and constant encouragement. For useful feedback and suggestions, I wish also to thank Jennifer Hornsby, Michelle Ciurria, Muhammad Ali Khalidi, and Matteo Grasso.

2. Frankfurt-type Compatibilism and the Disappearing Agent Objection

According to Frankfurt (1988), one acts freely and responsibly, when one acts on a first-order desire that is in accord with a correspondent second-order desire/volition. There is a sense in which, acting as he wants, the *willing addict* (who wants to will to take the drug) is free and responsible for taking the drug, while the *unwilling addict* (who desires to take the drug but does not want to will to take the drug) is not. Moral responsibility is understood in terms of "identification": "Even if the person is not responsible for the fact that the desire occurs, there is an important sense in which he takes responsibility for the fact of having the desire – the fact that the desire is in the fullest sense his, that it constitutes what he really wants – when he identifies himself with it" (Frankfurt 1988, p. 170).

Frankfurt-type compatibilism has been reformulated in several ways, in which different sorts of mental states play the leading role (e.g. Watson 2004, pp. 13-32; Bratman 2001). There are various reasons why these accounts, despite the powerful objections moved by their critics, have a widespread consensus. Much of their appeal resides in the fact that they seem to fit both the *Standard Story* in theory of action and the reductivist view in philosophy of mind, explaining how our actions are up to us (how we can control our conduct) without referring to irreducible agential properties. Mental events play the leading role and, in principle, nothing prevents their reduction to physical states.

However, it is not clear whether these accounts are able to explain control in a proper way. Indeed, it is often held that, both in their libertarian and in their compatibilist interpretations, they are victims of the syndrome of the *disappearing agent*, a version of the more general *luck objection* (Hume 1739; Pereboom 2004, 2012, 2014, 2015; Mele 2006): in the absence of a further explanation of how the choice is up to us, the decision occurs as a result of the causal factors already in place, and there is no way to support ordinary moral responsibility attributions. The fact that the agent might turn out to be a "passive bystander" of a string of mental events represented a serious issue for Frankfurt himself (1988, p. 54). What is doubtful is if identification, or something similar, is sufficient for filling the gap[1]. The problem is that control is hard to reduce to identification with specific mental states: by contrast, it seems that one can control one's choice if one is able to make a decision (at least partially) independently of the motivational force of one's mental states.

The situations characterized by the presence of contrasting motives help to stress this point (cfr. Frankfurt 1988, pp. 47-57). The following is a case characterized by opposed first-order desires: Roger wants to climb the *Mount Rushmore National Memorial* but – since there is a fine that discourages people from climbing – Roger opts for avoiding the risk. Now compare Roger with the unwilling addict. The difference rests on the fact that Roger can control his decision, while the unwilling addict does not have this power. The lack of sameness among the desires is not indicative by itself, but only to the extent that it might reveal the practical inability to exercise control over one's desiderative states. Only in such cases the absence of identification undermines moral responsibility. Even in situations of deep *ambivalence* – in which one is divided between opposing second-order desires/volitions – it is not the lack of identification by itself that does help to illuminate moral responsibility attributions. Consider a less mundane example, a Frankfurt-type version of the story of the *Lady of the Camellias*. Deciding to leave Alfredo under the pressure of his father, Violetta is divided between two opposed second-order desires: she both wants to be moved by the desire to spend her life with Alfredo, and by the desire to help him to have a better life. Being in a condition of ambivalence does not undermine her responsibility. From a phenomenological point of view, Violetta appears to be fully responsible because the decision belongs to her – a reasonable adult woman – independently of her identification with a specific mental state. Moral responsibility attributions depend on the internal structure of her choice in virtue of the fact that she appears to be an agent, who can master different desires, reasons and plans, and whose practical identity goes beyond the sum of her mental states.

[1] Bratman expresses a similar concern: "In some cases we suppose, further, that the agent is the source of, determines, directs, governs, the action and is not merely the locus of a series of happenings, of causal pushes and pulls" (2001, p. 311). Velleman explores the idea of the agent as a master of desire in a state or event-causal framework, looking for a kind of mental state that can play a role functionally identical to the role of the agent: "We must therefore look for mental events and states that are functionally identical to the agent, in the sense that they play the causal role that ordinary parlance attributes to him" (Velleman 1992, p. 475).

However, the idea of an "agent causing an action" does not fit a reductivist event-causal framework, and speaking up for agential properties seems slippery for a variety of reasons. Conceiving the agent as a peculiar substance capable of causally interacting with the physical dimension might not fit the naturalistic vision of the world also in a broadly construed naturalism (De Caro & Voltolini 2010, p. 76).

3. Some Moves towards Compatibilist Agent-causation

One strategy is to replace the event-causal framework with an agent-causal background. The defining claim of agent-causation is that agents are substances capable of causing decisions or intention-formations (Pereboom 2015). The adoption of the agent-causal perspective has traditionally characterized a branch of libertarian views on free will and moral responsibility (O'Connor 2000) but, more recently, it has been also advocated by some compatibilists. For example, Nelkin (2011) and Markosian (1999, 2012) both proposed compatibilist approaches to agent-causation, which deny that it is incompatible with determinism.

Markosian develops a hybrid account, where agent-causation coexists with event-causation inside a materialistic conception. An action is morally free *iff* it is caused by an agent, and the agent is morally responsible *iff* he is the cause of that action. Admitting *double causation* – according to which the very same event can be produced by two independent factors – an action freely produced by an agent can also be produced in an event-causal way. It might be objected that, if it is only the event-causal stream that is deterministic (while the agent-causal one is indeterministic), the account fails to provide a compatibilist version of agent-causation (Pereboom 2015). Otherwise, if both are deterministic and are causing the very same event, either (a) the physical occurrences characterizing the event-causal stream are not sufficient by themselves and the interaction with the agential causal powers should be explained or (b) the physical occurrences characterizing the event-causal stream are sufficient and the agential causal powers appear to be redundant (and, if the physical realm is complete, there seems to be no reason for admitting extra causal powers [Bennett 2003]). One of the burdens of such a view is that the analysis of the structure of the choice-making process – the core of Frankfurt-type compatibilism – partially loses its centrality. No matter the circumstances of choice, the action is free because an agent produces it (Markosian 2012, p. 384). Then Markosian – with the questionable assumption that, if the action is morally wrong, then it has to be morally free – has to admit that also a brainwashed individual (like Patriot Kid, who shoots the president after being kidnapped and manipulated by Martians [Markosian 1999, p. 272]) is morally responsible.

Nelkin instead provides a unified account, in which the only form of causation that exists is substance causation (Lowe 2008), whose effects are determined: given the kind of substance the agents are, and the circumstances in which the action occurs, the choice is deterministically produced. Nelkin adopts a distinction made by O'Connor (2000; cfr. Dretske 1993) between *structuring* and *triggering* causes. While reasons are the structuring causes of a choice (structuring one's propensities), the agent, with his specific causal power, is the triggering cause that settles the final decision. In our story, Violetta, with her specific causal power, makes a choice, which turns out to be determined. A source of doubt is that – once one's propensities have been already structured – it is not clear how the power to settle the (determined) final decision is to be understood in a way that might preserve one's ability to control one's own conduct. Despite some obscurities, deterministic agent-causation might represent a promising path. However, for many – at least without further clarifications – the assumption that the agents cause events remains controversial or even unintelligible, and turns out to be a sort of *ignotum per ignotius* explanation.

4. Lynne Baker's *Reflective-Endorsement*

A different perspective is sketched by Lynne Baker. In her work, Baker addresses directly the problem of moral responsibility recruiting Frankfurt's event-causal framework, which might be revised in the light of an explicit reference to the robust form of the *first-person perspective on agency*, defined as "the capacity to think of oneself, conceived in the first person, as the object of one's thought" (2013, p. XIX), and intended as incompatible with a third-person ontology[2]. Why is this "defining characteristic of persons" (2013, p. 201) – who "can consider the reasons" they "have and choose to act on them" (2013, p. 202) – supposed to improve Frankfurt-type approaches?

[2] "Frankfurt, Velleman, and Bratman [...] all speak of an agent's reflective participation in her action as if reflective participation [...] is compatible with a third-person ontology. Many philosophers do not acknowledge that the first-person perspective presents a problem for scientific naturalism" (Baker 2013, p. XVII).

The most peculiar aspect of Frankfurt's hierarchical view consisted in the idea that, in order to be morally responsible, one should be able to conceive the mental states leading to action as one's own. In Baker's *Reflective-Endorsement*, it is this essential capacity that gives people the limited amount of control – the ability to consider the reasons we have and to act on them – that might save the day for compatibilism[3]. More precisely, an "agent is morally responsible for an action if he endorses the beliefs and the desires on which he acts: When he affirms them as his own [...], he is morally responsible for acting on them" (2013, p. 205). According to Baker, the appeal to the first-person perspective is not to be intended merely as a reference to the practical unity of the subject, but implies an ontological commitment. As mentioned earlier, one intriguing aspect of Baker's proposal is that it is committed to preserve an event-causal framework. But how is the concept of an "event" to be understood? Following Kim, Baker interprets an event as an object's having a property at a time (2007, p. 97). What should be abandoned is rather the reductivist spirit according to which conscious mental events are reducible to physical states. In virtue of being (emergent) upper level properties-instances, mental states are irreducible to lower-level physical properties-instances, turning out to be independently causally efficacious[4].

The relation between the two orders is conceived in terms of *constitution*: given certain *favorable circumstances*, the higher-level properties-instances are constituted by, but not reducible to, the lower level ones, as a cat is constituted by, but not reducible to, the sum of its particles[5]: "The Constitution View, applied to property-instances, allows intentional phenomena to have causal efficacy" (Baker 2011, p. 13). For making sense of the moral realm, some first-person properties should be admitted in our ontology: "Property P is a first-person property if either (1) P entails that whatever exemplifies it has the capacity to interact consciously and intentionally with the environment and/or (2) P entails that whatever exemplifies it can conceive of herself as herself* in the first-person" (Baker 2013, p. 172).

Without considering the traditional objections towards the anti-reductivist program in general, a doubt I wish to explore in the remaining part of the paper regards constitution and its dependence on some favourable circumstances. Irreducible (and causally efficacious) emergent properties are produced by their subatomic constituents, given the presence of the relevant circumstances[6]. Differently from *supervenience* (which is necessary and independent of contextual factors), constitution occurs only if the microphysical constituents are accompanied by the relevant circumstances, so that "although a constituting property-instance does not supervene on its constituting property-instances, it may supervene ultimately on its subatomic constituters *together with* the microphysical supervenience base of all the circumstances in which the instance of the constitution relation obtains" (Baker 2013, pp. 219-220). Since Baker's near naturalism leaves the door open for the truth of the *causal-closure thesis* (*ibidem*), the emergent properties are constituted by their microphysical particles plus the relevant circumstances that, in turn, are also constituted by their microphysical particles. Then (even though that particular lower-level event does not necessitate that particular higher-level event), one might object that the upper-level properties and, in particular, the first-person perspective – the *locus* where the non-biological (Baker 2000, p. 17) discontinuity between human and non-human animals takes place – turn out to be a practical (epiphenomenal?) stance with no really independent causal powers, a lens through which one regards oneself as a unity, but without having a grasp of the ontological reality.

My last concern regards the kind of moral responsibility that is in question. The direction taken by Baker to escape the *disappearing agent objection* is quite promising: the implicit reference to the first-person point of view – which is hidden in Frankfurt's approach – is thus made explicit and it is now possible to account for cases of opposing desires and ambivalence, in which one is in control of one's choice because one refers the opposing mental states to oneself. Nevertheless, even incompatibilists usually do not deny that, also given the truth of determinism, one can endorse one's beliefs and desires, or think about the origins of one's mental states, thus forming *a sort of* first-person perspective, and having the impression of being able to shape the causes of one's choices. However, nothing proves that this picture, which fits a certain phenomenology of agency, is not a *post-factum* illusory reconstruction (despite Baker's denial: "The first-person perspective cannot be acquired by neural manipulation" [2013, p. 202; see also Baker 2006]), or a "center of narrative gravity", to use Daniel Dennett's words (1992). To dismiss these worries, Baker claims that her core concerns diverge, for example, from those that inspire the *theories of confabulation* in cognitive sciences (Carruthers 2013. See also e.g. Wegner 2002; Preston & Wegner 2005; Marraffa & Paternoster 2013), which are mainly focused on the idea that, given for example the *opacity* of introspection (Carruthers 2011), we might be mistaken "about the sources of our first-order beliefs" (Baker 2013, p. 64) or about the content of our inner life. Baker's theory rather concerns the question: "under what conditions can we have beliefs about our beliefs at all?" (2013, p. 64): having a robust first-person perspective, or "conceiving of oneself as having a perspective" (2013, p. 82) is meant to be the basic requisite for having a inner life, something that cannot be understood in terms of a misleading rationalization and self-ascription of mental states. Yet, does my awareness of my inner life – no matter the possible lack of insight into my first-order mental states – represent a strong basis for moral responsibility attributions? The problem with compatibilism does not seem to be that one might be unable to conceive of oneself as oneself. The limit is rather that – once accepted that our choices are determined by factors that are beyond us – we could hardly make sense of the concept of *accountability*[7] that, for many, is what moral responsibility is supposed to be.

Nevertheless, even though it is unlikely that Baker's account proves successful against traditional incompatibilist worries, it opens a thought-provoking line for those who share compatibilist intuitions and, more generally, for those – including myself – who are inclined to think that moral responsibility has (much) to do with identification with motives mediated through practical reasoning.

[3] The *Reflective Endorsement* view is articulated as follows.
(RE) A person S is morally responsible for a choice or action X if X occurs and:
(1) S wills X,
(2) S wants that she* will X [i.e., S wants to will X],
(3) S wills X because she* wants to will X, and
(4) S would still have wanted to will X even if she had known the provenance of her* wanting to will X.
Where the fourth condition specifies that the agent would not repudiate her desires given that she is aware of their provenance, and the "*" identifies the first-person perspective (Baker 2013, p. 204).

[4] Baker vindicates *commonsense causation* (as making something happen, giving rise to something): "An object x (or a property instance) has causal powers if and only if x has a property F in virtue of which x has effects" (2007, p. 98. See also Baker 2011, pp. 12-13). About Baker's emergentism, see instead Baker 2013, p. 220; 2007, p. 237.

[5] About mental causation and constitution, see Baker 1993, 2007, 2013. For a different account of constitution ("*the made up of* relation"), inside a non-reductivist framework, see Pereboom 2011, pp. 135-141.

[6] "Whether or not constitution obtains depends in large measure on the circumstances" (Baker 2013, p. 209).

[7] At least for certain interpretations, the idea that one deserves to be blamed and praised in virtue of the choice one made (cfr. Watson 2004, pp. 260-288; Pereboom 2001, p. XX).

REFERENCES

Baker, L.R. (1993), "Metaphysics and Mental Causation", in J. Heil & A. Mele (eds.), *Mental Causation*, Clarendon Press, Oxford, pp. 75-96;
Baker, L.R. (2000), *Persons and Bodies*, Cambridge University Press, Cambridge;
Baker, L.R. (2006), "Moral Responsibility Without Libertarianism", *Noûs*, 42, pp. 307-330;
Baker, L.R. (2007), *The Metaphysics of Everyday Life*, Cambridge University Press, New York;
Baker, L.R. (2011), "First-Personal Aspects of Agency", *Metaphilosophy*, 42(1-2), pp. 1-16;
Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, New York;
Bennett, K. (2003), "Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It", *Noûs*, 37(3), pp. 471-497;
Bratman, M. (2001), "Two Problems About Human Agency", *Proceedings of the Aristotelian Society*, New Series, 101, pp. 309-326;
Carruthers, P. (2011), *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford University Press, Oxford;
Carruthers, P. (2013), "Mindreading the Self", in S. Baron-Cohen, H. Tager-Flusberg & M. Lombardo (eds.), *Understanding Other Mind: Perspectives From Developmental Social Neuroscience*, Third Edition, Oxford University Press, New York, pp. 467-486;
De Caro, M. & Voltolini, A. (2010), "Is Liberal Naturalism Possible?", in M. De Caro & D. Macarthur (eds.), *Naturalism and Normativity*, Columbia University Press, New York, pp. 69-86;
Dennett, D.C. (1992) "The Self as a Center of Narrative Gravity", in F.S. Kessel, P.M. Kohl & D.L. Johnson (eds.), *Self and Consciousness: Multiple Perspectives*, Lawrence Erlbaum Associates, Hillsdale, NY, pp. 103-115;
Dretske, F. (1993), "Mental Events as Structuring Causes of Behavior", in J. Heil & A. Mele (eds.), *Mental Causation*, Clarendon Press, Oxford, pp. 121-136;
Frankfurt, H. (1988), *The Importance of What We Care About: Philosophical Essays*, Cambridge University Press, New York;
Frankfurt, H. (1999), *Necessity, Volition, and Love*, Cambridge University Press, New York;
Horgan, T. (2007), "Mental Causation and the Agent-Exclusion Problem", *Erkenntnis*, 67, pp. 183-200;
Hume, D. (1739/1888), *A Treatise of Human Nature*, L.A. Selby-Bigge (ed.), Oxford University Press, Oxford;
Lowe, E.J. (2008), *Personal Agency: The Metaphysics of Mind and Action*, Oxford University Press, Oxford;
Markosian, N. (1999), "A Compatibilist View of the Theory of Agent Causation", *Pacific Philosophical Quarterly*, 80, pp. 257-277;
Markosian, N. (2012), "Agent Causation as the Solution to All the Compatibilist's Problems", *Philosophical Studies*, 157, pp. 383-398;
Marraffa, M. & Paternoster, A. (2013), *Sentirsi esistere. Inconscio, coscienza, autocoscienza*, Laterza, Roma-Bari;
Mele, A. (2006), *Free Will and Luck*, Oxford University Press, New York;
Nelkin, D. (2011), *Making Sense of Freedom and Responsibility*, Oxford University Press, New York;
O'Connor, T. (2000), *Persons and Causes: The Metaphysics of Free Will*, Oxford University Press, New York;
Pereboom, D. (2004), "Is Our Conception of Agent-Causation Coherent?", *Philosophical Topics*, 32, pp. 275-286;
Pereboom, D. (2011), *Consciousness and the Prospects of Physicalism*, Oxford University Press, New York;
Pereboom, D. (2012), "The Disappearing Agent Objection to Event-Causal Libertarianism", *Philosophical Studies*, 169, pp. 59-69;
Pereboom, D. (2014), *Free Will, Agency and Meaning in Life*, Oxford University Press, Oxford;
Pereboom, D. (forthcoming 2015), "The Phenomenology of Agency and Deterministic Agent-Causation", in H. Pedersen & M. Altman (eds.), *Horizons of Authenticity in Phenomenology, Existentialism, and Moral Psychology: Essays in Honor of Charles Guignon*, Springer, New York;

Preston, J. & Wegner, D.M. (2005), "Ideal Agency: On Perceiving the Self as an Origin of Action", in A. Tesser, J. Wood & D. Stapel (eds.), *On Building, Defending, and Regulating the Self*, Psychology Press, Philadelphia, pp. 103-125;
Velleman, J.D. (1992), "What Happens When Someone Acts", *Mind*, New Series, 101(403), pp. 461-481;
Watson, G. (2004), *Agency and Answerability: Selected Essays*, Oxford University Press, Oxford;
Wegner, D.M. (2002), *The Illusion of Conscious Will*, MIT Press, Cambridge (MA).

ALAN MCKAY
The Queen's University of Belfast, Northern Ireland
alanmckay33@yahoo.co.uk

CONSTITUTION, MECHANISM, AND DOWNWARD CAUSATION

abstract

I develop an account of ordinary physical causation as productive, causally closed, and operating via mechanisms. This picture entails rejection of Baker's claims that intention-dependent properties are independently causally efficacious and share the lower-level physical causal nexus. However, I suggest that Baker's constitution account has the resources to overcome these difficulties, and that intention-dependent causal relations are constituted by lower-level ones.

keywords

Constitution, manifest image, mechanism, downward causation

1. Introduction

In the final chapter of *Naturalism and the First-Person Perspective* (Baker 2013, pp. 207-234), Lynne Rudder Baker builds upon the causal arguments developed in her earlier work (e.g., Baker 2000, 2007) as part of her constitution account of reality. In that account, Baker distinguishes those objects and properties that are intention-dependent (ID) from other, lower-level, non-ID objects and properties. ID properties are either propositional attitude properties – believing, etc. – or properties whose instances presuppose that there are entities that are bearers of propositional attitudes (Baker 2007, pp. 11-13), such as the property of being an economic recession. ID objects are either such entities (i.e., persons) or objects, like houses or computers, whose existence presupposes the existence of the former. ID objects and properties are constituted, in favourable circumstances, by the lower-level, non-ID ones. However, Baker (2013, p. 217) also contends that, like all properties and property-instances, mental and other ID properties are nevertheless *physical*.

It is central to Baker's anti-reductive causal arguments that ID causal property-instances are real and capable of independently causally affecting the objects and properties of the non-ID, physical world. Thus she claims that there is downward causation, whereby mental contents have physical effects, and she presents empirical data which she believes support this claim (Baker 2013, pp. 220-233). Baker's theoretical argument for downward causation is based on two claims that are, I argue, false and in any case mutually incompatible; firstly, that the causal powers of ID property-instances are independent of those of their constituting property-instances (Baker 2013, pp. 216), and secondly, that ID and lower-level causes, both being physical property-instances, belong in a single causal nexus, allowing inter-level causation (*ivi*, pp. 217; 231-233). Further, I will argue that on an account, which I will develop, of causal relations amongst the objects that make up the furniture of the everyday world, the idea that mental content, *qua* content, has effects in the physical world is incoherent. Nevertheless, I will claim, Baker's constitution account itself contains the resources to provide a robust and satisfying account of mental causation.

2. The Constitution Account and Independent Causal Efficacy

Clarification of my proposals requires a brief review of the relevant aspects of Baker's constitution account. Constitution, according to Baker, is a relation of unity without identity, a category that lies between identity and separate existence without being either. The constitution account, which presupposes that reality contains multiple hierarchical ontological levels or layers, is developed most fully in Baker (2007) as the basis of a defense of the reality of everyday objects and properties and their causal powers. Here I discuss only property constitution, according to which an instancing of a lower-level property in an object constitutes, in the presence of favourable circumstances, an instancing of a higher-level, for example ID, property in that object. This higher property-instance acquires, in virtue of its constitution in the favourable circumstances, novel and irreducible causal powers not possessed by its constituting property-instance alone.

Favourable circumstances, in Baker's technical sense (Baker 2007, pp. 160-161), are extrinsic or relational properties that must be instantiated if the constituting property is to constitute the higher property in question. So, to introduce one of Baker's examples, an instance of hand-raising, in favourable circumstances, constitutes an instance of voting. In this case, the favourable circumstances comprise the hand-raising's being deliberately performed as a voting, by a competent person, in an environment in which there is agreement, within a suitable background cultural milieu, that a ballot is in progress in which hand-raising counts as voting. In different circumstances the same hand-raising might have constituted something else, say a call for attention, or nothing at all.

Crucially, Baker insists also that the identity of the constituting thing is *subsumed* in the identity of what it constitutes. "As long as *x* constitutes *y*, *y* encompasses or subsumes *x*" (Baker 2000, p. 33), so that "*x* has no independent existence" (*ivi*, p. 46). The hand-raising *is* the voting – the "is", not of identity, composition, or predication, but of constitution.

Baker's claim that constituted property-instances, such as being a voting, are endowed with novel and irreducible causal powers is encapsulated in the Principle of Independent Causal Efficacy (ICE) (e.g., Baker 2013, pp. 216):

> An irreducible higher-level property-instance (x's having F at t) has independent causal efficacy if and only if
>
> (1) x's having F at t has an effect e, and
> (2) x's having F at t would have had the effect e even if its constituting property-instance had been different, and
> (3) x's having F at t confers causal powers that could not have been conferred by its constituting property-instance alone.

Baker (2007, pp. 115-116) offers an example in support of (ICE): Let

> V be Jones's voting against Smith at t
> P be Jones's hand going up at t
> V* be Smith's getting angry at Jones at t'
> P* be Smith's neural state at t'
> C be circumstances that obtain at t in which a vote is taken by raising hands
> Suppose V is constituted by P and V* by P*.

In the example it is assumed as a premise that Jones's voting *causes* Smith's anger. Baker's (2007, pp. 106) justification of this assumption, on the grounds of the practical indispensability of such causal claims in everyday life, is a key motivating factor in her rejection of Jaegwon Kim's arguments against non-reductive physicalism, and especially of his principle of causal/explanatory exclusion (Kim 1993, pp. 250; 1998), which states that there is no more than one complete and independent cause (or causal explanation) of any event. If Kim's arguments are accepted, Baker points out (2007, pp. 106-110), this would threaten not only the independent causal efficacy of mental content but also that of a huge range of non-mental ID properties, such as being a driver's licence or being a delegate, and for her this amounts to a *reductio ad absurdum* of Kim's position.

Baker claims that V's causing V*, in the example, is independent of any lower-level causal relation, thus vindicating (ICE), since, first, V could have been constituted differently, for example if votes were cast electronically, and still have caused V*, and second, although the causal powers of P alone are purely lower-level, P's constitution of V in favourable circumstances gives V the new power of causing Smith's anger.

The notion of cause that underpins Baker's claims here is a metaphysically undemanding one. Essentially, on her view, wherever a causal explanation is available and a counterfactual dependence of an explanandum on an explanans can be shown, a cause is also to be found (Baker 1993). I will now put forward an account of causal relations among the ordinary physical objects and substances that comprise our world that, I believe, calls Baker's account of ID causation into question.

3. Causation in the Manifest Image

Baker's rejection of the principle of causal/explanatory exclusion (Baker 2007, pp. 99-102) trades on the possibility that a fundamental microphysical causal level – the level at which true causal relations must be located, according to the exclusion principle – may not exist. I will not try to counter this argument because, I contend, this hypothetical level is not the appropriate place in which to look when we are seeking a clarification of mental and ID causation in the everyday world whose existence Baker's arguments in *The Metaphysics of Everyday Life* (2007) are aimed at establishing.

We should look, rather, at causality as it concerns the ordinary objects, with their properties and relations, that make up the perceptible, non-ID macroscopic world in which we live, together with some of its well-understood extensions into the microscopic. This is the world that corresponds to what Sellars (1991, pp. 1-40) called the manifest image of man in the world. My claim is that no matter how problematic the notion of causation may be at a fundamental level, there exist objectively real causal relations among these observable physical entities, 'objectively' being understood in Baker's (1995, pp. 232-236) sense of *recognition-independence*, in that facts about these causal relations generally do or do not obtain independently of any individual's or community's beliefs about them.

Sellars himself opposes the scientific image to the manifest, and claims that the occupants of the former are the only true existents. But, as many have pointed out, this very claim, as well as all other claims, is made from the standpoint of the manifest image. Baker's argument for the reality of the world of macroscopic objects is based on practical necessity, her idea that "metaphysics should not swing free of the rest of human enquiry ... [it] ... should be responsive to reflection on successful cognitive practices, scientific and nonscientific" (Baker 2007, p. 15). Philosophers such as McDowell (2000) and Davidson (2001) further argue for a *transcendental* link between our very possession of the conceptual capacities we do and the existence of the world revealed to us through perception.

I would argue, then, that the manifest image is the natural home of our causal claims and beliefs about the world, and that it is within the manifest image that we should expect to locate the relevant distinctions among and constraints on those beliefs. We have, I suggest, a deep and intuitive understanding of what is and is not causally possible within the manifest world. We know, for example, that macroscopic objects cannot change their spatial location from *a* to *b* without passing through space between *a* and *b*. As de Muijnck puts it (2003, p. 46), if we cannot find any physical influences connecting alleged cause and effect, we would sooner suspect coincidence than "action at a distance" – that is, than some kind of magical cause-like process. I will use these notions to distinguish a basic category of causation within the manifest image that I call "manifest physical causation". Further, I contend that our 21st Century manifest image includes objects, properties, and relations belonging to the special physical sciences, as in the biochemical example in the next section. This claim is justified, I believe, because even though such things as genes and neurotransmitters are visible only by special techniques, not only are their existence and properties so well-established empirically as to be effectively beyond doubt, but they clearly participate in the same causal nexus as more familiar, macroscopic entities.

Ordinary causal-explanatory claims, descriptions, and explanations contain multiple instances of the use of “cause”, “because”, and their cognates which, when they cite causes and effects, move freely among mental, non-mental ID, and non-ID items. In this everyday causal discourse we usually do not distinguish either between causation and causal explanation (Beebee 2004, p. 293), or among events, states, objects, facts, or negative facts, as causal *relata*. But when we unpack this causal discourse, I will argue, we can distinguish a more basic category of causal statement. Causal claims that I categorize as manifest physical, like

> a lightning strike caused the forest fire, or
>
> local electrical depolarization of the axonal membrane causes opening of voltage-gated sodium ion channels

are distinguishable, I claim, from ID and mixed ID/physical causal claims such as

> excessive sub-prime mortgage lending caused the recession,
>
> he purposely threw the ball that smashed the window, or
>
> human economic activity causes climate change

in a number of crucially important ways. It is important, moreover, to emphasise that our understanding of these differences is grounded in our intuitive grasp, based upon shared experience, of how things generally work in the non-intentional world around us.

Firstly, manifest physical causal statements are free of allusions to normativity or related properties that are connected with our interests, such as meaningfulness or goal-directedness. Secondly, as remarks such as de Muijnck’s, above, suggest, we have every reason to think that these causal relations form a single, closed causal nexus. My inclusion of an example from neurophysiology in the category of manifest physical causation is justified, I believe, because we cannot nowadays seriously doubt the existence of such entities as neurons or axons, or that their properties are components of a single shared causal nexus, even though they are not strictly part of the world of the manifest image in its pre-scientific form. While it is true that our understanding of special physical sciences such as neurophysiology probably does not reflect the nature of reality as postulated by fundamental physics, nevertheless within the context of the manifest image, this understanding is *homonomic*, in Davidson’s (1980, p. 219) sense, with our intuitive grasp of the workings of the macroscopic world. And this understanding, applied to, say the workings of mechanical, biological, or meteorological processes, includes the tacit conviction that they proceed entirely without any influence from outside the physical causal nexus. Even when we consider human agency, whatever our view of mental causation, Tyler Burge is surely right that we do not think of mental causes “on a physical model – as providing an extra ‘bump’ on the effect” (Burge 1993, p. 115).

4. Manifest Physical Causation: Production and Mechanism

On my account manifest physical causation is causation in a *productive* sense. Thus when a manifest physical causal relation is instanced we understand that there must occur a transfer of energy of some kind – mechanical, electromagnetic, or chemical, say. This implies, firstly, that these causal relations are instantiated in virtue of *intrinsic* properties of the causes, and secondly, that an appropriate kind of *spatio*-temporal connection must exist between cause and effect (Hall 2004).

In contrast, the criteria by which we identify ID or mixed ID/physical causal relations are much less rigorous, being mainly based on the requirement that there be a counterfactual *dependence* of effect upon cause. Manifest physical causes, of course, also show counterfactual dependency, but the difference is that

in their case the counterfactuals are grounded in properties of the manifest physical world.

Wim de Muijnck (2003) and Ned Hall (2004) acknowledge the differences between the dependence and production accounts of causation and believe that they mark an unavoidable duality in our concept of causality (de Muijnck 2003, pp. 41-50). Each of these authors independently claims that we need *both* concepts because there are some imaginable causal scenarios, such as pre-emptions, which resist analysis in terms of counterfactuals, and others, such as instances of causation by omission, that resist analysis in terms of production; thus, it is claimed, neither can provide a univocal account.

The biggest barrier to acceptance of the productive account has been the problems of causation by omission (or disconnection) and causation *of* omission (or prevention). For example, Schaffer argues that “causation by disconnection is causation full force” (Schaffer 2000, p. 289). The production approach cannot accommodate causation by disconnection, he claims, since the latter “involves no persistence line between disconnector and effect, but rather the severing of one” (*ivi*, p. 291). The hallmarks of productive causation, intrinsicality and *spatio*-temporal connection, are absent. Schaffer points out, for example, that when a victim is shot through the heart, the cause of death is *prevention* of oxygen from reaching the brain.

I would argue, however, that this merely seems to be a case of causation by disconnection. The example is a contextual and interest-bound description of manifest physical events, framed so as to meet our explanatory needs. If we analyze the process, not as a death by shooting, but at a lower, or simpler, level of description – if, that is, we bracket our natural tendency to think of the life-death contrast as the all-important explanandum, we find we can describe the process in terms of changes in intracellular metabolism without alluding to disconnections or omissions at all. I claim that all instances of manifest physical causation are capable of description purely in productive terms.

The reason references to phenomena like omissions and preventions feature in descriptions of manifest physical causal systems is that when those systems’ physical parts are arranged in suitable ways they constitute causal *mechanisms*. Glennan defines a mechanism as “a set of systems or processes that produce phenomena in virtue of the arrangement and interaction of a number of parts” (Glennan 2009, p. 315) and goes on, “discovering a mechanism is the gold standard for establishing and explaining causal connections” (*ibidem*). There seems to be increasing recognition that the study of mechanisms, rather than discovery of laws, is an appropriate line of inquiry for the philosophy of the special physical sciences. Craver and Bechtel (2007) give an account of mechanisms in neurophysiology that emphasizes the contrast between intralevel causation and interlevel constitution. Although their notion of constitution is not Baker’s technical one, there are clear parallels; the suitable arrangement of parts might be said to be the favourable circumstances whereby an aggregate of parts constitutes a mechanism.

I claim, then, that manifest physical causation is norm-free, causally closed, productive, intrinsic, and involves the operation of mechanisms. In contrast, an ID causal relation such as Jones’s voting making Smith angry is neither norm-free, productive, intrinsic, or mechanistic in anything like the same sense, and in light of this it seems that Jones’s voting, as a higher, constituted, and *ex hypothesi* independent causal power, has no place in the manifest physical causal nexus.

5. Manifest Physical Causation and Independent Causal Efficacy

Baker’s argument, above, for the independent causal efficacy of constituted, ID property-instances, appears valid, but depends on acceptance, on the basis of reasons that are external to the argument, of the premise that Jones’s voting, V, is indeed the cause of Smith’s anger, V*. Yet I think many would agree that the validity of this premise is just what is at issue. Can the argument itself establish its validity?

Baker claims that V’s causing V* is independent in the sense that it does not depend on any lower-level causal process. But such a process undoubtedly exists; Jones’s hand-raising, P, causes light rays to travel to Smith’s retinae, whence neural events are initiated that lead to the instantiation of Smith’s neural state, P*. Call this causal chain or mechanism P&ae's causing P* (ae for additional events). A causal relation between V and V*, however, cannot be inferred from P&ae's causing P*; nothing at the ID level corresponds to the manifest physical, mechanistic causal chain component "ae". And on the constitution account, the instantiation of P*, caused by P&ae, guarantees, in favourable circumstances, that of V*, so that, from the perspective of the argument, there seems to be no *need* for V to cause V*.

Further, *ex hypothesi*, Jones's hand-raising, P, and his voting, V, are both physical property-instances. So V's independent, irreducible power of causing V* must be a *physical* causal power. But in the constitution sense, V *is* P – it is just P in the presence of certain relational properties, which, according to the account, confer on it extra physical causal powers. If the account of manifest physical causation I have given is correct, Baker's account leaves the nature and origin of these new physical powers, and how they could be efficacious in the same causal nexus as the lower-level powers, quite mysterious.

I conclude that Baker's version of higher causal efficacy cannot work. Her insistence that ID property-instances are physical, and hence that ID causation is of the same basic kind as lower-level causation, obscures deep differences between the two. Baker's claim that all property-instances are physical seems to be based upon the assumption that the constitution relation of unity without identity dictates that constituted entities be of the same general kind as their constituters (Baker 2007, p. 161). I think, however, that the relational qualities that ID property-instances acquire *via* the favourable circumstances of their constitution are such that to insist that these instances are physical, despite their lacking the marks of manifest physicality and causality that I have identified, is just to deprive the term "physical" of any useful discriminatory ability.

6. Constituted Causation

Nevertheless, I agree with the commonsense view that, say, Jones's voting does indeed cause Smith's anger. A way of protecting our ordinary intuitions about ID causation, I propose, is to claim that not just ID causes and effects, but the causal *relations* between them, are constituted by manifest physical causal relations in favourable circumstances. Thus, on this proposal, the causal relation P&ae's causing P*, in the presence of circumstances that are essentially the same as those favouring the constitutions of V and V*, *constitutes* the causal relation V's causing V*. The former relation just is the latter in the constitution sense of "is", but it is transformed, in the presence of its personal and cultural relational milieu, from a mere manifest physical relation into a vastly enriched, multi-faceted ID causal relation. Further, in line with Baker's constitutional claims and our intuitions, the ID relation *subsumes* the physical one, thus vindicating our claim that it is the *real* causal relation.

ID causation, on this account, belongs in a quite different causal nexus from manifest physical causation, a nexus whose operations are constrained, not by the laws governing energy transfer or physical mechanisms, but by such factors as inference, justification, purpose, and desire. ID and manifest physical causes do not interact directly. Causation is a diachronic, purely *intra*level relation, while the physical and ID levels are connected through the synchronic relation of constitution.

This allows an alternative to Baker's (2013, pp. 220-233) interpretation of an empirical study (Anon. 2000) which found a correlation between hippocampal size and navigation experience in London taxi drivers. Baker claims that the study shows that downward causation occurs between learning, an ID property, and these physical, hippocampal changes. On my account, however, learning is constituted by other neural changes which cause the hippocampal effects, and this causal relation constitutes a purely ID causal relation between the learning and increased navigational ability.

REFERENCES

Anon., *Taxi Drivers' Brains "Grow" on the Job*, Internet Report, BBC News, 2000, http://news.bbc.co.uk/2/hi/science/nature/677048.stm;

Baker, L.R. (1993), "Metaphysics and Mental Causation", in J. Heil & A. Mele, (eds.), *Mental Causation,* Oxford University Press, Oxford, pp 75-95;

Baker, L.R. (1995), *Explaining Attitudes*, Cambridge University Press, Cambridge;

Baker, L.R. (2000), *Persons and Bodies: A Constitution View*, Cambridge University Press, Cambridge;

Baker, L.R. (2007), *The Metaphysics of Everyday Life*, Cambridge University Press, Cambridge;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;

Beebee, H. (2004), "Causing and Nothingness", in J. Collins, N. Hall, L. Paul (eds.), *Causation and Counterfactuals*, MIT Press, Cambridge, Mass., pp. 291-308;

Burge, T. (1993), "Mind-Body Causation and Explanatory Practice", in J. Heil & A. Mele, (eds.), *Mental Causation,* Oxford University Press, Oxford, pp 97-120;

Craver, C., Bechtel, W. (2007), "Top-down causation without top-down causes", *Biology and Philosophy*, 22, pp. 547-563;

Davidson, D. (1980), *Essays on Actions and Events*, Clarendon, Oxford;

Davidson, D. (2001), "The Emergence of Thought", in *Subjective, Intersubjective, Objective*, Clarendon, Oxford, pp. 123-134;

De Muijnck, W. (2003), *Dependencies, Connections, and Other Relations: A Theory of Mental Causation*, Kluwer, Dordrecht;

Glennan, S. (2009), "Mechanisms", in H. Beebee, C. Hitchcock, P. Menzies (eds.), *The Oxford Handbook of Causation*, Oxford University Press, Oxford, pp. 315-325;

Hall, N. (2004), "Two Concepts of Causation", in J. Collins, N. Hall, L. Paul (eds.), *Causation and Counterfactuals*, MIT Press, Cambridge, Mass., pp. 225-276;

Kim, J. (1993), *Supervenience and Mind*, Cambridge University Press, Cambridge;

Kim, J. (1998), *Mind in a Physical World*, MIT Press, Cambridge, Mass.;

McDowell, J. (2000), "Experiencing the World", in M. Willaschek (ed.), *John McDowell: Reason and Nature*, Transaction Publishers, New Brunswick, pp. 3-17;

Sellars, W. (1991), "Philosophy and the Scientific Image of Man", in *Science, Perception and Reality*, Ridgeview, Atascadero, Cal., pp. 1- 40.

TREASA CAMPBELL
New Europe College, Bucharest
treasacampbell@gmail.com

A HUMEAN INSIGHT INTO THE EPISTEMIC NORMATIVITY OF THE BELIEF IN THE SELF

abstract

Baker (2013) showcases the complexity of responses on both sides of the debate concerning the ontological status of the first-person perspective. This paper seeks to orientate the debate about the first person perspective away from an existence problem and back to a justified belief problem. It is argued that the account of our belief in the self, which emerges from Hume's descriptive epistemology, opens up the possibility of attributing a form of non-evidential justification to belief in selves.

keywords

First-person perspective, Hume, non-evidential justification, naturalism

In her book *Naturalism and the First-Person Perspective* Lynne Rudder Baker (2013) surveys both reductive and non-reductive versions of naturalism concluding that none of these versions of naturalism recognises first-person properties in their ontological inventory of what exists. Furthermore, she argues that the attempts to naturalise these properties through reduction or elimination fail. A first-person perspective is a conceptual capacity to attribute first-person references to ourselves. For Baker it is this capacity to think of ourselves in this first-personal way that distinguishes us persons from other beings. Baker argues that this capacity to form complex first-person thoughts has implications for a naturalistic ontology. For example, in reducing cognitive first-person perspective to a complex phenomenal first-person perspective naturalist thinkers such as Metzinger (2004) find no place for the subjects of experience in their ontology. On this view, the belief that a self carries out the act of cognitive self-reference is not epistemically justified. Baker's book showcases the ingenuity and the complexity of responses on both sides of the debate concerning the ontological status of the first-person perspective. While recognising the value of such debates, this paper seeks to advance a different approach to the understanding of the human capacity of generating self-concept beliefs. Instead of approaching the problem through the confines of ontological naturalism, it will be argued that naturalism within epistemology provides the resources to facilitate the affirmation of belief in the self.

In Section One, drawing on elements of Hume's descriptive account of belief formation, it will be demonstrated that belief in the self belongs to a class of beliefs called 'natural beliefs'. This class of beliefs contains beliefs that are universal and unavoidable features of how we engage in the world. Descriptive accounts, which explain our capacity to think of ourselves in this first-personal way, can provide us with explanation, even with instrumental justification, but have widely been thought to fall far short of anything resembling epistemic justification[1]. Baker, for example, sees no philosophical relevance in appealing to descriptive accounts of the mechanisms underlying the first-person perspective. Even if the sub-personal sciences can provide us with knowledge about the mechanisms underlying the first-person perspective, Baker strongly rejects the idea that knowledge of such

[1] The kind of epistemic justification appropriate within a naturalised epistemological landscape is an issue of considerable contention. Given the difficulties in merging the descriptive with the prescriptive, many contemporary naturalists are willing to abandon the idea that there are any epistemic norms (See Papineau 1993; Churchland 1995; Knowles 2003).

mechanisms can supplant or replace knowledge of the phenomena that they make possible. Section Two seeks to show that in connecting mechanisms to justification, the descriptive account provided in Section One opens up new paths to re-evaluate the claim that the belief that a self/person carries out the act of cognitive self-reference is not epistemically justified. Specifically, it will be argued that the account of our belief in the self, which emerges from Hume's descriptive epistemology, opens up the possibility of attributing a form of non-evidential justification to this belief. This sets aside the ontological question and settles instead for a naturalised epistemic justification of our belief in the self.

1. Hume's Descriptive Insight into Belief in the Self

Since for many an ontological worldview lies at the heart of naturalism's philosophical project, it is not surprising that the debate concerning the first-person perspective is played out in terms of naturalistic ontology. Yet the stated goal of this paper is to orientate the debate about the first-person perspective and its implications away from the existence problem and back to a justified belief problem. Kornblith (1985) emphasises that the naturalistic approach to epistemology marks itself out from the traditional view by insisting that the question 'How ought we to arrive at our beliefs?' cannot be answered independently of the question 'How do we arrive at our beliefs?'[2]. If we place this commitment at the heart of our investigations into the first-person perspective we will see that it is possible to open up new pathways for justifying our belief in the self. In turning away from ontology we can still contribute to discussions about the first-person perspective. The prize is no longer the ontological trophy of an affirmative existence claim but rather an epistemological award of the status of justified belief.

As we will see, some elements of Hume's descriptive account of belief formation provide insights into the nature of belief in the self. In opening the floodgates to naturalist readings of Hume, Kemp Smith argued that the traditional sceptical interpretation of Hume overlooked what was basic to Hume's positive philosophical achievement, namely a new doctrine of 'Natural Belief'. Though Hume never used the expression 'natural belief', there is general agreement that such a class of entities exists for Hume, and discussion of them has become central to those who consider Hume's main concern to be the revealation of non-intellectual resources, located within our human nature, which enable us to interpret and respond to our experience (Garrett 1997; Kemp Smith 1983; Stroud 1977; Strawson 1989). Natural belief in Kemp Smiths' strict sense is a belief which is not supported by evidence or philosophical argument, is determined by psychological propensities of human nature, and is irresistible. Kemp Smith gives the following set of natural beliefs: belief in the body, in causal action, in the identity or unity of the self and in the external world. These phenomena exposed by Hume, are not the product of reasoning, they are unavoidable, universally held, and necessary as a precondition of action (McCormick 1993, p. 106). They cannot be justified rationally but are impossible to give up; no amount of reasoning can eliminate them.

It is not simply that they are beliefs which are immediate and unreflective, since this would only mark them out from those beliefs that are based on reflection. We have many unreflective beliefs that are best classed as irrational beliefs. What marks out 'natural beliefs' is that they are unavoidable and universal. But in what sense are natural beliefs unavoidable and universal? Gaskin explains what it means for a belief to be unavoidable in terms of the belief being "a necessary per-condition of action" (Gaskin 1974, p. 286). Emphasising Hume's claim that such beliefs are "inseparable from the species", McCormick (1993) characterised the universality criterion in terms of those beliefs "which *necessarily* arise given the kind of creatures we are" (McCormick 1993, p. 107). It is clear that the vast majority of beliefs will fail to satisfy these stringent criteria. Indeed, most would not satisfy one of them, and as a result the set of beliefs which satisfies the criteria for being 'natural' is extremely small.

[2] For example the requirement of total evidence cannot be implemented given our capacities for information processing (see Baç 2007).

Although this small group of beliefs is not the result of a conscious rational assessment of evidence, common experience reveals that they cannot be dislodged except in brief moments of "philosophical melancholy and delirium" (T 175; T1.4.7.9; SBN 269)[3]. But Hume explains that this is not enough to discount them as universal and unavoidable. Hume's work famously demonstrated that it is possible to doubt the existence of the external world when one is engaged in deep reflection. However, Hume argues that it is impossible for this doubt to last and that this is why no amount of philosophy can entirely eradicate the belief. He calls for us to see these phenomena as instinctual features of our being which are "inseparable from human nature, and inherent in our frame and constitution" (T 371; T 3.3.1.17; SBN 583) and as such they are indispensable despite their lack of rational grounds. Hume himself never used the term 'natural beliefs' to refer to the small number of unavoidable, indispensable and irresistible mental features which he discusses. However, at key points in his texts, when he deals with the phenomena which commentators have termed 'natural beliefs', Hume chooses to use the term 'natural instincts'[4].

> It seems evident, that men are carried, by a *natural instinct* or prepossession, to repose faith in their senses [...] (EHU 113; EHU 12.1.7; SBN 151; my italics).
>
> There is a great difference betwixt such opinions as we form after a calm and profound reflection, and such as we embrace by a kind of *instinct or natural* impulse, on account of their suitableness and conformity to the mind. (T 142; T1.4.2.51; SBN 214; my italics).

In regularly referring to 'natural instincts', Hume continually highlights the innate primacy of the phenomena to which he is referring. We are asked to let go off the idea that these phenomena are beliefs and accept them in their true form as "a species of natural instinct, which no reasoning or process of the thought and understanding is able either to produce or prevent" (EHU 39; EHU 5.1.8; SBN 47).

What Hume has revealed are universal, unavoidable instincts in accordance with which all experience is processed. That I operate in the world as if I am a continuous and distinct person, is not the consequence of any belief which I affirm. Rather, it is a result of universal, unavoidable capacities to which my humanity binds me. It is a fundamental component of how we operate in the world and of how we form beliefs. We have the instinctual capacity of forming beliefs about the self, and this capacity is constitutive and regulative of the way in which we think about ourselves and about the world. Although the findings of Hume's descriptive account identify natural beliefs to be non-rational, to call these beliefs either irrational or unreasonable is problematic given that they are indispensable for human action.

The descriptive account of mechanism which Hume is engaged with would today be the remit of the cognitive scientist; indeed Fodor describes Hume's *Treatise* as "the foundational document of cognitive science" (Fodor 2003, p. 134). Such descriptive accounts are seen as having no bearing on the philosophical discussions of the first-person perspective. The reason for this is clearly stated by Baker (2013) when she characterises the interest of the cognitive scientists in the first-person perspective as consisting not in eliminating or reducing it but in ascertaining its reliability as a cognitive faculty. Regardless of the outcome of such an assessment these investigations cannot show how impersonal science can accommodate it. Baker has indicated that her concern is upstream from such psychological and epistemological matters. But Hume's descriptive account here is not concerned with the contents of our inner lives, or with how reliable we are in reporting our reasons for thinking as we do. We will see in Section Two how the descriptive findings impact on the normative options

[3] Abbreviations used for works by David Hume:
- T *A Treatise of Human Nature* (2011), D. Fate Norton, M.J. Norton (eds.), Oxford University Press, Oxford.
- EHU *An Enquiry Concerning Human Understanding* (2006),T.L. Beauchamp (ed.), Clarendon Press, Oxford.

[4] I have discussed this shift from belief to instinct in relation to causal belief in Campbell 2006.

we can appeal to. In identifying universal and unavoidable aspects of how we engage in the world, Hume's descriptive account of belief formation points to psychological and epistemological matters which are central to the question of 'Under what condition can we have beliefs about our beliefs at all?'. However, Hume's main concern is not what is required for us to have self-concept beliefs, but rather, given that we universally and unavoidably do have them, what are the implications for epistemic norm building? How are we to get from the descriptive findings of an investigation into belief generation to normative recommendations? This will be explored in the following section.

2. Normative Recommendations

In order to carry any kind of force, a normative philosophy needs to be based on the realities of how human beings do in fact operate in the world. That I universally and unavoidably operate in the world as if I was a self has implications for the normative status of the belief in the self. There is a real sense in which a description of our cognitive abilities is essential to establish any genuine epistemic norm. Harold I. Brown, in his article "Psychology, Naturalized Epistemology and Rationality", characterised the danger of discarding such descriptive findings as follows:

> If we attempt to proceed *a priori* we may well end up with norms that have no legitimate force for human beings because they make demands on us that we cannot possibly fulfill. In other words, we need an account of the *appropriate epistemic norms* for human beings; an account of what is normatively rational, requires a prior account of what is rational in the descriptive sense – an account of what cognitive abilities human beings have available (Brown 1996, p. 20).

According to Hume, natural beliefs are universal unavoidable features of how we operate in the world; as such, any account of appropriate epistemic norms must incorporate them. As a specific feature of human beings, these natural beliefs determine how we pursue the epistemologist's tasks. If we attempt to construct norms in a vacuum, disregarding such permanent and irresistible aspects of how we engage in the world, then the norms we establish will have no import for human beings. Goldman also makes it clear that cognitive science is relevant to certain epistemological questions, stating that "to the extent that human epistemic attainments critically depend on human cognitive endowments, those endowments are relevant to epistemology" (Goldman 2002, p. 146).
In the introduction to the *Treatise* Hume goes further, emphasising the importance of providing a 'science of man', since there is no question of any importance which "can be decided with any certainty, before we become acquainted with that science" (T 4; T Intro; SBN xvi). We cannot insulate our understanding of the justification of our belief in selves from the implications of answers to the question 'How do we arrive at our beliefs?'. The effect of identifying universal and unavoidable features of how we engage in the world must ripple out into how we form normative theories in this area. As Brown states, "epistemic norms that are based on a particular account of our cognitive abilities become suspect if that account is rejected, and norms that require us to do what is beyond our capabilities are surely unacceptable" (Brown 1996, p. 31).
The natural beliefs themselves and many of the content specific beliefs they give rise to are often cited by Hume as having a clear instrumental value. For example, he states that if we jettison our customary transition from causes to effects, a foundation of all our thoughts and actions, we would immediately "perish and go to ruin" (T 148; T 1.4.4.1; SBN 225). In doing so, he has provided a clear end for any chain of means-end reasoning seeking for instrumental normativity for this natural belief[5]. Indeed, Audi (2002) suggests that "broadly Humean versions of instrumentalism are among the most plausible contenders to represent instrumentalism as a contemporary naturalistic position in the theory of practical reason" (Audi 2002, p. 235). Contemporary advocates of naturalised epistemology

[5] Instrumental approaches to the problem of induction such as Reichenbach's (1963) bestow our faith in induction with a kind of practical rationality.

frequently commit themselves to instrumental teleological theories of normativity[6]. But is such instrumental reasoning trivial and inadequate as a normative theory? There is a difference between forming and retaining beliefs for epistemic reasons and forming and retaining beliefs for instrumental or pragmatic reasons. Even if normativity can be retained by appealing to instrumental norms contingent upon our aims, the instrumentalist still requires an account of the normative force of those aims. Thomas Kelly has argued that "one cannot immunize oneself against the possibility of acquiring reasons for belief by not caring about the relevant subject matter" (Kelly 2003, p. 628). The realms of epistemic and pragmatic justification operate in accordance with different requirements. While pragmatic responses may establish that our natural beliefs have a kind of practical rationality or instrumental justification, it cannot establish grounds in an epistemic sense. On the pragmatic side, it is clear that we do form beliefs, but without epistemological justification it remains unclear if we should form beliefs in this way. If, as Hume describes it, we are absolutely and necessarily determined to follow our natural beliefs, then there seems to be little grounds for the substitution of 'do form beliefs' with 'should form beliefs'. Hume's difficulty is with epistemic justification, not with instrumental justification.
Given that our descriptive account has revealed that these natural beliefs are unavoidable there might be a temptation to appeal to some form of 'ought implies can' justification for such beliefs. While noting the application of the principle of 'ought implies can' in ethics, Weintraub (2003) questions the validity of its use in epistemological assessments, arguing that, when it comes to epistemic criticisms, 'ought implies can' is not a plausible precept. Although the language of blame (with terms like responsibility, culpability and reproach) is present in many formulations for epistemic justification[7], Weintraub argues that there is no reason to equate 'epistemically unjustified' with 'morally blameworthy'. On this reading, one can be epistemically unjustified without being morally blameworthy. As Weintraub states

> A person who is psychologically bound to believe is absolved from (moral) guilt as is a person who is compelled to perform some action. But if he believes 'compulsively', and cannot be swayed by reason, he is deemed irrational, the more so the stronger the grip of his compulsion (Weintraub 2003, p. 371).

In those cases in which I am compelled to believe without the required evidence, I may not be morally blameworthy but I still remain epistemically unjustified.
In emphasising our inability to sustain doubt in natural beliefs, Hume's account opens up a more promising approach to the justification of natural beliefs. According to the descriptive account Hume provides, natural beliefs are placed beyond doubt. In the case of the capacity to attribute first-person references to ourselves this capacity delineates the scope of any engagement in the world. This understanding of the belief in the self as universal and unavoidable opens up the possibility of assigning to it a form of non-evidential justification. In the contemporary epistemological landscape we can find advocates of the Wittgensteinian notion of a "hinge proposition" also marking out propositions that are neither true nor false but cannot be coherently doubted. Drawing on a line of thought extracted from Wittgenstein's *On Certainty* (1969, §§ 341-343) hinge epistemology has sought to address sceptical challenges to the epistemic credentials of our beliefs of hinge propositions. Wittgenstein wrote that:

> The questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, are as it were hinges [die Angeln] on which those turn (OC 341).

[6] For example the following thinkers all adhere to some variant of the idea that the normativity of epistemology is simply the normativity of instrumental reason: Kitcher 1992, Kornblith 1993, Laudan 1990.
[7] See BonJour (1985, p. 8) and Kornblith (1982, p. 243).

Teasing out such fragments has led to many different readings of how we are to understand the nature, role and justification of hinge propositions.
One prominent approach accepts that hinge propositions cannot be evidentially justified but appeals to non-evidential warrants. In this context Crispin Wright (2004) has distinguished between ordinary evidential justification and non-evidential justification which he calls ‘entitlement’. This approach expands the narrow notion of epistemic rationality, which confines it only to evidentially warranted propositions. It would require another paper to fully trace the various formulations of hinge propositions and I do not have the space here to provide a detailed account of this debate. The goal of the paper is to demonstrate that if we accept Hume’s descriptive account of belief in the self as a natural belief then this vein of normative argumentation is opened up. We then have the prospect of developing a non-evidential justification for our belief in the self. While much work remains to be done in advancing this line of argument, it nevertheless holds out the prospect of not just insulating belief in the self from scepticism but also of placing it on a knowledge footing. This can be achieved circumventing the issue of ontology.

3. Conclusion

The empirical findings of Hume’s investigation into our belief-forming mechanisms conclude that the belief in the self is a natural belief. Regardless of which ontological story we tell about the first-person perspective, if human beings universally and unavoidably function as persons, as exemplified by their capacity to form individual content specific self-concept beliefs, than our epistemology must take this into account in assessing the validity of self-concept beliefs. This is a case in which empirical findings about our constitutive psychological mechanisms demonstrate how natural beliefs may be warranted even if not supported by justificatory arguments. As we have seen, descriptive explorations can open up new paths for assessing the normative status of belief in selves. Such natural beliefs cannot be justified in the sense that they are not supported by positive discursive argument. Nevertheless, given their status as natural beliefs they can appeal to non-propositional justification similar to that of Wittgenstein’s hinge propositions. This approach opens up the prospect of developing warrant for our belief in the self even if we have no ontological assurance of the existence of something like ‘the self’.

REFERENCES

Audi, R. (2002), “Prospects for a Naturalization of Practical Reason: Humean Instrumentalism and The Normative Authority of Desire”, *International Journal of Philosophical Studies*, 10(3), pp. 235-263;
Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, New York;
Baç, M. (2007), “Epistemological Naturalism, Skeptical Threat and The Question of Normativity In Post-Apocalyptic Times”, *Filozofia*, 62(7), pp. 590-600;
BonJour, L. (1985), *The Structure of Empirical Knowledge*, Harvard University Press, Cambridge MA;
Brown, H. (1996), “Psychology, Naturalized Epistemology, and Rationality”, in R. Kitchener and W. O’Donohue (eds.), *Philosophy and Psychology*, Sage Publications, London, pp. 19-32;
Campbell, T. (2006), “Human Philosophy: Hume on Natural Instincts and Belief Formation”, in E. Di Nucci and C. McHugh (eds.), *Content, Consciousness, and Perception: Essays in Contemporary Philosophy of Mind*, Cambridge Scholars Press, London, pp. 62-72;
Churchland, P. (1995), *The Engine of Reason, The Seat of The Soul: A Philosophical Journey Into The Brain*, MIT Press, Cambridge MA;
Fodor, J. (2003), *Hume Variations*, Clarendon Press, New York;
Garrett, D. (1997), *Cognition and Commitment in Hume’s Philosophy*, Oxford University Press, New York;
Gaskin, J. (1974), “God, Hume and Natural Belief”, *Philosophy*, 49(189), pp. 281-294;
Goldman, A. (2002), “The Sciences and Epistemology”, in P. Moser (ed.), *The Oxford Handbook of Epistemology*, Oxford University Press, Oxford, pp. 144-176;
Hume, D. (2006), *An Enquiry Concerning Human Understanding*, T. Beauchamp (ed.), Clarendon Press, Oxford;
Hume, D. (2011), *A Treatise of Human Nature*, D. Fate Norton and M.J. Norton (eds.), Oxford University Press, Oxford;
Kelly, T. (2003), “Epistemic Rationality As Instrumental Rationality: A Critique”, *Philosophy and Phenomenological Research*, 66(3), pp. 612-40;
Kemp Smith, N. (1983), *The Philosophy of David Hume*, Garland, New York;
Kitcher, P. (1992), “The Naturalists Return”, *The Philosophical Review*, 101(1), pp. 53-114;
Knowles, J. (2003), *Norms, Naturalism, and Epistemology*, Palgrave Macmillan, Basingstoke;
Kornblith, H. (1985), “What is Naturalistic Epistemology?”, in H. Kornblith (ed.), *Naturalizing Epistemology*, MIT Press, Cambridge MA, pp. 1-13;
Laudan, L. (1990), “Normative Naturalism”, *Philosophy of Science*, 57(1), pp. 44-59;
McCormick, M. (1993), “Hume on Natural Belief and Original Principles”, *Hume Studies*, 19 (1), pp. 103-116;
Metzinger, T. (2004), *Being No One: The Self-Model Theory of Subjectivity*, MIT Press, Cambridge MA;
Papineau, D. (1993), *Philosophical Naturalism*, Blackwell, Oxford;
Reichenbach, H. (1963), *The Rise of Scientific Philosophy*, University of California Press, Los Angeles;
Strawson, G. (1989), *The Secret Connexion: Causation, Realism and David Hume*, Clarendon Press, Oxford;
Stroud, B. (1977), *Hume*, Routledge and Kegan Paul, London;
Weintraub, R. (2003), “The Naturalistic Response to Scepticism”, *Philosophy: The Journal of the Royal Institute of Philosophy*, 78(305), pp. 369-386;
Wittgenstein, L. (1969), *On Certainty*, Basil Blackwell, Oxford;
Wright, C. (2004), “Warrant for nothing (and foundations for free)?”, *Aristotelian Society Supplementary Volume*, 78(1), pp. 167-212.

BIANCA BELLINI
Università Vita-Salute San Raffaele
biancabellini@gmail.com

TOWARDS A FAITHFUL DESCRIPTION OF THE FIRST-PERSON PERSPECTIVE PHENOMENON: EMBODIMENT IN A BODY THAT HAPPENS TO BE MINE

abstract

This article aims at providing a faithful *description of the first-person perspective phenomenon. After clarifying what makes a description faithful, it will argue that Perry's and Baker's theories alone do not offer such description. Nevertheless, they offer some interesting insights which, along with the phenomenological attitude, contribute to the formulation of a faithful description. This is why this article focuses on these two specific authors from a phenomenological perspective.*

keywords

Phenomenological attitude, bodily self, first-person experience, Leib, Körper

1. First and Third-Person Perspective Experiences

Following E. Mach's example (Baker 2013, p. 38), let us imagine getting on a bus and seeing a shabby-looking man at the far end. We will probably think something along the lines of '*That* is an unkempt person!'; however, when we suddenly realize that the person we are looking at is our own image reflected in the bus mirror, then we will think something along the lines of '*I* am the unkempt person!'. The core argument of Baker's *Naturalism and the First-Person Perspective* is that "versions of naturalism, without first-person properties, are in error" (Baker 2013, p. XXII). Mach's example enables us to dive right into the first-person perspective phenomenon. The crucial question that this article will attempt to answer is how one can describe the experience *exemplified* by Mach's case in a *faithful* way. Therefore, how the first-person perspective can be described in a faithful way.

J. Perry provides a similar example: "I once followed a trial of sugar on a supermarket floor [...] seeking the shopper with the torn sack to tell him he was making a mess [...]. Finally it dawned on me. I was the shopper I was trying to catch" (Perry 1993, p. 167). In order to describe this baffling feeling, one can distinguish two propositions formulated by the careless shopper at two different times: 1) 'Who is making a mess?'; 2) '*I* am making a mess!'.

This article will argue that Perry's theses enable one to formulate only a *partially* faithful description of the first-person perspective phenomenon[1]. Perry's approach, in order to be *entirely* faithful, needs to be completed: and here is where Baker and the phenomenological attitude come to the rescue.

2. Epistemological and Pragmatic Priority of the First-Person Perspective

Let us consider the conceptual categories through which Perry tries to describe the careless shopper's experience (Perry 1993, 2007[2]), which is very similar to that of Mach's example. Perry argues that the information obtained by the subject through the first-person proposition ('I am making a mess!') cannot be rephrased into a third-person perspective ('J. Perry is making a mess') without missing an *essential* information[3]. Furthermore, the two propositions reflect a distinction between two different ways of picking up information about people. Firstly, a 'self-informative way', that is, the ordinary first-person way ('I am making a mess'); secondly, a 'role-based way', that is, a third-person reformulation ('J. Perry is making a mess').

[1] This phenomenon is phenomenologically conceived as the *subjective* and *personal* side of every intentional structure.
[2] See also Borges 1964.
[3] *Cf.* also Sartre 1943, Part III, Chapter I (especially § 4).

Only after the realization that one is the careless shopper (or, in the other example, the unkempt person), the information gained in the first-person perspective is *linked* to the information gained in third-person perspective and the two shoppers' beliefs are no longer 'detached'. This 'linkage' causes the change in beliefs. Similarly, this change determines Perry to stop following the trail and rearrange the torn sack. Perry, in order to explain this change in behaviour, devises the distinction between *belief* and *belief state*. According to him, in fact, indexical beliefs are paramount in explaining behavior and making predictions. This distinction concerns the difference between the way one believes a belief, first or third-personally, and the belief's content. It is the belief state that explains behavior and that needs to be inherently indexical. This entails that explanations of the careless shopper's case or of the unkempt person's case cannot be given in terms of *what* is believed, but have to include *how* the belief is believed.

Perry's argument leads us to claim that – according to Perry – the impossibility of translating a first-person perspective assertion into a third-person perspective, is epistemological and pragmatic. This means that the first-person perspective has only an epistemological significance, as far as it is the cognitive perspective that persons can adopt on themselves, and a pragmatic one, as far as it is the action perspective. Can Perry's description be conceived as faithful in respect to the first-person perspective phenomenon? Before answering this question, it is necessary to identify which criteria a description should meet in order to be conceived as faithful.

3. The Phenomenological Attitude Provides the Criteria of Faithfulness

The phenomenological attitude towards a given phenomenon ensures the faithfulness of the phenomenon's description: it is an attitude consistent with the phenomenological *epoche*. This means that, when approaching a given phenomenon, it is necessary to bracket what one knows about the concerned thing, apart from what appears of it within his/her *experience* and *seeing*. As M. Scheler has clearly stated, phenomenology is neither the name of a new science nor a new method. Phenomenology is "an attitude of spiritual seeing in which one can see [*er-scahuen*] or experience [*er-leben*] something which otherwise remains hidden" (Scheler 1973, p. 137) and what "is seen and experienced is *given* only in *the seeing and experiencing act itself*" (Scheler 1973, p. 138). This entails that:

> A philosophy based on phenomenology must be characterized first of all by the most intensely vital and most immediate contact with the world itself, that is, with those things in the world with which it is concerned, and with these things as they are immediately given in experience, that is, in the act of experience and are 'in themselves there' only in this act (Scheler 1973, p. 138).

Therefore – I argue – it is possible to regard the description of a given phenomenon as faithful if and only if [4]:

a. It is consistent with the immediate experience that one has of this phenomenon.
b. It is consistent with the phenomenon's appearance and transcendence: every kind of thing has a specific way to appear and to transcend its appearance. For example, people have a specific way to appear – that is, physiognomy – and to transcend their appearance – that is, the knowledge of a person. The transcendence of the phenomenon is not its reality, but its entirety.
c. It is consistent with the essential traits emerged from a phenomenological seeing. Seeing is, on the one hand, a phenomenological seeing, that is, an approach to the phenomenon that 'puts into brackets' all the previous information one has about it, without deriving from an immediate contact with it. On the other hand, seeing is looking for those essential traits which make this phenomenon exactly what it is.

[4] A phenomenological background is indispensable to grasp the real meaning of these criteria (Husserl 1913, Scheler 1916 and, more particularly, Conni & De Monticelli 2008, De Monticelli 2000), that otherwise could be misunderstood; for the same reason, the knowledge of Baker's theses is necessary to understand the following paragraphs (see, especially, Baker 2000, 2007, 2013).

4. Perry's Account Does Not Meet All the Criteria of Faithfulness

Now, in the light of these criteria of faithfulness, does Perry's analysis about the first-person perspective satisfy them[5]? It is possible to evaluate the faithfulness of Perry's account only by putting into brackets Perry's position, *i.e.* a third-person naturalistic ontology. This allows us to individuate the consistency of Perry's analysis with the first criterion. In fact, the experience of being the careless shopper is well explained by Perry's description: one who had this experience would in fact consider Perry's explanation faithful to his/her own experience.

Nevertheless, the second criterion is not satisfied. The *appearance* of the first-person perspective phenomenon, in fact, also includes a physical component, the body, that is the physical and personal bearer of the first-person perspective. This physical component is an *essential* trait of the first-person perspective, but none of Perry's conceptual categories enables us to identify the fundamental role played by the personal body – nor does Perry recognize other essential features of the first-person perspective phenomenon. This implies that Perry's analysis does not meet the third criterion and therefore we can conclude that his description is unfaithful.

5. Ontological Priority of the First-Person Perspective: Can Baker's Account Be Conceived as Faithful?

The essential features that Perry fails to recognize seem to be acknowledged by Baker, especially the ontological priority of the first-person perspective, the distinction between the two stages of the first-person perspective and, therefore, the notion of 'self-concept' (Baker 2013). For this reason, the conceptual categories through which Baker argues her theses contribute, in a valuable way, to the formulation of a faithful description of this phenomenon. At close analysis, Baker's approach towards the first-person perspective phenomenon seems to consider the priority of the first-person perspective as *ontological* as well as epistemological and pragmatic. This ontological priority prevents first-person perspective sentences from being translated into third-person perspective. Ontology is conceived by Baker as the 'inventory' of all that exists: it "includes every object and property needed for a complete description of reality" (Baker 2013, p. 169). Baker conceives the first-person perspective as an ineliminable and irreducible property necessarily belonging to ontology: although ontological naturalism tries to rid reality of the appearance of first-personal phenomena by naturalizing them, the first-person perspective cannot be naturalized (Baker 2013, pp. 28, 30).

As I mentioned earlier, the distinction between the two stages of the first-person perspective and, therefore, the notion of 'self-concept' are *essential* traits of the first-person perspective phenomenon. In order to understand the necessity of these two features to achieve a faithful description, let us consider how Baker analyzes the experience of the unkempt person that we saw earlier.

> He did not realize that he* was the unkempt person referred to: He was referring to himself without realizing that it was himself* he was referring to. Soon, Mach realized that it was himself whom he was looking at [...]. And because Mach had a robust first-person perspective, with that realization came a raft of others [...]. In general, once a person has a robust first-person perspective, then his simple assertions using 'I' are connected to 'I*' sentences (Baker 2013, pp. 38-39).

This description seems to be *more* faithful than Perry's. Perry, in fact, identifies only *some* essential traits of the first-person perspective phenomenon, such as the epistemological and pragmatic priority of the first-person perspective, the distinction between different ways of picking up information about people, the idea of linkage and the dichotomy between belief and belief state. However, these features are not enough to make the description of this phenomenon faithful. These features are essential, but they are not the *only* ones.

[5] The notion of faithfulness is comparative. If a description does not meet all the criteria of faithfulness, then it is unfaithful: however, it can be *more* or *less* faithful in comparison with other descriptions of the same phenomenon.

6. Baker's Criticisms Against Perry's Theses

Baker investigates Perry's position in chapter 3 of *Naturalism and the First-Person Perspective*. Baker's description of the careless shopper's case seems to be *more* faithful than the one offered by Perry. This is because Baker makes use of her own conceptual categories concerning the distinction between the two stages of the first-person perspective and the 'self-concept'[6]. I will now briefly consider the two main criticisms that Baker addresses to Perry's analysis.

> (1) Perry does not give a non-first-person account of how the two notions of himself [...] became linked, and (2) Perry attempts to construe first-person phenomena in terms of ways of knowing associated with identity; but ways of knowing associated with identity are insufficient for first-person phenomena without a capacity to conceive of oneself as oneself* in the first person. So, there is no real reduction of the first person (Baker 2013, p. 52).

While the second criticism is very solid, the first one, at close analysis, seems to miss the mark. Perry does not give a non first-person account of what linkage is and how it happens; this means that his account ends up being inadequate to his reductive needs. More importantly, his account turns out to be *partially* consistent with a faithful description of the first-person perspective phenomenon, because it identifies some essential traits of this phenomenon. In fact, if Perry's naturalism – along with a phenomenological attitude – is put into brackets and if Perry's argument about the pragmatic and epistemological priority of the first-person perspective is acknowledged as faithful, then it becomes evident that one cannot demand from Perry to provide an account of the *linkage* in a third-person perspective.

7. Baker's Account Is Not Consistent with All the Criteria of Faithfulness: the Phenomenological Distinction Between *Leib* and *Körper*

Perry's account is not consistent with all those criteria of faithfulness discussed above, and the same applies to Baker's account. Her description, in fact, does not satisfy the second and the third criterion, because it does not embrace an *essential* trait of the first-person perspective phenomenon, that is, the *phenomenological* distinction between *Leib* and *Körper* (Husserl 1952, § 35-41)[7]. For this reason, her description is not consistent with the phenomenon's appearance and transcendence and with the essential features emerging from a phenomenological seeing.

Baker rightly argues that the first-person perspective needs to be embodied (Baker 2013). Yet, the way in which Baker conceives this embodiment seems questionable. As I said earlier, Perry does not recognize the significance of the physical component, whereas, according to Baker, the bearers of the robust first-person perspective are embodied human persons. A human person is in fact necessarily *constituted* by a body. According to Baker, the subject of experiences is the whole person, which is *constituted* by a whole body and which is not reducible to his/her brain, mind or body (Baker 2013, p. 142). This aspect turns out to be fundamental for a faithful description of first-person bodily experiences.

The crucial point is the thesis according to which "although we are essentially embodied, we do not essentially have the bodies that we now have" (Baker 2013, p. 142). This entails that the person has essentially *a* body, not the body belonging to him/her. Granted that the body has to provide the mechanisms supporting robust first-person perspective, this body can be made of anything. Going down this road, one might therefore end up to be constituted by non-organic bodies. In fact, Baker says:

> What is required for our continued existence is the continued exemplification of our first-person perspectives, along with some kind of body that has mechanisms capable of doing what our brains do (Baker, 2013: p. 142).

[6] The experience of being the unkempt person and that of being the careless shopper are well explained by Baker's description: one who had these experiences would in fact consider Baker's explanation faithful to his/her own experience. This entails that Baker's approach meets the first criterion.

[7] This distinction is broadly articulated, for example, by M. Merleau-Ponty, D. Legrand, A. Mandrigin, J. Kiverstein, A. Noë, M. Matthen and T. Metzinger.

This implies that we are *fundamentally* persons, who *necessarily* are embodied and who have the body that we have only in a *contingent* way. In short, according to Baker's theory:

> We are constituted by our bodies, and the bodies that constitute us now are organisms. With enough neural implants and prosthetic limbs [...] we may come to be constituted by bodies that are partially or wholly nonorganic [...]. The property of being me is the property of being *this* exemplifier of a first-person perspective. It is being this exemplifier of a first-person perspective that makes me *me* (Baker 2013, pp. 149, 155).

Basically, Baker's thesis seems to lack a definition of the *constraints* of the body's variability, which are aimed at preserving the personal identity. Baker's approach seems to lack a phenomenological distinction between *Leib* and *Körper.*

8. The First-Person Perspective Is Embodied in That Body That Is *Mine*

The absence of this phenomenological distinction makes Baker's description of the first-person perspective an *unfaithful* account of this phenomenon. This aporia in Baker's theory can be clearly understood if one considers the *embodied* experience deriving from playing a particular kind of game that will now be illustrated and which bears some similarities with Mach's and Perry's examples.

Let us imagine playing this game: sitting around a circular table, we lay our hands on it and then move our right hand to the right of our playmate's left hand so that one's left hand is located to the left of our playmate's right hand. Now, in turn, everybody has to lift up their hands so that all the hands are lifted up in succession one after the other. Playing this game is rather puzzling, because one is easily tricked into forgetting to lift his/her hand up at his/her turn and, instead, wait for the playmate sitting next to him/her to lift his/her hand up: it is as if one feels that one's *own* hand is not his/her, but his/her playmate's. It is possible to describe this baffling feeling derived from playing this game by distinguishing two propositions, formulated by the forgetful playmate at two different times: 1) 'Why is the other playmate not lifting his/her hand up?'; 2) 'Ah, it is my turn, I am not lifting my hand up!'. The analogy with Perry's careless shopper is evident and here drawn on purpose: 1) 'Who is making a mess?'; 2) 'I am making a mess!'.

The experience of being the forgetful playmate cannot be described in a faithful way neither through Perry's conceptual categories nor through Baker's. Surely, some conceptual categories conceived by Perry are useful to *partially* describe the forgetful playmate's experience, but none of them enables us to find out about the fundamental role of the personal body. Similarly, Baker's proposal also fails to shed light on the baffling feeling associated with the game case. Her approach allows us to grasp only some essential traits of the first-person perspective and does not allow one to understand how the forgetful playmate's experience is *firstly* an embodied experience, which *inherently* concerns the phenomenological distinction between *Leib* and *Körper.* The difficulty in explaining this feeling makes the peculiarity of this game case clear, especially when compared with Perry's case of the careless shopper: the latter can be clearly illustrated by Baker's theory, demonstrating how the impossibility of translating an experience lived in first-person perspective ('Ah, I am the messy shopper!') into a third-person perspective ('J. Perry is the messy shopper') is ontological, epistemological and pragmatic. However Baker's conceptual categories do not allow one to fully understand what happens in the game case, *i.e.* why the players do not immediately lift their hands up at their turn.

In the game case it seems impossible to understand the reason of that puzzling feeling unless we are minded to recognize that it is not sufficient that the first-person perspective is embodied in *a* body. It is necessary that the first-person perspective is embodied in the body that is *mine*: not a body that happens to be mine and could be someone else's, but a body that is mine. The first-person perspective is embodied in *my* body, in that personal body that is the bearer of the personal perspective on the 'world-of-life'. To ignore this aspect means to abolish the distinction between *Leib* and *Körper*, that is, the distinction that enables us to understand how the assertion 'Ah, it is my turn, I am not lifting up my hand!' would loose its meaning if we did not establish the constraints of the body's variability[8].

9. Body's Appearance and Transcendence

In Baker's view the person is just embodied in a body whose main task is supporting the first-person perspective. This implies that the *physiognomy* of the personal body is completely disregarded: the person is not essentially characterized by his/her body. However, the physiognomy of the body is what strikes us first when looking at someone. Baker's account, despite being better than Perry's description, is unfaithful to the essential traits emerging from a phenomenological *seeing* of the phenomenon's appearance itself. A specific phenomenological attitude therefore, along with the distinction between *Leib* and *Körper*, enables one to discover that faithful description that this article sets out to find. Nevertheless, the phenomenon's transcendence itself suggests that the individuation of the first-person perspective's essential features cannot be limited to this research; as a matter of fact, it demands a continuous exercise of phenomenological seeing. The faithfulness of this description can only be gradually achieved: phenomenology's primary task is a continuous attempt to comprehend the essential traits of every phenomenon.

To ignore the distinction between *Leib* and *Körper* means to ignore the distinction between body's appearance and transcendence. The body as a physical object represents the transcendence of the body's immediate appearance, *i.e.* the experienced body. This priority of the experienced body is clearly explained by Husserl: "*der Leib zugleich als Leib und als materielles Ding auftritt*" (Husserl 1991, p. 158), that is to say, the experienced body appears *immediately* as an experienced body and as a physical thing. It is the experienced body that can be conceived as *Leib* or *Körper*. *Leib* and *Körper* are two sides of the same coin: the first-person perspective is necessarily embodied in a *Leib*, which, necessarily, is a *Körper*. Without the notions of *Leib* and *Körper*, it seems impossible to formulate a faithful description of what happens to the forgetful playmate. Quite differently from what happens in the example of the careless shopper's and that of the unkempt person, the game case involves the personal body in a more specific way. A description of the first-person perspective that aspires to be faithful to the phenomenon itself has to examine this first-person bodily experience.

10. No First-Person Perspective Without Bodily Self

To conclude, a faithful description of the first-person perspective phenomenon has now been given, thanks to the dialectic examination of the phenomenological attitude and of Perry's and Baker's theses. It has been argued that there is no first-person perspective without *my* body and there is no bodily self without the first-person perspective: the body in which the first-person perspective is embodied is *my* body. As long as one acknowledges this main feature, it is possible to formulate a thoroughly faithful description of a given experience, such as in the case of the careless shopper, of the unkempt person and of the forgetful playmate.

[8] The game case is partly similar to the phenomenon of the rubber-hand illusion: a further investigation focused on the scientific counterpart of the argument here presented could support these claims with empirical evidence.

REFERENCES

Baker, L.R. (2000), *Persons and Bodies: A Constitution View*, Cambridge University Press, Cambridge;

Baker, L.R. (2007), *The Metaphysics of Everyday Life: An Essay in Practical Realism*, Cambridge University Press, Cambridge;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, New York;

Borges, J.L. (1964), "Borges and I", in *Labyrinths: Selected Stories and Other Writings*, New Directions, New York, pp. 246-247;

De Monticelli, R. (a cura di) (2000), *La persona: apparenza e realtà. Testi fenomenologici 1911-1933*, Cortina Editore, Milano;

De Monticelli, R. & Conni, C. (2008), *Ontologia del nuovo: la rivoluzione fenomenologica e la ricerca oggi*, B. Mondadori, Milano;

Husserl, E. (1913), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Erstes Buch, Allgemeine Einführung in die reine Phänomenologie*, Verlag von Max Niemeyer, Halle a.d.S.;

Husserl, E. (1952), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Zweites Buch, Phänomenologische Untersuchungen zur Konstitution*, M. Nijhoff, Den Haag;

Perry, J. (1993), "The Problem of the Essential Indexical", in *The Problem of the Essential Indexical and Other Essays*, Oxford University Press, New York, pp. 33-52;

Perry, J. (2007), " 'Borges and I' and 'I' ", The Amherst Lecture in Philosophy, Lecture 2, (http://www.armherstlecture.org/);

Sartre, J.P. (1943), *L'etre et le néant: Essai d'ontologie phénoménologique*, Gallimard, Paris;

Scheler, M. (1916), *Der Formalismus in der Ethik und die materiale Wertethik. Neuer Versuch der Grundlegung eines ethischen Personalismus*, M. Niemeyer, Halle;

Scheler, M. (1973), "Phenomenology and the Theory of Cognition", in *Selected Philosophical Essays*, Northwestern University Press, Evanston, pp. 136-201.

PATRICK ELDRIDGE
Katholieke Universiteit, Leuven
patrick.eldridge@kuleuven.be

OBSERVER MEMORIES AND PHENOMENOLOGY

abstract

This paper explores the challenge that the experience of third-person perspective recall (i.e. observer memories) presents to a phenomenological theory of memory. Specifically this paper outlines what Husserl describes as the necessary features of recollection, among which he includes the givenness of objects in the first person perspective. The paper notes that, on first sight, these necessary features cannot account for the experience of observer memories as described by Neisser & Nigro (1983). This paper proposes that observer memories do not so much entail a shift of perspective as they do a process of self-objectification and as such do not break with the phenomenological emphasis on the first person perspective.

keywords

Edmund Husserl, Ulric Neisser, Georgia Nigro, observer memories, episodic memory, perspective, self-objectification

1. Introduction

The philosophical questions that we pose about consciousness today are inextricably linked to questions concerning perspective. Whatever is conscious is thought to experience the world from its own first-person perspective. Yet, is it reasonable to speak of conscious experiences from a third-person perspective? We speak often enough of adopting other perspectives but does such talk have any philosophic or scientific weight? Cognitive psychologists Ulric Neisser and Georgia Nigro famously investigated perspective in memory and made a distinction between observer memories and field memories (Nigro & Neisser 1983). Field memories are recollections from the first-person point of view. Thus, when we remember, we re-experience the event from the original perspective we had when we first witnessed it. According to Neisser and Nigro, there are also memories where we are spectators of ourselves. They call these observer memories and claim that they are recollections from the third-person perspective. The person recollecting sees his or herself from the outside, participating in some event or other. The authors describe the distinction between the two forms of memory as a difference of vantage point (*ibid*., pp. 467-469).

This form of memory seems to pose a problem to philosophers who insist that the first-person perspective is a necessary feature of mental phenomena. Husserl is perhaps the thinker of the first-person perspective *par excellence* and we may wonder if new empirical findings upset Husserl's phenomenological account of memory, since he held there was apodictic evidence that the first-person perspective characterizes all conscious experience. Indeed, cognitive scientists have found that there are several factors that motivate a change of perspective in memory. Neisser and Nigro point to the purpose, emotional quality, and level of self-awareness of a memory determining whether it will have the field or observer perspective (*ibid*., pp. 481-482). Other studies have shown that aging has an impact on the frequency of observer memories (Piolino et al. 2002) and there is some agreement that observer memories are re-constructions of past events rather than copies of them (on the 'construction' model of memory, see Conway & Pleydell-Pearce 2000). The aim of this paper is to investigate the challenges and opportunities that this form of recollection offers to Husserl's phenomenological analyses of memory and to decide whether observer memories offer insuperable obstacles for intentional analysis.

2. What is a Phenomenology of Recollection?

Phenomenological research into recollection consists in the attempt to determine the intentional structure of the conscious experience of remembering, i.e. the way that the mind refers to or is 'about' transcendent objects when it remembers them. It also consists in determining the structures of remembered objects with respect to their thinkability, i.e. the conditions of their possible experience. When it comes to the analysis of intentional experiences, Husserl claims that what is directly revealed in reflection may not be enough and that a comparison within reflection is often necessary (Hua XIX/1, pp. 462)[1]. In his lectures on inner time-consciousness Husserl compares and contrasts recollection with perception. Perception and recollection have roughly the same temporal structure – they both have a privileged now-phase, they both have running-off phases. For example, whether I hear or remember a melody, the experience has a unity and flow in duration; the past notes do not completely disappear and the current note is fresh. In recollection there is a temporal present, a Now, but it is a remembered, re-presented Now that has elapsed. Thus, there is a discrepancy between the now *that* I recollect and the now *in which* I recollect, unlike perception where there is simultaneity between the perceived object and the perception's execution (Hua X, pp. 40-45).

To clarify this talk of discrepancy and simultaneity, we should note Husserl's distinction between intentional consciousness and inner consciousness. Intentional consciousness refers to acts that mean transcendent objects in different ways, e.g. perception, phantasy, signification. Inner consciousness is a pre-reflective self-experiencing – a self-awareness that the ego has of its own intentional activity. It is the non-explicit awareness that I own the copyright on the activities and sufferings of my consciousness. The traditional term is apperception. Through perception I experience external objects, but I also experience the act of perception immanently in inner consciousness. When perceiving some object it is given to me as being vividly now and the act of perception is also given to me apperceptively (pre-reflectively) as being now. Thus if I remember some object right now, I am running through an elapsed perception, and the object's now-phases have already run their course. Thus, there is a discrepancy between the object's now and the apperceptive now in recollection, but for perception they are simultaneous.

This discrepancy results from the doubleness of recollection. According to Husserl's analyses, recollection exhibits a double intentionality (Hua X, pp. 53-55, 57-59). Yet the doubleness of the intentionality does not yield a double-object. When I recollect, I thematically intend the object of my former perception and I implicitly intend that perception, which was originally experienced in inner consciousness. The external experience is necessarily nested in consciousness by means of an internal experience. The quality of 'having been perceived' is an essential determination of the recollected appearance. Thus Husserl says that recollection is constituted by a double intentionality. The act of perception is intentionally *implied* but not thematically posited in recollection, unlike the remembered object. For example: 'I remember my snow-shovel' does not mean 'I remember having perceived my snow-shovel' as that would signify an act of reflection. It rather means 'I see the snow-shovel as having been'. In the now the rememberer sees the not-now. Expressed more technically, I intend the same object, I execute that perception again, and this 'again' expresses how retentions have modified that intentional act. Here retention refers to a phase of the living present that both preserves and *de*-presents phases of my intentional acts as they trail off into the past.

Since recollection has a double intentionality, and intentionality is structured by inner time-consciousness, recollection also has a double flow; I experience both the initial perception's elapsed now and recollection's actual now. Despite this doubleness, it belongs to *one* stream of consciousness. Conscious acts like perception endure in inner consciousness but they also succeed each other and remembered-perceptual acts appear as having a certain co-ordination in that succession. Husserl holds that recollection is an experience integrated into one conscious life by virtue of its determinate horizon of protentions (Hua X, pp. 52-53).

[1] All Husserl references are to the volumes of the *Husserliana* editions followed by the page number.

Just as intentional acts have their retentions, so they have their protentions, which refer to the phase of the living present that is open to what is coming next. Whereas the protentions of perception emptily and indeterminately anticipate what is coming, the protentions of memory have been determined. The moments that I protend in one phase of memory are identical with the moments I retend in a subsequent phase of memory, which does not hold for perception. More plainly: the next moment of perception is a possibility while the next moment of recollection is an actuality, a determinate part of my unitary, flowing conscious life. Thus, remembered events are posited in a temporal context. In remembering an object, I implicitly intend the perception with its obscure temporal surroundings, which are nothing other than the moments of my life.

To sum up this sketch of the phenomenology of recollection we can say that: perception gives me the 'now as now' by virtue of the simultaneity of its object, the object's presentation, and the inner awareness of that presentation. To reproduce an elapsed now I must bring about a modified perception, which is what we call recollection. The thematic focus of this recollection is the percei*ved* but we also implicitly intend the percei*ving*. In recollection I attend to the former perception's object and the object's now, which is posited *in relation* to the actually present now[2].

3. What Challenge do Observer Memories Present to Phenomenology?

Can phenomenologists make sense of observer memories? It seems that the phenomenologist would hold that recollections must in principle be field memories. If recollection is re-experiencing what we originally perceived and if perception is necessarily in the first-person perspective, then it should follow that recollections are in the first-person perspective. Husserl's analyses show that the body bears the zero point of orientation for perceptual exploration. The horizon of perceptual space opens up around me. If I move my head then there is a shift of the object – the object stays where it is, but it now appears more to the left in my visual field. I apperceive my own possibilities of movement and the profiles of the objects around me as being correlated. Given the decisive role that the apperception of the body as a zero point plays in Husserl's analyses of perception, and given that recollection for Husserl is a quasi-re-perceiving, we must ask: what is the zero point of observer memories? What sort of perspective organizes appearances in observer memories?

The phenomenologist could deny that they are really recollections. Instead one might say that observer memories are knowledge of the past rather than recollections. This would then dismiss observer memories as being merely event-specific knowledge accompanied by confused quasi-perceptual elements. The strategy is not a satisfying one, given that observer memories appear as a species of episodic remembering rather than semantic knowing (on this distinction see Tulving 1972).

Perhaps observer memories have a distorted self-intention. Brough argues that Husserl's own logic dictates that recollection requires a triple, not double intentionality. When recollecting we intend: i) the object originally perceived in the act; ii) the perceptual act executed; and he adds iii) a past segment of the absolute time-constituting consciousness. To illustrate: I recall an object and to do so I implicitly intend the perception that constituted the object and to do that I even more implicitly intend the inner temporal experience that constituted the perception. Brough says: "to recall the elapsed act without representing the flow through which I first experienced it, would be tantamount to recalling an act which belonged to no one" (Brough 1975, pp. 60)[3]. We might have grounds to say

[2] The foregoing outline is quite meagre. It only presents what is necessary for the subsequent analyses, leaving out the important analyses of affection, motivation, and fulfilment in memory.

[3] This is a presentation and not an endorsement of Brough's position. An alternative view can be found in Zahavi (2003), who posits that the pre-reflective self-awareness of the act is nothing other than its temporalizing. There are great merits to this latter position but the issue of whether or not inner consciousness is adequately described in analyses of time-consciousness is thorny. Brentano (2008, pp. 144-152) for one held that the apperception of one's intentional activity included propositional and even affective dimensions.

that observer memories are recollections in which one intends an object and a perception but fails to properly couch those intentionalities in an intention of the flow that constitutes the synthetic unity of the stream of consciousness. We can then say that observer memories are conditioned by a distortion of recollection's nested structure. The theme of self-consciousness is crucial but I would stress that observer memories do not have a failed self-intention but rather an original and peculiar form of self-intention. I propose that observer memories are genuine forms of recollection that involve a self-objectification.

4. What is Self-objectification?

This theme is a vast and multifarious one. I will omit Husserl's considerations concerning the empirical ego as a psycho-physical reality. Instead, I will restrict myself exclusively to forms of self-objectification that speak directly to the question of perspective. I will start with a straightforward case of self-objectification in perception, moving on to inner consciousness, then, taking these together, I will attempt a sketch of self-objectification in observer memories.

Already in perception we grasp ourselves as objects. Specifically in touch there occurs a twofold apprehension: I feel the *object's* tactile features and I feel the localization of *my* sensations and movements (Hua IV, pp. 79-84). I touch the object, but I am also touched by the object, i.e. by touching objects I can discover objective features of my hand. Even in perception my body is constituted for me both as a means and as a transcendent object of external intuition. With respect to movement, I apprehend myself as initiating certain movements and as suffering other movements. Husserl's famous distinction between *Leib* and *Körper* shows how my body is constituted for my consciousness as both a lived body and some extended matter (Hua IV, pp. 157-160). Here self-objectification means apprehending oneself as a thing with its exposed surfaces and its externality to intentional animation.

This awareness of one's body as something object-like, however, is not sufficient to explain the sort of self-objectification in observer memories. This tactile manner of self-objectification is a feature of perception and would as such be common to field memories insofar as they are re-perceivings. The self-object in observer memories is unlike this basic tactile self-objectification in two regards. First, in observer memories I take a distance from that which I apprehend as my body, whereas in touch I merely change attitude or apprehension with respect to my body. Second, in observer memories I apprehend my objectified self not just as a body, but rather as a person who I once was. Observer memories are not limited to merely remembering objective and causal features of my self. The self-objectification specific to observer memories requires further analysis.

Returning to our theme of inner-consciousness, we note that in recollection, the initial perception and its object have been representationally modified, but what of the inner consciousness involved? Husserl tells us that every experience is either impressional (i.e. inwardly presentational) or re-presentational. On the one hand, Husserl says that to every consciousness of something immanent (every impressional, inner consciousness) there corresponds a re-presentational consciousness of the same (to every sensed red there is a possible phantasmal red). On the other hand, every re-presenting is, in turn, couched in an impressional, inner consciousness; every conscious act is impressionally experienced. On the other, other hand (!), among such impressional experiences some are present *as* re-presentations. We must juggle three demands: 1) all consciousness involves inner, impressional consciousness; 2) any consciousness has a possible, representifying modification that corresponds to it; and 3) impression and representation are mutually exclusive terms (Hua XXIII, pp. 301-312). To clarify the stakes and theme, I find that in phantasy, for instance, I represent some event that is not really an event in my life. The phantasied event does not happen to me – it happens to a phantasied me. It is not a lived, impressional me but a modified me. I experience the act of phantasy and so it has a place in my stream of consciousness but it is present there as a foreigner. This is in stark contrast to perception, in which I fully identify with the one perceiving, where my experience of the perception is impressional through and through. It is doubtful that Husserl was ever fully satisfied with the analysis of inner consciousness in recollection. Upon recollecting, consciousness folds back upon itself, yielding an experience that is temporally structured not by one centre but by two poles (two 'nows'). This doubleness both defines and obscures my self-experience in remembering – a dual nature that is difficult to clarify. On this background, what can say about self-awareness in observer memories?

If we really attend to observer memories, what proves to be truly salient in them is not the shift of perspective noted by Neisser and Nigro, but rather something that they missed: the introduction of a new element, the inclusion of an objectified self. The event I present in an observer memory belongs to my past, but it could not have happened like that. There is a self-object in the memory, one that I identify with, but there are certain irresolvable discrepancies. For example: if I remember shovelling snow from an observer perspective it is possible that I see the surface of my eyes. Yet, I could not have seen the surface of my eyes when I was outdoors in the snow.

Furthermore, it is misleading to say that one adopts a third-person perspective when having an observer memory. When I recollect myself shovelling, I recollect this self in front of me, at a certain distance and angle. Thus even my objectified self correlates to a zero point of orientation. The way that Husserl conceives of the connection between perspective and the body is not at all straightforward. For Husserl, even when I imagine a jabberwock it is given with an orientation to me – it is to the left of me, it is galumphing away from me. Thus my experience of the imagined jabberwock with its profiles is also correlated to my perspective although in this case we cannot speak of real, localized eyes that are really seeing (Hua IV, pp. 55-58). Therefore, observer memories cannot be characterized in terms of a shift of perspective but rather the constitution of a self-object that results in a complex awareness of perspectives. It is complex because not only is this self-object given to me in some form of memorial first-person perspective, I grasp this self-object precisely as something that has its own zero point of orientation. I apprehend my objectified self as having a certain perspective on the shovel.

There is no shift outside of the first-person perspective then. Our conceptual analysis based on intentionality finds that the common description of empirical-psychological research on observer memories errs where it says that observer memories are given in the third-person perspective. There is, however, a complex perspective due to the constitution of a self-object. With respect to self-awareness or apperception in an observer memory we can say that it is: A) impressional because I experience it as really belonging to my past; B) representational because there are elements that I never really experienced; and most intriguing C) it is exteriorized[4]. The hard problem of self-awareness in observer memories is how to think the unity of the apperception of the remembering with its exteriorized self-awareness. In all recollection as such we implicitly present a lapsed self, who is removed from who we are now.

[4] To admit both propositions A and B is tantamount to contradiction for Husserl. This points to the need for a revision of the old distinctions within inner consciousness. The main task of this revision would be to avoid conceptual contradictions while respecting the tension that defines the experience of recollection, i.e. the tension between memory's dreamlike nature and the way it presents events as 'hard facts' that have irrevocably and irreversibly passed. Perhaps the problem lies in Husserl's overly centrist conception of inner consciousness, which cannot handle heterogeneity. The path towards a reconciliation, then, lies in the direction of a more complex model that accounts for the way that, in remembering, consciousness doubles back on itself.

The objectified self in an observer memory seems to give intuitive content to this sense of removal. The object-self in an observer memory is connected but not contiguous with my self now. It is me 'out there'. An observer memory objectifies the self-alienation conditioned by changes over time.

I will close this section with a few hypotheses concerning the benefits of observer memories, guided by the question: what does the self-object contribute to my conscious life? Aside from lacking certain lived, first-person attributes, can we not say that such a self-representation accomplishes something that the field-memory does not? It might be the case that the objectified-self acts as a stand-in, an actor who plays me on the stage of my past. This actor can explore and experience things while I safely watch from my seat. I can represent events at a distance without being affected by them in the same way that reliving them would entail. From a clinical, pathological perspective, this self-object might be akin to Deleuze & Guattari's (1991, pp. 60-81) *personnage conceptuelle*. Just as Zarathustra can tell me things that Nietzsche cannot, just as the transcendental ego can experience things that Kant cannot, so my self-object can relive disturbing experiences I cannot. Indeed, empirical research has shown that voluntarily changing recollections from field memories to observer memories decreases levels of affectivity (Robinson & Swanson 1993). It is precisely the benefit behind self-objectification that is difficult to explain in transcendental phenomenology. Why should I be psychologically vulnerable to my own past? Why should the transcendental ego feel any danger coming from its previous constitutional achievements? At any rate, there are also non-pathological explanations of self-objectification. There is likely a link to what cognitive psychologists call 'verbal overshadowing'. When one recounts an experience again and again, the narrative elements of the story may start to seep into the recollection. There would be a story-self in the recollection of the original event – a strange mixture of the 'I' that a story-character utters and the lived 'I'. This calls for further empirical research.

5. Conclusion

I will restate my initial question: do observer memories pose a problem for Husserl's phenomenological account of recollection? I have argued that observer memories do not break or even escape Husserl's account of recollection (i.e. they are not a salient counter-example). I hope also to have shown that Husserl has provided us with the distinctions and concepts to produce knowledge about observer memories in phenomenological description. Thus, rather than a proof against Husserl's philosophy of recollection, I believe the observer memory phenomenon makes a strong case *for* Husserl's foundational insight that self-identity and pre-reflective self-consciousness are vital structuring elements of mnemic experience. What the observer memory does reveal, however, is that self-consciousness is ubiquitous yet evasive, moving on a spectrum from immediate, immanent self-identification to quasi-exterior-representation.

REFERENCES

Brentano, F. (2008/1874), *Psychologie vom Empirischen Standpunkt*, Ontos, Frankfurt;
Brough, J. (1975), "Husserl on Memory", *Monist*, 59(1), pp.40-62;
Conway, M. & Pleydell-Pearce, C. (2000), "The Construction of Autobiographical Memories in the Self-Memory System", *Psychological Review*, 107(2), pp. 261-288;
Deleuze, G. & Guattari, F. (1991), *Qu'est-ce que la philosophie?*, Minuit, Paris;
Husserl, E. (1952), *Ideen zur einen reinen Phänomenologie*, *Zweites Buch*, Hua IV, Nijhoff, The Hague;
Husserl, E. (1969), *Zur Phänomenologie des inneren Zeitbewusstseins*, Hua X, Nijhoff, The Hague;
Husserl, E. (1980), *Phantasie, Bildbewusstsein, Erinnerung*, Hua XXIII, Nijhoff, The Hague;
Husserl, E. (1984), *Logische Untersuchungen*, *Bd. II*, Hua XIX/1, Nijhoff, The Hague;
Neisser, U. & Nigro, G. (1983), "Point of View in Personal Memories", *Cognitive Psychology*, 15, pp. 467-482;
Piolino, P., Desgranges, B., Benali, K., & Eustache, F. (2002), "Episodic and Semantic Remote Autobiographical Memory in Aging", *Memory*, 10, pp. 239-257;
Robinson, J. & Swanson, K. (1993), "Field and Observer Modes of Remembering", *Memory*, 1, pp. 169-184;
Tulving, E. (1972), "Episodic and Semantic Memory" in E. Tulving & W. Donaldson (eds.), *Organization of Memory*, Academic Press, New York, pp. 381-402;
Zahavi, D. (2003), "Inner Time-Consciousness and Pre-reflective Self-awareness" in D. Welton (ed.), *The New Husserl*, Indiana University Press, Bloomington, pp. 157-180.

GAETANO ALBERGO
Università degli Studi di Catania
gaetanoalbergo@yahoo.it

THE FIRST-PERSON PERSPECTIVE REQUIREMENT IN PRETENSE

abstract

According to Lynne Baker we need to investigate the performances to understand if someone has a first-person perspective. My claim is that language has not the main role in the formation of epistemic states and self-consciousness. In children's performances, we have evidence for a self-consciousness without "I" thoughts. We investigate if it is possible to understand the difference between a case of false belief and one of pretense. My aim is to demonstrate that pretense is not a proto-concept but a first-person fact, endowed with a rich phenomenology.

keywords

Awareness, pretense, non-conceptual point of view, agency, intentionality

1. Introduction

According to Lynne Baker: "To have a robust first-person perspective, one must be able to manifest it" (Baker 2013, p. 154). Well, how can children manifest their first-person perspective? If we consider this ability as our *desideratum*, I think it is instructive to compare it to its closest early manifestation, i.e., pretend play. It seems that intentionality is a necessary condition for the activity of pretense. In the first part of the paper I investigate if it is possible to understand the difference between a case of false belief and one of pretense. Against the notion of *prelief*, an early status of indistinction between pretence and belief, our claim is that awareness is not dissociable from the first person perspective. However, a phenomenological approach must be supported by an epistemology that can explain how the mind is sensitive to the *refractory* nature of the world. The theory of agency allows us to highlight some relevant differences between first and third-person perspective. Then, we try to understand the limits of a theory that attributes to the language the main role in the formation of epistemic states and self-consciousness. Pretense, as an early manifestation of a set of pre-reflective self- and social cognition abilities, represents evidence for a self-consciousness without "I" thoughts.

2. Why Should We Distinguish between First and Third-person Perspective?

Angeline Lillard (2001) has observed that every pretense act involves certain features, several of which are defining and necessary: there must be an animate pretender and a reality that is pretended about, a mental representation of an alternative situation must be involved and projected onto the reality, and, this is the Austin's requirement, action must be intentional. Without intention there is no pretense, as Searle has noted too: "One cannot truly be said to have pretended to do something unless one intended to pretend to do it" (Searle 1975, p. 325). Finally, "full awareness" of the actual situation and the represented one is required. I think that Baker's suggestion to investigate the performances brings us on the right track. However, at the outset, we need to better understand the ways one can talk about the "manifestation" of an ability. So, some insist on stressing the difference between imagination and explicit behavior emphasizing a more evident transparency and the intentional nature of mental states compared to simple behavior. As a consequence, transparency of the imaginative states is regarded as a logical prerequisite to understand the limits of the principle of "semantic innocence", whereby the semantic value of a referential expression ought to remain constant inside and outside the scope of a verb of attitude like "believe". Imagination would be the gateway to intensional contexts, namely those contexts in which two expressions with the same extension cannot be substituted *salva veritate*. I think that, to avoid confusing the transparency of the imagination with the opacity of the mind-world epistemic relation, it would be better to consider the transparency of imagination from the perspective of an epistemology of understanding, letting epistemic states, such as belief and desire, to pertain to an epistemology of knowledge[1]. Mind, here, *should* learn to keep track of the world. In imagination the world offers us props, but what you must keep track of is different from the truth. Now, we know, pretense and imagination are not overlapping phenomena. Normally pretense acts are visible and children align their pretense responses with action. Yet, pretense is well adapted because it is an activity able to combine features of acting with the epistemic ones of the intentional attitude[2].

Pretense is not a fact about what happens to the body in acting, but, on the other side, it is neither, as for Baker, a rudimentary, by default, first-personal perspective. It is, rather, a first-person fact, endowed with a rich phenomenology. Nevertheless, some have considered it a mere proto-concept, not a full-developed mental concept. Theorists will naturally balk at referring to children's "belief" at all, so for example, according to Joseph Perner (1994), at the age of three children possess a concept, "prelief" (or "betence"), in which the concepts of pretense and belief coexist undifferentiated. The concept of prelief allows the child to understand that a person can "act as if" something was such and such (for example, as if "this banana is a telephone") when it is not. At the age of four, they understand that, like the public representations, inner representations can also misrepresent states of affairs. This hypothesis lends itself to several criticisms. The idea of an early lack of distinction between pretense and false belief contains a confusion between ascriptions in the first and third-person. For example, a first element that demonstrates the implausibility of the argument of indistinction is the recognition that engaging in pretense involves a certain degree of awareness that one is dealing with a not-real situation. Lynne Baker has strongly highlighted the connection between awareness and first-person perspective. Furthermore, whoever does not distinguish between the point of view of the first- and the one of third-person, is mistaking by way of making the risk of representational abuse something more than a mere logical possibility. To say that an observer can also confuse a wrong action for a case of pretense does not mean that from the first-person perspective he is unable to distinguish between reality and fantasy. Indeed, the first situation seems quite common among children under the age of three, and this is easy to explain if, as noted above, we use the concept of acting "as if". Wendy Custer (1998), for example, in a series of studies with three years old children, has used images of people engaged in the action of fishing but catching a boot instead of a fish. In the pretense condition researchers described a man as pretending to fish. Then, two drawings with "thought pictures" were presented. In the first one the man was thinking to catch a fish, in the other one the real situation was depicted. Children were asked to choose which one represented what the man had in his mind during pretense. Custer reported the high percentage of correct answers even among children of three years, i.e., the tendency to choose the thought picture with the hooked fish in the pretense condition. These results could suggest a mentalistic interpretation, i.e., one might conclude that behind these performances lies the understanding that pretense comes from thoughts entertained by the minds of the characters. Nevertheless, it is not difficult to give an alternative explanation of deflationary kind. Before using meta-representational hypothesis, we may give an account of the results in this kind of test using the ability of the subjects to recognize that pretending is different from reality, and this would explain the choice of combining the thought picture containing a false thought with the pretense condition. Our hypothesis has the effect of showing how children usually see the wrong actions as cases of pretense. However, this does not mean that those who make mistakes are acting according to the same relation to the world that is supposed in the act of pretending. We can try to understand the difference between these two perspectives considering both as our 'constructions'. A construction is usually something that we realize for a purpose. For example, when we observe an action, we can also imagine the consequences that usually, *ceteris paribus*, accompany it. Among our constructions some may be true, some may be false. In addition, every fiction turns out to be a construction. Nevertheless, the reverse is not valid because, by definition, no fiction is true. If this removes, at least at the conceptual level, the possibility of confusion-dementia, however, it only provides us a tautological solution to the issue of the alleged confusion between an incorrect action and pretense from the perspective of the third-person. We could indeed recognize that, from the perspective of third-person the one who observes maintains a positive attitude towards the set of things observed. There is some kind of regularity between our actions and the world. As a matter of fact, our actions are usually in keeping with the world because our beliefs are quite often in agreement with it.

Now, to be conscious is to be conscious *of* something. We can imagine someone who has not acquired the Brentanian 'intentionality of the mental', because he is not able to put together objects of thought and mental orientations. He would be unable to think that something is 'so and so' because he would not have achieved the three-term relation between 1) subject, 2) propositional attitudes, and 3) contentful thoughts. However, it would be always possible to ascribe to this subject some kind of consciousness, at least the one of getting experience of objects, and, consequently, the self-consciousness obtained by distinguishing objects in the world and the experience of them. To be able to have different mental attitudes towards intentional objects would then be secondary to the ability to achieve different physical orientations towards real objects. The notion of agency is more primitive than thought.

3. Agency and Cognitive Development

Piaget thought that children develop the self-world dualism through exercising agency, because their actions would become progressively more spontaneous, differentiated, and integrated. According to James Russell, this position would be a modest proposal, because its claim is that interaction with objects is a necessary feature of mental development and self-awareness. We usually work with the idea of an organism equipped with representational capacities supposed to be the *explicantia*, the starting-point having its *explicanda* in successful interactions with objects out there. Instead, from the Piagetian point of view, the theoretical starting-point will be an acting and sensing organism, of course not a pure agent, within a world of objects. This does not mean that we do not need representations, it is just that they lose their priority in the relation of the mind with the world. This is captured by James Russell when he writes that the question for the representational theorist is "What kind of representational medium or content must be innately present or must develop if this is to become the mind of a successful thinker and agent?" (Russell, 1996, pp. 75-76). Yet, focusing on the responsibility for our own actions and on the experience of the constraints that reality sets on what we can experience, the question becomes "What does this organism have to be able to *do* in relation to objects if it is to develop an adequate representational system?". Cognitive development is not only a matter of representing how things are out there. Representations distinguish between a subject and a world of objects, but activity is necessary to establish the self-world dualism. Experiencing the refractoriness of reality is necessary for subjectivity and self-awareness, because making a contribution to the object of experience, paired with the phenomenological value of participation, allows us to develop a subjective mental life set off from an objective reality. Moreover, if we assume the dependency of subjectivity and self-awareness on agency we can also understand the obvious conceptual links between considering others as rational beings and considering them as agents. The

[1] Following a rationalist tradition in meta-knowledge, we could formulate an approximate difference between transparency of imagination and opacity of full epistemic states in terms of weak and strong transparency:

- An epistemic state E is weakly transparent to a subject S if and only if when S is in state E, S can know that S is in state E;
- An epistemic state E is strongly transparent to a subject S if and only if when S is in state E, S can know that S is in state E, and when S is not in state E, S can know S is not in state E.

[2] See Albergo 2012, 2013; Harris 2000.

giving and asking for reasons in practical reasoning presupposes that there is not a mere passivity in relation to putative objects of knowledge. This does not mean that minds may be known entirely from the outside. Knowledge of our own actions might not have a representational character. Being an agent is an intrinsically first-person fact, it is known immediately and non-observationally. As Thomas Nagel puts it, there is "a clash between the view of action from the inside and *any* view of it from the outside. Any external view of an act as something that happens [...] seems to omit the doing of it" (Nagel 1979, p. 189). Saying that the subject-attitude-content triad is not a form of primary behavior does not mean that children conceive of others in behavioristic terms. To perceive others as agents means, at least in a modest form, to recognize them as endowed with minds, and it is possible to perceive others as agents only if we experience being agents in the first-person. Nevertheless, experiencing one's agency does not require the concept of agency. It is not the problem of ascribing a mental category to others after picking oneself out as the referent for a predicate. Here, predicate ascription, the germinal form for the following I-thoughts, is not at issue. However, it is not enough just to enunciate such a conceptual claim, we should also look for experimental confirmations. For example, Andrew Meltzoff (1990) showed that one year old babies are able to take a third-person perspective in relation to their own actions. The way in which they recognize when they are imitated drives us to hypothesize the existence of two parallel abilities: the one of realizing that it is their activity what is reproduced in the behavior of others, and the one of being able to project agency in others[3].

A first important contribution that the present theory offers us consists in recognizing the necessity of separating the points of view of the first- and the third-person. Recognizing that situations of direct experience lead to a wealth of mental activity much greater than that involved, at least in these early stages, in the processes of attribution to others, it also means having good reasons to keep a distance from the supporters of the *theory-theory* approach[4]. To develop a theory means to increase our own capacity to access to the relevant theoretical concepts, whose existence is assumed in a disembodied way. Children's access to their own mental life in first-person is just a special case of access to the notions of belief, pretense, or, in general, of mind. Access to these notions is independent by the epistemic perspective. One of the most common result of this kind of solution is to put different proto-abilities under a single category, with the consequence of having to explain mistaken performances with strange *ad hoc* hypotheses, such as Perner's idea of prelief, failing to recognize that they are actually two different competences. When working with proto-concepts concerning explanatory theories it would be advisable to make sure that attributions to child were constrained. It is not just a matter of how fine-grained you are prepared to be. About pretense, a very clear example of this kind of error is offered to us by Leslie and Happé, as in their opinion "solitary pretense comes out simply as the special case in which the agent of PRETEND is self" (Leslie & Happé, 1989, p. 210). These approaches forget that the subject has developed the ability to understand pretense in others because he himself is able to pretend. The theory of agency, therefore, allows us to support the hypothesis of the difference of the two perspectives even on the logical level. We need to recognize that being able to pretend in first-person is a necessary condition for being able to recognize it in the others. Now, we may ask whether it is also a sufficient condition. It is not only a conceptual problem, rather, formulating it allows us to go back to the problem from which we started, in order to clarify how it is possible that children assimilate false actions of others to cases of pretense.

[3] To avoid any possible confusion between the hypothesis illustrated and the thesis supported by the theorists of simulation by reflection, it is enough to recognize that we are talking about abilities that demarcate a pre-theoretical competence, while for the simulationists, there is no a theory that should be acquired but only a set of concepts that need to be developed.

[4] According to some authors concepts do not capture only salient features in objects. Children would develop concepts as mini theories using knowledge about causal mechanisms, teleological purposes, hidden features, and a biologically driven ontology (see Carey 1985; Keil 1989).

The theory of agency allows us to consider children under three years of age as subjects capable of conceiving the other as an agent and not merely as producer and consumer of representations. Action maintains a relation with the idea of experience that is not present in the explanations based on the passivity of subjects in relation to putative objects of knowledge. In addition, it is possible to interpret the tendency to see errors as cases of pretense as an example of the attribution of the ability to change the nature of perceptual inputs at will, as Piaget would say. It also allows us to put this kind of performance within a growing executive ability rather than an emerging theory. Thus, the first-person perspective in pretense is only a necessary condition for the attribution in third-person. Its insufficiency, if we put in these terms the inability to exclude from its extension erroneous situations, turns out to be an additional argument in favour of the existence of a non-observable, irreducible element, belonging to the first-person.

4. A Non-conceptual Point of View

With respect to self-consciousness, I think that Baker's insistence on the inescapable role of language goes beyond what is justified by the facts. Early pretense is just an example of this problem. For example, according to Jose Bermudez (1998) when we say that without language there would be no self-consciousness, we meet two circularities. The first is that the ability to entertain thoughts with self-consciousness precedes the competence with the pronoun "I", while the hypothesis of language first makes the ability to entertain I-thoughts dependent on the linguistic competence. Moreover, and this is the second circularity, the circular dependency of the two abilities makes it difficult to explain also how we can become self-conscious. So, if each of the two capacities presupposes the other one, what do we learn first, to entertain I-thoughts or to use I-sentences? For Bermudez there are also kinds of non-conceptual self-consciousness, something similar to self-recognition. These forms of thought would not be based on any linguistic mediation. If the thought may be non-linguistic, then even self-consciousness can be non-conceptual. Thus, the competence in the pronoun "I" is far from being regarded as a prerequisite of self-conscious thought. We know that Bermudez adopts Peacocke's idea that there are non-conceptual contents that are not representational in nature. For example, the function of objects in pretense games may be a good example of this kind of content. The pretender projects a minimum image onto the real situation, therefore realizing real departures from reality. So a stick becomes a horse, a sceptre, a sword[5].

Looking for a primitive form of self-consciousness in infants lacking language, Bermudez relies upon scientific data about growing abilities, such as reaching-behavior, object-focused attention and pointing. According to Bermudez, a lot of evidence pushes a primitive self-consciousness back into pre-linguistic stages of human development. So, in order to abandon the idea that self-consciousness is a matter of having "I" thoughts (thoughts immune to error through misidentification) we need to investigate the evolutionary path starting from its lowest stages, such as "the capacity to feel sensations" and agency. In fact, Bermudez argues that:

> Distinguishing self-awareness involves a recognition of oneself as a perceiver, an agent, and a bearer of reactive attitudes against a contrast space of other perceivers, agents, and bearers of reactive attitudes. It can only make sense to speak of the infant's experience of being a performer in the eyes of the other if the infant is aware of himself as an agent and of his mother as a perceiver (Bermudez 1998, pp. 252-253).

Many abilities usually related to "I" thoughts of language users would be detectable into a broadened non-conceptual point of view, but I think we need to take into account a further two cognitive abilities, so far forgotten and not in plain sight in Baker's account. First, we need conscious memory.

[5] See Gombrich (1963).

Without it, a child cannot distinguish his experience from permanent features of the environment which instantiate a given experience. You can find your way back to a particular place by sheer luck or, as Bermudez recognized, because you consciously remember it. A continuous present gives way to a temporally extended point of view. Moreover, this point of view also depends on "basic inductive generalizations at the non-conceptual level".

5. A Two-ply Account of Self-consciousness

I would add that it is not obvious that if we had an explanation of how we can entertain I-thoughts, then we would have explained everything there is to explain. What would remain to be explained is the phenomenal side of self-consciousness that is not reducible to the introspective accessibility to information. For example, according to Pietro Perconti (2008) the Thought-Language principle is wrong because it does not distinguish between the phenomenal aspects and the cognitive ones of self-awareness. Having the ability to refer to my-self my own mental states and being aware of them belongs to the cognitive aspect of self. The feeling of being yourself is instead something that has to do with the phenomenal aspects of the matter. To explain how we get I-thoughts is a psychological issue, but it leaves out the phenomenology linked to them. The notion of non-conceptual content allows us to introduce the idea of nonrepresentational properties, that is to say, a kind of sensational properties that an experience has in virtue of what it is like to have that experience.

It can be concluded that between the simple consciousness and self-consciousness would be appropriate to recognize intermediate states, in order to avoid reducing the first to the mere ability to intentionally generate relevant stimulus-response correlations, therefore making self-consciousness a function of language with the consequence, for example, of not attributing self-consciousness to people suffering from speech disorders, such as aphasia. Consciousness is not just a matter of ability to discriminate environmental stimuli and to select from a range of possible responses, but it is also a matter of being aware of this experience and feeling something while being in this state of awareness.

REFERENCES

Albergo, G. (2012), "Does Ontogenesis of Social Ontology start with Pretense?", *Phenomenology & Mind*, 3, pp. 120-126;

Albergo, G. (2013), *L'impegno ontologico del pretense*, Rivista di Estetica, 53, pp. 155-177;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, New York;

Bermúdez, J.L. (1998), *The Paradox of Self-Consciousness*, MIT Press, Cambridge (MA);

Carey, S., (1985), *Conceptual Change in Childhood*, MIT Press, Cambridge (MA);

Custer, W.L. (1998), "A Comparison of Young Children's Understanding of Contradictory Mental Representations in Pretence, Memory, and Belief", *Child Development*, 67, pp. 678-688;

Gombrich, E. (1963), *Meditations on a Hobby Horse and Other Essays on the Theory of Art*, Phaidon, London.

Harris, P.L. (2000), *The Work of the Imagination*, Blackwell, Oxford;

Keil, F.C. (1989), *Concepts, Kinds, and Cognitive Development*, MIT Press, Cambridge (MA);

Leslie, A.M., & Happé, F. (1989), "Autism and Ostensive Communication: The Relevance of Metarepresentation", *Development and Psychopathology*, 1, pp. 205-212;

Lillard, A. (2001), "Pretend Play as Twin Earth: A Social-Cognitive Analysis", *Developmental Review*, 21, pp. 495-531;

Meltzoff, A.N. (1990), *Foundations for a Developing Conception of the Self*, in D. Cicchetti & M. Beeghly (eds.), *The Self in Transition*, Chicago University Press, Chicago, pp. 139-164;

Nagel, T. (1979), *Mortal Questions*, Cambridge University Press, Cambridge;

Perconti, P. (2008), *L'autocoscienza*, Laterza, Roma-Bari;

Perner, J., Baker, S., & Hutton, D. (1994), "Prelief: The Conceptual Origins of Belief and Pretence", in C. Lewis & P. Mitchell (eds.), *Children's Early Understanding of Mind: Origins and Development*, pp. 261-286, Erlbaum, Hillsdale NJ;

Russell, J. (1996), *Agency, Its Role in Mental Development*, Taylor & Francis, Erlbaum (UK);

Searle, J. (1975), "The Logical Status of Fictional Discourse", *New Literary History*, 6, pp. 319-332.

GIUSEPPE LO DICO
Università Cattolica di Milano
giuseppe.lodico@unicatt.it

INTROSPECTION ILLUSION AND THE METHODOLOGICAL DENIAL OF THE FIRST-PERSON PERSPECTIVE

abstract

This paper will provide an evaluation of the Self/Other Parity Account, according to which introspection is an illusion and the data coming from it are unreliable for justifying theories. The paper will argue that the foundation of this account is based upon an a priori denial of the first-person perspective, considered as an obstacle to a full naturalization of psychology, that affects both the choice of the methods of inquiry and the interpretation of the empirical data.

keywords

Introspection, first-person, third-person

1. It is Better not to Trust the Subject!

As a psychologist, I wonder why how the epistemological status of my discipline is still a matter of controversy. I think that the very question on the table is nothing but whether it is a naturalized science, a social science, or something hybrid between these two ones. I also think that researchers working in (experimental) psychology tend to consider themselves as natural scientists, no more and no less than physicists or biologists. In this sense, they tend to argue that their object of inquiry – the mind – is something inter-subjectively observable through inter-subjectively validated methods, no more and no less than what is argued by physicists or biologists. This approach can be called *objective* or *third-person* and has allowed psychology to gain results and credit. In this sense, it can be defined as a fruitful approach. But the open question here is whether this approach is completely true, that is, whether it tells the whole story about the mental. Personally, I think it does not and I want to face this problem by discussing a specific issue: the place of introspection in psychology.

I here define introspection as an empirical method of enquiry through which subjects are able to learn and then verbally report about their own currently going on, or very recently past, mental states (Schwitzegebel 2010). It is worth noting that, even though the term introspection rarely occurs in recent textbooks of psychology or research methodology (Hurlburt & Heavey 2001, p. 401), we can find traces of it under various aliases, such as *verbal report* or *self-report*, as stated more than sixty years ago by the historian of psychology E. Boring (1953, p. 163). I am sure that those familiar with psychological literature have no difficulties to say that many experiments or review papers refer to data coming from verbal reports. The point is how researchers consider these data for the justification of their theories, especially in comparison with other kinds of evidence. My personal opinion is that verbal reports have no credibility among the majority of psychologists. For example, let us consider the following quotation from Philip Johnson-Laird, a prominent authority in the field.

> It is impossible to establish the veridicality of subjective reports. At worst, they may be fraudulent [...]; at best, they may be misleading, because none of us has access to the wellsprings of thought (Johnson-Laird 2006, p. 27).

As I said above, similar statements are quite common in psychological literature. These statements are formulated as showing a mere empirical fact generally accepted by the scientific community:

introspection is a sort of illusion, since people have many mistaken notions about their introspective information and its value (Pronin 2009, p. 3). For example, in a documented review ranging from many research areas such as social psychology (*ivi*, pp. 15-26 and pp. 49-51), developmental psychology (*ivi*, pp. 45-46), and neuroscience (*ivi*, pp. 48-49), the psychologist Pronin reports a large body of empirical evidence that aims at demonstrating that the *introspection illusion* can be a source of danger, since it "causes problems. It can foster conflict, discrimination, lapses in ethics, and barriers to self-knowledge and social intimacy" (*ivi*, p. 2). On the basis of the results reported in the review, she individuates four components of the illusion (*ivi*, pp. 4-6):

1. Introspective weighting: When people have to assess themselves, they generally tend to be too confident of their introspections.
2. Self/other asymmetry: When people have to assess the others, they generally do not rely upon introspection.
3. Behavioral disregard: People generally tend to disregard observable behavior when they have to assess themselves, and to take it in full consideration when they have to assess others.
4. Differential evaluation: People generally tend to take into great account their own introspections and to underestimate those of the others.

As we can see, Pronin argues that introspection is a potential source of biases and errors (*ivi*, p. 15) and thus hopeless as a method of scientific inquiry. So, she derives the methodological claim of mistrusting introspection and verbal reports from a large body of empirical evidence. This methodological claim implies the preference for non-introspective methods of any sort in psychological research, such as behavior observation, non-conscious priming, and brain neuro-imaging "[...]in order to pursue the goal of understanding mental experience" (*ivi*, p. 55). Thus, only third-person and inter-subjectively validated methods are allowed in the understanding of the mind: because of the biases and errors that introspection can provoke, researchers cannot trust what experimental subjects tell about their own point of view.

2. The Self/Other Parity Account

It is important to stress that, in order to qualify a method as introspective, it must meet, among the others, the so-called first-person condition. On this condition, introspection aims at generating knowledge, judgments, or beliefs about one's own mind and no one else's (Schwitzgebel 2010). In other words, this condition implies that, for making introspection, a person must adopt "[a perspective from which one thinks of oneself as an individual facing a world, as a subject distinct from everything else" (Baker 1998, p. 328). This does not mean simply that a person must possess a certain perspective towards the world and her thoughts (*ivi*, pp. 328-329). Rather, it means that she must possess the ability "to conceptualize the distinction between oneself and everything else there is" and "also to conceive of oneself as the bearer of those thought" (*ivi*, p. 330). That is, she must have a strong or robust (and not a weak or rudimentary) first-person perspective (Baker 1998, pp. 331-332; Baker 2013, pp. 147-150).

This implies that it is postulated an asymmetry between the way we can know our own mind and the way we can know others' minds: people cannot directly know others' minds through introspection, but they can indirectly know them only by making inferences from the observation of others' overt behavior. In other words, because of the first-person condition, introspection provides a sort of privileged access to the mind: the owner of mental states can have a better understanding of them than other people and thus, at the methodological level, psychologists can trust her when they ask about what happens in her mind. However, it is clear that this picture clashes with the conclusions reached by researchers emphasizing the presence of an introspection illusion. In fact, if we take a look at the four components of the illusion above mentioned, it appears evident that it is the first-person point of view that leads people to commit evaluation and judgment errors and thus to make the data coming from introspection scientifically unreliable. Roughly speaking, the four components aim at showing that the first-person perspective is so entrenched in subjectivity that it can lead to give unreliable interpretations of both inner mental states and outer behavioral phenomena.

For this reason, various researchers appear to endorse a position that is counterintuitive from the standpoint of common sense: the so-called Self/Other Parity Account. This account points out that people can have a reliable and adequate comprehension of their own mind only on the basis of the same processes through which they acquire knowledge of the others' minds rather than through introspection. Thus, according to the simplest version of the Self/Other Parity Account, the first-person condition cannot be met (Schwitzgebel 2010): this means that reliable introspections cannot be met in any way and people can know both their own mental states and those of others only indirectly.

It is clear that these arguments echo the old criticisms moved by the psychologist Watson against the late 19th and early 20th Century introspectionism in his 1913 behaviorist manifesto. However, I think it would be quite unfair to define the supporters of the Self/Other Parity Account as behaviorists, in spite of the similarities between them. In one of the most important contributions in favor of the Self/Other Parity Account, the developmental psychologist Gopnik strongly rebuts the charge of re-proposing a version of the old-fashioned behaviorism. In fact, differently from behaviorists, she stresses that internal psychological states do exist and that the discovery of their nature is the very aim of psychology: in this sense, the Self/Other Parity Account can be defined as a truly mentalist approach. She goes further by pointing out that there are also "[...] full, rich, first-person psychological experiences of the Joycean and Woolfian kind" (Gopnik 1993, p. 12). However, similarly to behaviorists, she points out that the first-person experiences cannot be considered as the genuine causes of people's thoughts and behaviors: this is so because first, people have internal psychological states, observe the behaviors and the experiences they lead to both in themselves and others; second, they build up theories about the causes of those behaviors and experiences that postulate the adoption of the first-person perspective; third, as a consequence, they experience the first-person perspective. This position is close to that proposed in one of the most-cited and controversial papers in psychology, that is, Nisbett and Wilson 1977, as explicitly claimed by Gopnik (1993, p. 9). The crucial argument of Nisbett and Wilson article can be briefly summarized in the two following points (Wilson 2002, p. 106):

1. Most emotions, judgments, thoughts, feelings, and behaviors are caused by an unconscious mind – in Gopnik terms, the *internal psychological states*.
2. Because people cannot have any conscious and first-person access to the unconscious mind, the conscious mind confabulates reasons – in Gopnik terms, *build up theories* – to explain emotions, judgments, etc.

A famous example in support of Nisbett and Wilson's view comes from the results of a selection task in which the participants were required to choose between four consumer products that were actually identical and to verbally justify their choice (Nisbett & Wilson 1977, pp. 243-244; see also Newell & Shanks 2014, p. 5). As a result, it was found that the participants tended to select the right-most of the four alternatives without mentioning the position as a justification of their choice. Rather, for this justification, they *built up a theory* based upon certain attributes of the chosen product. Thus, according to Nisbett and Wilson, the subjects' missed report about position effects on choice is evidence in favor of the dissociation between (unconscious) third-person psychological states and (conscious) first-person psychological experiences.

I think that the quotation of Johnson-Laird proposed above follows the same line of reasoning of

Gopnik and Nisbett and Wilson: most of our mental life is unconscious and, since verbal reports are not the product of any genuine introspection, but rather *post-hoc* theories of what is supposed to happen in the mind, they cannot be considered as reliable tools in psychological research. Thus, the Self/Other Parity Account appears to deny the use of introspective reports in the justification of psychological theories because they are irremediably biased by the subject's first-person point of view. In this sense, such a point of view seems to preclude the possibility to have reliable data at disposal.

3. Is the Empirical Evidence in Favor of the Introspection Illusion Really Grounded?

What amazes me of the literature in favor of the Self/Other Parity Account is the amount of empirical evidence reported for its justification (see Nisbett & Wilson 1977; Gopnik 1993; Wilson 2002; Pronin 2009). In this sense, as I pointed out above, the prevalence of an unconscious mind over the conscious appears to be an empirically well-grounded scientific theory. In fact, the claim that our unconscious mental states play a significant role in the determination of thoughts and behaviors seems to be empirically confirmed and generally accepted by the scientific community. However, all the scientific theories must be continually revised and put into question and psychological ones are not exceptions. Recently, the psychologists Newell and Shanks have proposed a critical review of the role of the unconscious mind on decision-making and have reached conclusions different from those of the supporters of the Self/Other Parity Account. The focus of their work is on the methods used to test whether experimental subjects are conscious or not of the mental processes involved in the determination of behavior during decision-making tasks (Newell & Shanks 2014, pp. 1-2). They point out that, in decision-making tasks, it is made a comparison between a behavioral performance and a conscious assessment based on subjects' verbal reports (*ivi*, p. 3): researchers infer that a mental state occurs unconsciously if the subjects' behavioral performance is clearly guided by this mental state but their verbal reports do not reflect it in any way. According to their proposal, in order to be reliable in assessing the presence or the absence of consciousness, a test must meet four criteria (*ivi*, pp. 3-4, Table 1):

> *a. Reliability: the assessment test must be unaffected by those factors that do not influence the behavioral performance.*
> *b. Relevance/Information: the assessment test must consider only the amount of information relevant to the behavioral performance or the decision in question.*
> *c. Immediacy: the assessment test must occur concurrently or as soon as possible after the behavioral performance to avoid possible lapses or distortions.*
> *d. Sensitivity: the assessment test should occur under optimal retrieval conditions.*

It is important to note that the idea behind their criteria for assessing consciousness dates back to two papers written by Shanks himself and the psychologist St. John in 1994 and in 1997. In these papers, Shanks and St. John provide a criticism to what they call the *Thesis of Implicit Knowledge and Learning*, according to which the most of people's knowledge is the primary cause of their behavior but it cannot be represented into consciousness. Further, also the learning of this knowledge takes place unconsciously both at the time of learning and at the time of retrieval (St. John & Shanks 1997, p. 164). According to their criticism, most studies in favor of the Thesis of Implicit Knowledge and Learning use invalid tests of consciousness, that is, tests clearly violating criteria (b) and (d) listed above (Shanks & St. John 1994, pp. 73-75 and p. 377; St. John & Shanks 1997, p. 167). For this reason, they conclude that the empirical evidence in favor of the Thesis of Implicit Knowledge and Learning is not as grounded as it might appear at a first sight (Shanks & St. John 1994, p. 367 and p. 394; St. John & Shanks 1997, pp. 162-163). In this sense, the paper by Newell and Shanks 2014 can be seen as an application of Shanks and St. John's (1994 and 1997) work to the area of decision-making – where the Thesis of Implicit Knowledge and Learning seems to be prevalent. However, the 2014 paper goes beyond the conclusions reached in the 1994 and 1997 papers. In fact, in the older articles, Shanks and St. John aim only to show that the reviewed studies using tests of consciousness violate the criteria (b) and (d). Instead, in the newer article, Newell and Shanks also show that the studies using tests of consciousness that respect the four criteria above described "either demonstrate directly that behaviour is under conscious control or can be plausibly explained without recourse to unconscious influences" (Newell & Shanks 2014, p. 19).

Thus, the points moved by Shanks and colleagues seem to overturn the picture sketched by the supporters of the Self/Other Parity Account. In fact, at the empirical level, they argue that the data coming from verbal reports, if adequately treated, cannot be defined as illusory or confabulatory in any way and can be legitimately used for justifying psychological theories. Instead, at the theoretical level, the mind seems to be much more conscious and introspectively accessible to a first-person perspective than many researchers can think, and the appeal to the unconscious in psychological theories often appears not to be justified.

4. Concluding Remarks

Now it is time to go back to the starting questions: should introspection and the data coming from it be eliminated from psychology? Should the methods of inquiry of psychology be limited only to the third-person and objective/inter-subjective ones? Is psychology a naturalized science? We have seen that Shanks and colleagues' work casts doubts on the supposed unreliability of the data coming from introspection and on the claim that most of the mind is unconscious (Newell & Shanks 2014, pp. 18-19). I think that their arguments are compelling and that the four criteria they propose should be met in the construction of every test for assessing consciousness. Personally, I think that these criteria are reasonable and not so difficult to be met and can provide a useful guide for evaluating the validity of the results of psychological research.

I believe that the most relevant conclusion of Shanks and colleagues' work can be summarized in this way: researchers should start to take subjects' introspective or verbal reports much more seriously than they actually do. However, to do this, they should also assume that the subject is able to adopt a first-person perspective allowing her to access her own mental states. This assumption seems to clash with the possibility of using only objective/inter-subjective and third-person methods for assessing psychological facts. That is, if we take a look all along the Newell and Shanks 2014 review, we can find that sometimes the authors must focus upon the single data obtained from a single participant for assessing their degree of consciousness and not only upon the results of the overall sample of subjects. This is clear, for example, in one of the studies that they review – and one of the few ones respecting the four criteria above discussed – that is, the Maia and McClelland 2004 paper on the re-examination of the Damasio's Somatic Marker Hypothesis, specifically when they discuss the results of two single participants, respectively number 36 and number 41 (Maia & McClelland 2004, pp. 4-5). As we can see, the adoption of the first-person perspective seems to imply the adoption of methods typical of an idiographic approach. This means that the primary goal of verbal or introspective reports is to provide an accurate description of a particular person's experiences, no matter whether they can be similar to or different from some or most other people's experiences (Hurlburt & Akhter 2006, p. 274). Of course, this does not appear to fit with an idea of psychology as a naturalized science, aiming at being nomothetic, objective/inter-subjective, and based on the average responses of a large sample of individuals to the introduction of some experimental manipulation in comparison with the response to certain control conditions (Hurlburt & Akhter 2006, p. 297; Barlow & Nock 2009, p. 19). To put it in another way, psychology cannot be limited to the study of the (universal) unconscious and sub-personal mechanisms necessary for a first-person perspective because the knowledge of these mechanisms cannot "supplant or replace knowledge of phenomena that the mechanisms make possible" (Baker 2007, p. 206). Personally, I do not want to argue that psychology should rebut the nomothetic approach in favor of the idiographic one: I believe that these two approaches should be viewed as the *methodological legs* of psychology, in spite of their irreconcilable differences, aims, and historical and philosophical traditions they come from (see von Wright 1971, Chapter 1). Thus, in my opinion, the denial of one of these two methodological tenets would lay psychology on the line to be incomplete. However, it is important to stress that these two approaches appear to be difficult to conciliate, since the explanations used in the idiographic approach cannot leave aside the adoption of the first-person perspective (see Baker 1998, pp. 336-337) and those used in the nomothetic one seem to work exclusively in a third-person perspective.

In conclusion, Shanks and colleagues' papers above considered suggest that there are no empirical reasons to reject the idea of a central role of the conscious mind in psychology. This is because the empirical results in favor of the unconscious mind appears to be theoretically affected by naturalistic presuppositions *a priori* dismissing the first-person perspective. Now, the question at play is no more empirical but philosophical/logical: can a naturalistic framework be a proper account for psychology? More precisely, if the acceptance of the first-person perspective appears to be undeniable for psychology, does the naturalism have the resources for coherently dealing with it? If not, the consequence should be to renounce to an idea of psychology as a fully naturalized science (see Baker 1998, pp. 336-337 and pp. 342-343 and Baker 2007) and to radically revise and reinterpret many psychological concepts and constructs.

REFERENCES

Baker, L.R. (1998), "The First-Person Perspective: A Test for Naturalism", *American Philosophical Quarterly,* 35, pp. 327–348;

Baker, L.R. (2007), "Naturalism and the First-Person Perspective", in G. Gasser (ed.), *How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society*, Ontos-Verlag, Frankfurt, pp. 203-226;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;

Barlow, D.H. & Nock, M.K. (2009), "Why Can't We be More Idiographic in Our Research?", *Perspectives on Psychological Science,* 4(1), pp. 19-21;

Boring, E.G. (1953), "A History of Introspection", *Psychological Bulletin,* 50(3), pp. 169-189;

Gopnik, A. (1993), "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality", *Behavioural and Brain Sciences,* 16, pp. 1-14;

Hurlburt, R.T. & Heavey, C.L. (2001), "Telling What We Know: Describing Inner Experience", *Trends in Cognitive Sciences*, 5, pp. 400-403;

Hurlburt, R.T. & Akhter, S.A. (2006), "The Descriptive Experience Sampling Method", *Phenomenology and the Cognitive Sciences*, 5, pp. 271-301;

Johnson-Laird, P.N. (2006), *How We Reason*, Oxford University Press, Oxford;

Maia, T.V. & McClelland, J.L. (2004), "A Re-Examination of the Evidence for the Somatic Marker Hypothesis: What Participants Know in the Iowa Gambling Task", *Proceedings of the National Academy of Sciences*, 101, pp. 16075-16080;

Newell, B.R. & Shanks, D.R. (2014), "Unconscious Influences on Decision-Making: A Critical Review", *Behavioral and Brain Sciences*, 37,pp. 1-61;

Nisbett, R.E. & Wilson, T.D. (1977), "Telling More than We Can Know: Verbal Reports on Mental Processes", *Psychological Review*, 84, pp. 231-259;

Pronin, E. (2009), "The Introspection Illusion", in M.P. Zanna (ed.), *Advances in Experimental Social Psychology*, 41, Academic Press, Burlington, United States, pp. 1-66;

Schwitzgebel, E. (2010), "Introspection", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, retrievable at http://plato.stanford.edu/entries/introspection/;

Shanks, D.R. & St. John, M.F. (1994), "Characteristics of Dissociable Human Learning Systems", *Behavioral and Brain Sciences*, 17, pp. 367-447;

St. John, M.F & Shanks, D.R. (1997), "Implicit Learning from an Information Processing Standpoint", in D.C. Berry (ed.), *How Implicit is Implicit Learning?*, Oxford University Press, Oxford, pp. 162-194;

Von Wright, G.H. (1971), *Explanation and Understanding*, Cornell University Press, Ithaca, United States;

Watson, J.B. (1913), "Psychology as the Behaviorist Views it", *Psychological Review*, 20, pp. 158-177;

Wilson, T.D. (2002), *Stranger to Ourselves. Discovering the Adaptive Unconscious*, The Belknap Press of Harvard University Press, Harvard, United States.

VALENTINA CUCCIO
Università degli Studi di Palermo
valentina.cuccio@unipa.it

THE NOTION OF REPRESENTATION AND THE BRAIN

abstract

The definition of the mechanism of Embodied Simulation is controversial. To account for this mechanism, Goldman and de Vignemont 2009 proposed the notion of mental representation in bodily format. In this paper I will offer arguments against the definition of mental representation in bodily format. To this purpose, I will specifically focus on the distinction between personal and subpersonal levels of explanation and on a first-person approach to the study of mental phenomena.

keywords

Embodied simulation, mental representation, first-person approach

1. Introduction

The term *representation* is, perhaps, one of the most contested expressions in the history of philosophy. Hundreds of pages would not suffice to summarize all of the different definitions and usages it has undergone in centuries of philosophical inquiry. Thus, my aim here must be far more limited. I will focus on some usages of this notion in the field of embodied cognition. Even in embodied cognition studies, the notion of representation has been seen in very different lights. Radical enactivists claim that we should get rid of this notion altogether, while many other supporters of embodiment believe that how the mind works cannot be explained without it. In fact, the use of the notion of representation in theories about human cognition has been considered as a demarcation line between radical and less radical embodied theorists (this distinction was introduced by Clark 1997; see also Chemero 2009; Alsmith & de Vignemont 2012). On one hand, radical embodied theorists claim that we can explain how the human mind works without resorting to mental representations (Kelso 1995; Port & van Gelder 1995; Thelen & Smith 1994; van Gelder 1995; Chemero 2009); on the other hand, exponents of moderate embodiment propose theories that include both representational and not representational explanations of human cognition (Barsalou 1999, 2008; Clark, 1997; Gallese & Sinigaglia 2011; Goldman & de Vignemont 2009).

I will not analyze all of the usages representation has undergone in embodiment literature here. Instead, I will focus on a more specific problem: to what extent can we define the mechanism of Embodied Simulation in terms of mental representations? Embodied Simulation is the activation of specific neural circuits of the brain that control actions, perceptions, the experiencing of bodily states or emotions when a person is not actively engaged in those actions, perceptions, bodily states or emotions. To give an example, the motor areas in my brain that control the action of grasping a cup will be activated not only when I effectively grasp a cup but also when I see a cup, when I observe someone else grasping a cup, when I read or listen to a proposition about someone grasping a cup or when I just imagine someone grasping a cup.

Today, the question about the legitimacy of the definition of the mechanism of Embodied Simulation in terms of mental representation is very urgent. The characterization of this mechanism in the current debate is quite controversial and, as a consequence, the definition of its role in human cognition is also quite controversial (see Mahon & Caramazza 2008 for a deeper discussion of this point).

In a recent paper, Goldman and de Vignemont (2009) proposed the notion of mental representations in bodily format. In the authors' words, these mental representations are identified with the activation of the mirror mechanism that gives rise to Embodied Simulation. This is clearly stated at the very beginning of their paper.

> We offer several interpretations of embodiment, the most interesting being the thesis that mental representations in bodily formats (B-formats) have an important role in cognition. Potential B-formats include motoric, somatosensory, affective and interoceptive formats. The literature on mirroring and related phenomena provides support for a limited-scope version of embodied social cognition under the B-format interpretation. (Goldman and de Vignemont 2009, p. 154).

The notion of mental representation in bodily format was later explicitly adopted by Gallese and Sinigaglia (2011) and a similar concept can also be found in Barsalou's idea of grounded symbols (Barsalou 2008). To what extent can we legitimately define Embodied Simulation in terms of mental representations? For the mechanism of simulation being considered as a mental representation, it would be necessary that we could clearly distinguish between the content and the format of the representation. Furthermore, it would also be necessary to identify the subject of the mental representation. I will suggest that neither of these criteria is matched by the notion of mental representation in bodily format. To this purpose, I will discuss some problematic issues related to the definition of this notion and I will specifically focus on the distinction between personal and subpersonal levels of explanation and on a first-person approach (Baker 2013) to the study of mental phenomena.

2. The Notion of Representation in Brain Sciences

The term *representation* has been widely used in brain sciences. In many cases, the use of this term does not imply any content-bearing state. To give an example, the use of this term has been fairly common in the somatotopic description of the brain. Somatotopy is the identification of the correspondences between areas of the brain and parts of the body. Somatopy tells us which part of the brain *represents*, namely *controls*, movements of the face, hands, legs and so on. These causal mappings between parts of the brain and certain areas of our body have been described and explained by means of an image, the Penfield *homunculus*, drawn in the human cortex. Although characteristics of the Penfield *homunculus* have been widely questioned, this terminology is still present in the debate. The *homunculus* drawn in the human cortex and its description in representational terms are just a visual metaphor of the correspondences between areas of the brain and parts of the body.

Cognitive neuroscientists also use the term representation when they refer to our motor repertoire. It is common, in this case, to say that neurons represent the goals of actions such as grasping or kicking. It is also common to talk about neurons that "represent" human faces or objects and so on. How can we interpret the use of the term representation in those cases? Are neuroscientists endorsing an intentionalist and representationalist account of what neurons do? Let us look closely at the case of motor neurons representing goals. A deeper analysis of this case will help us to show that the use of the term representation or of other expressions such as goals or intentions, that seem to ascribe mentalist and representationalist power to neural circuits, does not really imply a mentalist or representationalist explanation.

Neuroscientific evidence tells us that the motor cortex has a goal-centered organization (Umiltà et al. 2008). That is, there are neurons in the motor cortex that represent the goals of actions, independently of the specific movements we accomplish to carry out those actions. For example, Umiltà and colleagues (2008) showed that the "grasping" neurons in the F5 area of the motor cortex fire both when a monkey grasps an object by using normal pliers as well as when it uses inverse pliers. Normal and inverse pliers require hand-movement-patterns that are opposite from one another.

Physiology principles of correlational learning can explain how we build our motor repertoire and the goal-centered organization of the motor cortex. We know that neurons that fire together for a sufficient amount of time start to strengthen their mutual connection, thus creating a neural circuit. This principle is known as long-term potentiation (see Hebb 1949). By means of long-term potentiation and other phenomena of correlational learning, different areas of the motor cortex become wired together. That is, they are physically wired in a chain. Thus, neurons that usually fire in the F5 area of the monkey's brain when the monkey grasps an object with her fingers can also fire when the monkey grasps the object with normal or even inverse pliers. But to have this result, it is necessary to train the monkey. That is, it is necessary to create that physical chain of neurons. Umiltà and colleagues (2008, 2011) explain how this mechanism works:

> What could be the mechanism that allows a transformation of a goal into appropriate movements even when an opposite sequence of movements is necessary to achieve the goal? Our findings show that, after learning, the correct movement selection occurred immediately as soon as the monkey grasped one or the other type of pliers. This correct movement selection may be accounted for if one admits that goal-related F5 and F1g neurons are synaptically connected with two different sets of motor cortex neurons controlling the opening and the closing of the hand, respectively. These movement-related neurons, besides sending their output to the spinal cord, would also send a corollary discharge to the goal-related F5 and F1g neurons. In a natural setting, daily interactions with objects reinforce the connections that lead to the desired goal, thus selecting first those neurons that control hand opening and then those that control hand closure. After learning to use the reverse pliers, the opposite connections, reinforced by the success of the tool-mediated motor acts, prevail. As a consequence, the neurons that control hand closure are selected first, and those that control hand opening are selected subsequently (Umiltà et al. 2008, p. 2211).

It is clear that, in the case of neurons representing goals, we are describing physical chains of neurons that are synaptically connected and that are the result of correlational learning. In other words, this is an entirely mechanical and physical process. No *goal* is *represented* by those neurons that is in any sense different from a mechanical description. Similar results have also been observed in a study that recorded Motor Evoked Potentials to TMS from the right *opponens pollicis* of humans when using normal or reverse pliers or when observing others using the same tools, both with a specific goal (to grasp something) or without any specific goal (Cattaneo et al. 2009). According to the authors' interpretation of the data, the goal-centered organization of the motor cortex allows us to understand other people's goal-oriented actions by means of a kind of generalization. The observation of any grasping action, independently of the specific movement involved in the action, for example even when we are observing the use of tools such as inverse pliers, makes our implicit knowledge of real grasping available to us. Furthermore, according to the authors' interpretation, we can say that when we observe goal-oriented actions carried out by means of tools, the tool is incorporated in the observer's body-schema. This clearly explains why, when we observe people using reverse pliers with a goal, "grasping" neurons in the F5 motor cortex are activated by a pattern of movements that are the opposite of those of the real grasp or of a standard pliers grasp. In this case, reverse pliers become the distal effector that determines the activation of grasping neurons.

Should we endorse an intentionalist and representationalist interpretation of what neurons do? The fact that the same "grasping" neurons in the F5 motor cortex control flexor and extensor muscles being synaptically connected to both of them, as was suggested by Umiltà et al. (2008), and the hypothesis advanced by Cattaneo et al. (2009) about the incorporation of tools in the observer's body schema, which explained why "grasping" muscles are activated by the observation of grasping with reverse pliers that involve an opposite pattern of movement, seem to suggest that the use of the term representation in this case does not allow for a representationalist interpretation. In the first case, we have a physical chain of neurons that does not necessarily imply any representational relationship. In the second case, reverse pliers become a distal effector. The reverse pliers movement resembles the movement of a real grasping action and, when we observe or execute goal-oriented actions, pliers become our own fingers. Thus, when we observe someone else using reverse pliers to grasp something, the grasping neurons in the F5 motor cortex are activated because of the mechanism of simulation. As an alternative, we could describe these cases in terms of causal relations between physical events which, in turn, function as content vehicles. Thus, the goal-related F5 neurons would be bearers of a representational content. The point is, what would be the explanatory or predictive virtues of this representationalist explanation? What could this representationalist explanation add to our account of the motor system that a physical explanation cannot provide? So far, it seems that a representationalist explanation would not add anything to the picture we can sketch in physical terms.

Thus, the specific usages of the term representation discussed so far seem to be technical acceptations, internal to the neurophysiological and neuroscientific jargon. These usages do not necessarily commit neuroscientists to a representational theory of mind. However, recently and very often this technical acceptations of the term representation, in which the subject of the representation is a brain area or a particular neural circuit, is qualified as mental. Is this a correct move? Are those putatively *mental representations* the milestones on which we can construe our theory of language, social cognition, and so on? In the following section I will analyze one particular usage of the notion of mental representation in relation to neural facts, the notion of mental representation in bodily format proposed by Goldman and de Vignemont (2009).

3. Mental Representations in Bodily Format

Goldman and de Vignemont defined mental representations in bodily formats as those realized by means of the mechanism of Embodied Simulation (see the introduction to this paper). According to their definition, the activation of the hand-related areas of my motor cortex when I am looking at someone else grasping a cup is a mental representation, encoded in a motoric format, of the action of grasping a cup. Considering that the mechanism of simulation is not limited to motor areas of the brain but it is a widespread mechanism in our brains, in the same vein, we can have representations in somatosensory, affective or visual formats, and so on.

In a previous paper (Cuccio, submitted) I have already proposed two arguments against the definition of mental representation in bodily format. These arguments can be summarized as follows. First, we cannot define the mechanism of simulation as a representation because, in this case, it is not possible to distinguish between the content and the format of the representation. This distinction is implicitly present in our usages of the notion of representation and holds true even in very different philosophical traditions. The distinction between content and format is a necessary condition to define something as a representation and this distinction cannot be applied to the case of Embodied Simulation, where neurons firing are at the same time the format of the representation and its informational content. And, indeed, while Goldman and de Vignemont (2009, p. 155) make a clear distinction between content and format when they talk about representations with bodily content, when they define the notion of representation with bodily format, which is identified with the mechanism of simulation, they are no longer able to make such a distinction. The format of the representation is entirely identified by means of its content and the informational content cannot be truly distinguished from the format. This is evident from the examples the authors provide in their paper.

Second, it has been observed that in the case of neurons firing during the process of Embodied Simulation, we cannot clearly make a distinction between the role these neurons carry out in the circuit in which they are embedded and the putative information they should convey, since they are considered as a representation. In other words, a real occurrence of a phenomenon and the occurrence of its representation should differ, while in the case of the process of Embodied Simulation, they completely overlap (see Cuccio, submitted, for a deeper discussion of these arguments). In the next section, I will discuss another argument against the notion of mental representation in bodily format. Such argument is based on the distinction between the personal and subpersonal level of explanation (Dennett 1969) and on a first-person approach (Baker 2013) to the study of mental phenomena.

4. Personal and Subpersonal Levels of Explanation

The use of the term *mental* in the notion of mental representation in bodily format seems to be highly ambiguous. On the one hand, unless we are committed to a strong reductionist hypothesis, to qualify a process as a *mental* process seems to suggest that we are dealing with something that happens at the personal level, the level of the people experiencing and acting. The personal level is the level that we experience from a first person perspective. On the other hand, what Goldman and de Vignemont are referring to when they define mental representations in bodily format are the patterns of neural activation. The activation of neurons is a subpersonal physical process. We can only gain epistemological access to this kind of processes from a third-person perspective. Although subpersonal physical processes are constitutive of our experiences at the personal level (see Colombo 2013), if we describe our mental processes in the third-person perspective we will eliminate from our explanations our knowledge of these experiences in the first person perspective. Yet Goldman and de Vignemont (2009) propose their definition of mental representation in bodily format without subscribing to any form of reductionism. There seems to be a contradiction here. In fact, the authors seem to be far from proposing a redefinition of the human mind in biological terms. However, can patterns of neural activation be defined as mental processes without subscribing to reductionism?

It is worth noting here that, starting from the beginning of the 80s and during the 90s, researchers working in the Computational and Representational Theory of Mind proposed the idea of subpersonal mental representations as the building blocks of our cognition (Fodor 1998). These subpersonal mental representations were symbolic units of the Language of Thought that could be creatively combined according to syntactic rules and the principle of compositionality. Human cognition, it was suggested, can be entirely expressed in a propositional format.

Although this research paradigm had a great and long-standing influence on philosophical debate, it also faced some problematic aspects, particularly concerning the very same existence of those subpersonal mental representations. The question behind this problem was how a physical and mechanical system such as the brain can acquire representational content. Different solutions have been proposed to answer this question (e.g. Dretske 1981; Millikan 1984). However, all the programmes of naturalization of subpersonal mental representations had deep problems to solve such as: the problem of the indeterminacy of the content (how can a brain-state acquire a content and become a mental representation?); the problem of the decoupability of representations, which leaves very little explicative power to some of the solutions proposed; the problem of the gap between the subpersonal and the personal level of experiences; and the problem, originally proposed by Ryle, of the existence of a kind of practical, non-propositional, knowledge to ground propositionally structured knowledge. This kind of knowledge is not considered in the Representational and Computational Theory of Mind. Hence, this issue of the very same existence of subpersonal mental representations is highly problematic and still unsolved, even in the research paradigm that originally proposed this notion. This research paradigm has currently lost momentum, also in light of the embodied theories of cognition that are currently being proposed. Many empirical findings have widely shown that the existence of amodal symbols in the mind/brain is largely unfounded (see Barsalou 2008 for a discussion).

Let us now look closer at Dennett's distinction between the personal and subpersonal level of description. Daniel Dennett introduced the distinction between the personal and subpersonal level of explanation in 1969. To define it, he referred to the case of pain.

When we ask a person why he pulled his hand away from the stove, and he replies that he did so because it hurt, or he felt pain in his hand, this looks like the beginning of an answer to a question of behavioural control, the question being how people know enough to remove their hands from things that can burn them. The natural "mental process" answer is that the person has a "sensation" which he identifies as pain, and which he is somehow able to "locate" in his fingertips, and this "prompts" him to remove his hand. (Dennett 1969, p. 91).

The personal level of explanation pertains to "the explanatory level of people and their sensations and activities". The subpersonal level, on the other hand, concerns "the level of brains and events in the nervous system" (Dennett 1969, p. 93). In other words, personal level phenomena are those *mental* processes that characterise our life as subjects, as persons, while subpersonal phenomena are physical processes that we can describe in mechanical terms. Interestingly, in relation to the pain example, Dennett says:

> When we abandon mental process talk for physical processes we cannot say that the mental process analysis of *pain* is wrong, for our alternative analysis cannot be an analysis of pain at all, but rather of something else – the motion of human bodies or the organization of the nervous system (Dennett 1969, p. 93).

Thus, the personal level of explanation pertains to people. This is the level of beliefs and desires; this is the level of normative agents and moral responsibility.

Intuitively, this distinction is hard to eliminate. Lynne Baker (2013) proposes an argument against reductionism that is based on biological considerations. The argument runs as follows. Biologists say that the differences between human and other non-human primates are biologically insignificant. On the other hand, there is a tremendous difference between human and non-human primates in cognitive, socio-cognitive and communicative terms. From both these premises, it follows that we need to go beyond biology to make sense of this difference. A similar argument is presented by Michael Tomasello (1999) to explain the evolution of mankind. In his account, a solely biological explanation would not be enough to make sense of the extraordinary fast evolutionary path that led us to be human. We need to go beyond biology and also take into consideration the cultural dimension of human evolution (on the interaction between culture and biological evolution see also Deacon 1997).

The personal level is the level of our experiences. Our entire mental life, as well as all of our various forms of mental content, would be lost if we reduced mental processes to their subphysical mechanisms (Baker 2013) because we would abandon our first-person approach to these experiences. As Dennett says, if we talk about the physiology of pain we are not talking about pain anymore, about how it feels and how we react to it (similar arguments, Dennett acknowledges, can be already found in Ryle and Wittgenstein). If we abandon this distinction, then, we will inevitably and immediately loose the legitimateness of our first-person epistemological access to the world. This is the price to pay if we do not accept Dennett's distinction. On the other hand, to accept this distinction does not necessarily lead us to endorse a dualistic approach. And, in fact, Dennett, at least in his 1969 book *Content and Consciousness*, clearly claims an anti-dualist and anti-physicalist approach to the mind-body problem. We are what we are because we have the bodies that we have. That is, even if we cannot explain our personal level experiences in subpersonal terms and even if we cannot describe the activity of a physical subsystem in mental terms that is more isolated from the rest of the system, the qualifications that we ascribe to the whole system, the person, are necessarily dependent on the body that we have. Dennett's distinction is a distinction between levels of explanation; he is not claiming the existence of two different substances.

Things are different in many respects in Dennett's later works but I am not going to address this problem here. I will just keep that distinction between the personal and subpersonal levels of description on the table, as it was formulated in his 1969 book.

If we buy Dennett's definition of these two levels of explanation, then we cannot qualify the activation of the mechanism of simulation *per se* as a *mental* phenomenon. We cannot do that simply because we have already abandoned the level of people and their *mental* processes. We are talking about physical processes in the brain. Hence, as we have already seen, the only way to accept the definition of *mental* representation in bodily format and to be coherent would be to embrace a strong reductionist theory. That is, mental phenomena are physical phenomena that we can explain in terms of neural activity in the brain. Though, Goldman and de Vignemont do not seem to be committed to any reductionist hypothesis. Then, on the basis of these premises, it follows that their definition cannot be coherently accepted.

5. Conclusions

Having said that, what then about Embodied Simulation? Is the claim that Embodied Simulation cannot be considered as a mental representation equivalent to say that it is not necessary or not relevant in human cognition? My aim here is not to make such a claim. Embodied Simulation is a central and important mechanism in human cognition and a lot of empirical evidence supports this hypothesis. Empirical evidence suggests that Embodied Simulation is a part, at the subpersonal level, of the processes that allow us to comprehend other people actions or to understand language. Its role seems to be constitutive of the process of understanding and not merely causally correlated to it or just a side effect of the process of comprehension. In fact, when the mechanism of simulation is disrupted, for example artificially by means of TMS, the process of understanding is somehow impaired (see Pulvermüller 2013 for a discussion of the constitutive role of the mechanism of simulation during the comprehension of language).

It is then of paramount importance to rethink the mechanism of simulation and to describe its role without appealing to the notion of mental representation.

REFERENCES

Alsmith, A. & de Vignemont, F. (2012), “Embodying the Mind and Representing the Body”, *Review of Philosophy and Psychology*, Special Issue 3(1), pp. 1-13;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, New York;

Barsalou, L.W. (1999), “Perceptual Symbol Systems”, *Behavioral and Brain Sciences*, 22, pp. 577-609; discussion pp. 10-60;

Barsalou, L.W. (2008), “Grounded Cognition”, *Annual Review of Psychology*, 59, pp. 617-645;

Cattaneo, L., Caruana, F., Jezzini, A. & Rizzolatti, G. (2009), “Representation of Goal and Movements Without Overt Motor Behavior in the Human Motor Cortex: A Transcranial Magnetic Stimulation Study”, *Journal of Neuroscience*, 29, pp. 11134-11138;

Chemero, A. (2009), *Radical Embodied Cognitive Science*, MIT Press, Cambridge (MA);

Clark, A. (1997), *Being There: Putting Brain, Body, and World Together Again*, MIT Press, Cambridge (MA), London;

Colombo, M. (2013), “Constitutive Relevance and the Personal/Subpersonal Distinction”, *Philosophical Psychology*, 26(4), pp. 547-570;

Cuccio, V. (submitted), “Embodied simulation as bodily attitude. For a direct role of the body in language and cognition”, *Philosophical Psychology*.

Deacon, T.W. (1997), *The Symbolic Species: The Co-Evolution of Language and the Brain*, Norton, New York;

Dennett, D.C. (1969), *Content and Consciousness*, Routledge & Kegan Paul Humanities Press, New York;

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

Fodor, J. (1998), *Concepts: Where Cognitive Science Went Wrong*, Oxford University Press, New York;

Gallese, V. & Sinigaglia, C. (2011), “What Is So Special About Embodied Simulation?”, *Trends in Cognitive Sciences*, 15, pp. 512-519;

Goldman, A., & de Vignemont, F. (2009), “Is Social Cognition Embodied?”, *Trends in Cognitive Sciences*, 13 (4), pp. 154-159;

Hebb, D.O. (1949), *The Organization of Behavior: A Neuropsychological Theory*, John Wiley, New York;

Kelso, J.A.S. (1995), *Dynamic Patterns: The Self-Organization of Brain and Behavior*, MIT Press, Cambridge (MA), London;

Mahon, B.Z. & Caramazza, A. (2008), “A Critical Look at the Embodied Cognition Hypothesis and a New Proposal For Grounding Conceptual Content”, *Journal of Physiololgy of Paris*, 102, pp. 59-70;

Millikan, R. (1984), *Language, Thought and Other Biological Categories*, MIT Press, Cambridge (MA);

Port, R.F. & van Gelder, T. (1995), *Mind as Motion: Explorations in the Dynamics of Cognition*, MIT Press, Cambridge (MA), London;

Pulvermüller, F. (2013), “How Neurons Make Meaning: Brain Mechanisms for Embodied and Abstract-Symbolic Semantics”, *Trends in Cognitive Science*, 17, pp. 458-470;

Thelen, E. & Smith, L.B. (1994), *A Dynamic Systems Approach to the Development of Cognition and Action*, MIT Press, Cambridge (MA), London;

Tomasello, M. (1999), *The Cultural Origins of Human Cognition*, Harvard University Press, Cambridge (MA), London;

Umiltà, M.A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Jezzini, A., Gallese, V. & Rizzolatti, G. (2008), “When Pliers Become Fingers in the Monkey Motor System”, *Proceedings of the National Academy of Science USA*, 105, pp. 2209-2213;

Van Gelder, T. (1995), “What Might Cognition Be, if Not Computation?”, *The Journal of Philosophy*, 92(7), pp. 345-381.