ISSN 2280-7853

PHENO-
MENO-
LOGY
AND
MIND

PHENOMENOLOGY AND MIND

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

12

STUDIORUM UNIVERSITAS
FIRENZE
UNIVERSITY
PRESS

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

n. 12 - 2017

PHENOMENOLOGY AND MIND

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

NEW TRENDS IN PHILOSOPHY

Edited by Laura Caponetto and Bianca Cepollaro

FIRENZE
UNIVERSITY
PRESS

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

CC 2017 Firenze University Press
Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28, 50122 Firenze, Italy
www.fupress.com

Phenomenology and Mind. The Online Journal of the Faculty of Philosphy,
San Raffaele University, on-line: http://www.fupress.net/index.php/pam

Cognitive Neurosciences, Philosophy of Mind and Language, Logic (CRESA)

CONTENTS

INTRODUCTION

INVITED CONTRIBUTIONS

SUBMITTED CONTRIBUTIONS

CONTENTS

INTRODUCTION

INTRODUCTION

Laura Caponetto, Bianca Cepollaro
A Snapshot of a New Generation of Philosophers

LAURA CAPONETTO
Vita-Salute San Raffaele University
lauracaponetto@gmail.com

BIANCA CEPOLLARO
IFILNOVA, Lisbon
bianca.cepollaro@gmail.com

A SNAPSHOT OF A NEW GENERATION OF PHILOSOPHERS

Phenomenology and Mind, n. 12 - 2017, pp. 10-15
DOI: 10.13128/Phe_Mi-21101
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

A SNAPSHOT OF A NEW GENERATION OF PHILOSOPHERS

1. New Trends in Philosophy

This Special Issue of *Phenomenology and Mind* (“New Trends in Philosophy”) gathers the works of young philosophers from all over the world, from master students to PhD candidates, post-doctoral fellows, and young researchers. It aims to draw a picture of the directions in which philosophy is heading and provide a critical overview of some of the most interesting topics and methodologies in the current philosophical debate. The volume consists of four invited papers and nineteen contributed papers that were selected through a double-blind peer review process.

The Issue is a portrait of state-of-the-art research in many different areas of philosophy, with a particular focus on philosophy of language, philosophy of mind and psychology, (neuro) phenomenology, and moral philosophy. Further areas it deals with include philosophy of science, philosophy of mathematics, metaethics, political philosophy, and history of philosophy. Many contributions to the volume adopt an interdisciplinary stance, where philosophy engages in a fruitful dialogue with other disciplines: physics, neurosciences, cognitive sciences, psychology, linguistics, psycholinguistics, biology, and much more. Such a rich variety of perspectives offers an intriguing snapshot of a new generation of philosophers.

2. Contents

2.1. Invited Contributions

As invited authors for this Special Issue, we chose young and prominent philosophers from different fields who are representative of what good philosophy looks like nowadays. The section dedicated to invited papers opens the volume.

Valeria Giardino (CNRS/Laboratoire d’Histoire des Sciences et de Philosophie - Archives Henri-Poincaré) kick-starts the section with her paper “The Practical Turn in Philosophy of Mathematics: A Portrait of a Young Discipline”, where she discusses the so-called *philosophy of mathematical practice*, in relation to the practical turn in philosophy of science and in philosophy of mathematics. Giardino goes through various approaches to the practice of mathematics and discusses the possible replies to the question as to what counts as mathematical practice.

In her invited paper “New Wine in Old Bottles: The Kind of Political Philosophy We Need”, Beatrice Magni (University of Milan) investigates the meaning and the role of contemporary political philosophy by exploring the relation between philosophy and politics. In particular, she is interested in what kind of political practice is able to reconcile the normative commitments of political philosophy (Rawls, 2007) with its actual and feasible goals (Hall, 2015; Galston, 2007).

The third invited paper, “What Metalinguistic Negotiations Can’t Do” by Teresa Marques (LOGOS Group, University of Barcelona), tackles a hot question in philosophy of language and metaethics, namely the role of metalinguistic negotiation in normative and evaluative disagreement. *Contra* Sundell (2016), Marques argues that metalinguistic negotiations are neither necessary nor sufficient for genuine evaluative and normative disputes, as for Marques value talk requires stronger metanormative commitments.

The section dedicated to invited contributions ends with a paper in philosophy of time. In "The Myth of Presentism's Intuitive Appeal", Giuliano Torrengo (Centre for Philosophy of Time, Department of Philosophy, University of Milan) questions the intuitive appeal of presentism, the view according to which only what is present exists. While the intuitive character of such a view seems to constitute a main reason for taking it into consideration, Torrengo goes through the misconceptions on which such appearance of intuitiveness is based.

2.2. Submitted Contributions

The section dedicated to contributed papers features contributions from various fields in philosophy too: we observe a special focus on philosophy of language, philosophy of mind and psychology, (neuro)phenomenology, philosophy of science, moral philosophy, and history of philosophy.

In "Contextualist Answers to the Challenge from Disagreement", Dan Zeman (Institute of Philosophy, University of Vienna) offers a survey of the most recent contextualist answers to the *challenge from disagreement* raised by contemporary relativists. The challenge constitutes one of the main objections against contextualism in philosophy of language. While providing an overview of the latest dialectical moves of the debate, Zeman critically discusses the main strategies available to the contextualist and formulates some objections against them.

"How to Dispel the Asymmetry Concerning Retraction", by Diogo Santos (LanCog, University of Lisbon), discusses MacFarlane (2014)'s *assessment-sensitivity* and addresses the asymmetry concerning retraction identified by Ferrari & Zeman (2014). Although assessment-sensitivity predicts that speakers ought to retract previous assertions whose content is deemed false, MacFarlane argues that retracting is not necessarily "admitting fault" (2014, p. 110) – where the sense of not being at fault invoked is distinctively epistemic. Ferrari & Zeman (2014) identify an asymmetry between retractions involving predicates of personal taste and moral terms that MacFarlane's epistemic notion of "being at fault" cannot explain. In his contribution, Santos provides a way to dispel such an asymmetry and concludes that assessment-sensitivity needs no supplementation in order to account for it.

Simone Carrus (Vita-Salute San Raffaele University), in "Slurs: At-issueness and Semantic Normativity", discusses the interaction between slurs and denial as well as the ways in which denial is taken to be effective in targeting the descriptive and evaluative content of certain expressions. After dealing with some theoretical issues, Carrus applies the test of denial to an utterance extracted from a recent juridical case with the aim of investigating the content of slurs in non-standard uses.

In "Thomason (Un)conditionals", Andrés Soria Ruiz (ENS-PSL, Paris - Institut Jean Nicod) considers utterances of the form *if p, ~Kp* – such as "If my coworkers hate me, I have absolutely no idea" – known as *Thomason conditionals*. The author discusses the ways in which these sentences pose problems for epistemic theories of indicative conditionals. Soria Ruiz aims to show that Thomason examples are *not* in fact indicative conditionals, but alternative unconditionals, in the sense put forward by Rawlins (2013).

Paolo Labinaz (University of Trieste), in "Assertion and the Varieties of Norms", challenges Cappelen (2011)'s claim that assertion does not correspond to a speech-act category, as for Cappelen there is no satisfactory criterion to distinguish between utterances that are assertions and utterances that are not. While adopting an "Austin-inspired" framework, Labinaz claims that there are in fact some norms that can be seen as applying to assertions in a specific way.

Enrico Cipriani (University of Turin), in "Chomsky on Analytic and Necessary Propositions", discusses Chomsky's view on the analytic-synthetic distinction and on necessary propositions. Cipriani underlines how Chomsky's defense of such a distinction can hold only under the assumption of conceptual innateness. Furthermore, Cipriani notes that, in Chomsky's view, the distinction between necessary and contingent truths is determined by the structure of the conceptual system and its relations with other systems of common-sense understanding. But such a hypothesis, Cipriani argues, seems to be incompatible with Chomsky's own objection to Kripke's essentialism.

In "The Two-Way Relationship Between Language Acquisition and Simulation Theory", Hashem Ramadan (Boğaziçi University) draws a two-way connection between Simulation Theory and language acquisition. The idea is that, on the one hand, if an individual has better simulation capabilities, then she will be better when it comes to L2 acquisition; on the other hand, being exposed to different languages seems to lead to better simulation capacities and higher degrees of empathy. Drawing on an evolutionary explanation, Ramadan argues in favor of Simulation Theory over Theory Theory and discusses some studies involving children with ASD which provide support for it.

Marco Fenici (University of Florence), in "Rebuilding the Landscape of Psychological Understanding After the Mindreading War", addresses the intricate net of connected debates in philosophy and cognitive sciences about the onset, the development, and the nature of mindreading mechanisms. Fenici discusses the contribution of each debate and the ways in which philosophy and cognitive sciences have or have not fruitfully interacted thus far.

Alessandra Buccella (University of Pittsburgh), in "Naturalizing Qualia", puts forward an alternative to Hill (2014)'s naturalization of qualia. For Hill, perceptual qualia (*i.e.*, the ways in which things look from a viewpoint) are physical properties of objects and are relational in nature – that is, they are functions of objects' intrinsic properties, viewpoints, and observers. After analyzing the weaknesses of Hill's account, Buccella builds upon Chirimuuta (2015)'s *color adverbialism* and argues for a broadly adverbialist view of perceptual qualia.

"Carving Mind at Brain's Joints. The Debate on Cognitive Ontology", by Marco Viola (IUSS Pavia and Vita-Salute San Raffaele University), assesses the vexed mind-brain problem; in particular, he discusses the traditional hypothesis of a one-to-one mapping between mental states and neural activities and the shortcomings of this sort of "new phrenology". Viola explores two ways to avoid such weaknesses: the first endorses a many-to-many mapping model, whereas the second radically rethinks its *relata*.

Joana Rigato (Champalimaud Center for the Unknown, Lisbon), in her paper "Looking for Emergence in Physics", discusses a topic on which philosophers and physicists often talk past each other: *emergence*. Emergentism, in its different forms, is the view that certain features of reality (be they objects, properties or laws) are irreducible to the lower-level bases they emerge on. After going through some examples of emergence in (classical and quantum) physics, Rigato concludes that paradigmatic examples of discontinuity between models in physics can back the emergentist philosopher's case up against reductionist theories.

"Direct Social Perception of Emotions in Close Relations", by Andrea Blomqvist (University of Sheffield), explores the theory of Direct Social Perception with respect to perceiving the emotional states of our closest ones (spouses, friends, and family). Blomqvist argues that emotions are embodied and can be directly perceived. Moreover, she argues against a non-conceptual view of emotion recognition and claims instead that by attending to certain expressive patterns of emotions, we can learn "emotional concepts". This view predicts that we can directly perceive both basic and non-basic emotions of people we are close to.

In "Me, You and the Measurement. Founding a Science of Consciousness on the Second Person Perspective", Niccolò Negro (University of Milan) critically assesses the methodologies involved in the study of consciousness while discussing whether they adopt a first-, second- or third-person perspective. In particular, he argues that Integrated Information Theory is the approach that is most likely to account for a measure and a mathematical analysis of conscious experience.

Timothy A. Burns (Loyola Marymount University), in "Empathy, Simulation, and Neuroscience: A Phenomenological Case against Simulation-Theory", questions the claim that the discovery of mirror neurons provides empirical support for the simulation view of mindreading. In addition to formulating multiple objections against Simulation Theory, Burns draws on the works of Edmund Husserl and Edith Stein and proposes a phenomenological account to mindreading.

In "On Experiencing Meaning: Irreducible Cognitive Phenomenology and Sinewave Speech", John Joseph Dorsch (University of Tübingen) deals with the phenomenon of sinewave speech (*i.e.*, a synthetic acoustic signal that replaces the original human voice's formants with pure tone whistles). When subjects first hear sinewaves, all that they can discern are beeps and whistles; however, after listening to the speech from which the sinewave is derived, beeps and whistles actually sound like speech. Granted that the two episodes (whistles *vs.* speech) differ in their phenomenal character, Dorsch investigates whether and to what extent such an alteration in phenomenal character may provide evidence for irreducible cognitive phenomenology.

Joe Higgins (University of St. Andrews and University of Stirling - SASP) discusses the tension within cognitive scientific accounts of human selfhood between bodily processes and social processes in his paper "Embodied Mind – Ensocialled Body: Navigating Bodily and Social Processes within Accounts of Human Cognitive Agency". Drawing on a range of phenomenological and empirical insights, Higgins argues for the concept of an "ensocialled body", in which all organic bodily processes are at the same time also social processes.

In "Biology, Justice and Hume's Guillotine", Hugo de Brito Machado Segundo (Federal University of Ceará, UFC - Brazil) and Raquel Cavalcanti Ramos Machado (Federal University of Ceará, UFC - Brazil) discuss the role of philosophy in the investigation of moral sentiments and address the question as to how the discovery that moral sentiments have evolutionary origins interacts with the problem of *Hume's Guillotine*. The authors explore the ways in which certain features of human beings that can be accounted for in terms of natural selection are culturally promoted or discouraged.

"On Solidarity: Gramsci's Objectivity as a Corrective to Buber's I-It", by Ryan Adams (Franciscan University of Steubenville), addresses a dichotomy in human interactions: on the one hand, there is the merely objective *I-It* interaction; on the other, there is the intense intersubjective relationship of the *I-Thou*. Adams posits the *principle of solidarity* as a *non-I-Thou* relation that retains the dignity and the personhood of the Other in a way that still confines her/him to the "It" of Buber's *I-It* pair. In the second part of the paper, Adams shows how – in such a framework – solidarity functions as Gramsci's *Objectivity*.

The section dedicated to contributed papers is closed by "The Italian 'Difference': Philosophy Between Old and New Tendencies in Contemporary Italy", by Corrado Claverini (Vita-Salute San Raffaele University). The author discusses the legitimacy, the risks, and the benefits of the tendency of Italian philosophy to reflect on itself. Moreover, Claverini identifies the distinctive hallmarks of the Italian philosophical tradition from the Renaissance to today in "precursory genius", in ethical and civil vocation, and in the so-called *living thought*.

3. Acknowledgements

We are deeply grateful to all who supported this project.

First, we shall thank Valeria Giardino, Beatrice Magni, Teresa Marques, and Giuliano Torrengo for having accepted our invitation to contribute to this Special Issue: the whole project has

very much benefited from their excellent works as well as from their support and enthusiasm. We are grateful to all the many authors who submitted their works to *Phenomenology and Mind*. Special thanks to the hundreds of anonymous referees that reviewed submitted papers: evaluating an article is a job that scholars perform for free and anonymously, while both the authors and the journal greatly benefit from their comments and expertise. They did a great service to both accepted and rejected papers and we would like to express our deepest gratitude for their insightful comments and helpful suggestions.
Moreover, let us thank Roberta De Monticelli, editor-in-chief of the Journal, and the whole Editorial Board for choosing us as the guest-editors for this Special Issue: it has been a great opportunity and a great journey. Finally, we would like to thank all those who helped us putting this Issue together, by generously offering us their skills, time and experience: Claudia Bianchi, Francesca Boccuni, Roberta De Monticelli, Francesca De Vecchi, Francesca Forlè, Antonio Moretti, Massimo Reichlin, Elisabetta Sacchi, Roberta Sala, Sarah Songhorian. Heartful thanks to Claudia Bianchi for her precious support, her valuable and infallible advice and much more.

REFERENCES

Cappelen, H. (2011). Against Assertion. In J. Brown & H. Cappelen (Eds.), *Assertion: New Philosophical Essays*. Oxford: Oxford University Press, 21-48.
Chirimuuta, M. (2015). *Outside Color*. Cambridge, MA: MIT Press.
Ferrari, F., & Zeman, D. (2014), Radical relativism, retraction and “being at fault”. In S. Caputo, M. Dell’Utri, & F. Bacchini (Eds.), *New frontiers in truth*. Newcastle: Cambridge Scholar, 80-102.
Galston, W.A. (2010). Realism in Political Theory. *European Journal of Political Theory*, 9(4), 385-411.
Hall, E. (2015). How to do realistic political theory (and why you might want to), *European Journal of Political Theory*, 1-21, DOI: 10.1177/1474885115577820.
Hill, C. (2014). *Meaning, Mind, and Knowledge*. Oxford: Oxford University Press.
MacFarlane, J. (2014). *Assessment sensitivity: Relative truth and its applications*. Oxford: Oxford University Press.
Rawlins, K. (2013). (Un)conditionals. *Natural Language Semantics*, 21(2), 111-178.
Rawls J. (2007). *Lectures on the History of Political Philosophy*. Boston, MA: Harvard University Press.
Sundell, T. (2016). The tasty, the bold, and the beautiful. *Inquiry*, 59(6), 793-818.

INVITED CONTRIBUTIONS

INVITED CONTRIBUTIONS

Valeria Giardino
The Practical Turn in Philosophy of Mathematics: A Portrait of a Young Discipline

Beatrice Magni
New Wine in Old Bottles: The Kind of Political Philosophy We Need

Teresa Marques
What Metalinguistic Negotiations Can't Do

Giuliano Torrengo
The Myth of Presentism's Intuitive Appeal

VALERIA GIARDINO
CNRS/Laboratoire d'Histoire des Sciences et de Philosophie
Archives Henri-Poincaré
valeria.giardino@univ-lorraine.fr

THE PRACTICAL TURN IN PHILOSOPHY OF MATHEMATICS: A PORTRAIT OF A YOUNG DISCIPLINE

abstract

In the present article, the current situation of the so-called philosophy of mathematical practice is discussed. First, its emergence is evaluated in relation to the "practical" turn in philosophy of science and in philosophy of mathematics. Second, the variety of approaches concerned with the practice of mathematics and the new topics being now object of research are introduced. Third, the possible replies to the question about what counts as mathematical practice are taken into account. Finally, some of the problems that are still open in the philosophy of mathematical practice are presented and some possible new directions of research considered.

keywords

philosophy of mathematics, practical turn, philosophy of mathematical practice, mathematical practice

Phenomenology and Mind, n. 12 - 2017, pp. 18-28
DOI: 10.13128/Phe_Mi-21102
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

THE PRACTICAL TURN IN PHILOSOPHY OF MATHEMATICS

1. Introduction: The Practical Turn, From Science to Mathematics

Starting from the 1970s, philosophy of science has shown a tendency to move more and more away from the definition of abstract and general theories about scientific knowledge towards the investigation of the concrete work that scientists as practitioners in a particular scientific field are engaged in everyday. This change of perspective has commonly been labeled as the "practical" or "practice" turn:[1] science is not conceived anymore as true and justified belief that has to be examined, so to speak, *in vitro*, but as a mingle of different practices that should be considered *in vivo*, looking at the behaviors and the habits characterizing the people involved in them.

There were reasons for this new perspective to emerge. First, in reaction to views of science that were too disembodied: science cannot be science *from nowhere*; second, in opposition to the "rational reconstructions" that were typical of the philosophy of science of the beginning of the 20th century: science cannot either be science *from anywhere*. According to this new line of thought, the scientific enterprise is analogous to other *human* practices: it is historically and culturally situated, and the task of philosophy should be to clarify its specificities.

However, the consequences of such a practical "revolution" are still under discussion. First of all, was it really a revolution? And if this is the case, in which way was it revolutionary? What are the new objects of research? What is its methodology? In other words, where exactly has the practical turn led us?[2] These questions and analogous ones have been at the center of the recent debate, starting from famous proposals such as Kuhn (1962)'s notion of paradigm up to naturalistic approaches, which totally deny the possibility of *a priori* knowledge and embrace some form of empiricism or pragmatism.

Many of these issues go back to a crucial ambiguity concerning in general "practice-based" approaches. How is 'practice' defined? Is there just one practice or are there many? And if there are many, what are their common features?[3] As Salanskis (2014) has very conveniently summarized,

> if the very program of practice turn is to be significant, it must recommend looking at science in a specific way, which can then be contrasted with other ways. This absolutely

[1] For details about the introduction of this term, see Soler *et al.* (2014, p. 39, n. 1).

[2] For an in-depth discussion of these topics, see Soler *et al.* (2014).

[3] Incidentally, analogous difficulties may rise for the notion of "paradigm" as introduced by Kuhn (1962).

> requires that we have at our disposal a non universally encompassing notion of practice (p. 44).

Compared to the case of philosophy of science, a practical turn in philosophy of mathematics has happened later and only partially.[4] As Hersh (2005) puts it, in the 1970s philosophy of science was already in its renaissance, while most of philosophy of mathematics rather looked as "foundationalist ping-pong" (p. vii). An evident exception was Lakatos' *Proofs and Refutations* (1975), a book which, according to Hersh, is surely fascinating but was however virtually unknown at the time. Moreover, despite Lakatos' approach finds inspiration in Polya's previous work on problem-solving and is undoubtedly new and original, his book ends up being a sort of hybrid, caught between the before and after the practical turn. In fact, the dialogue between the characters in the book still offers a rational reconstruction of the Theorem of Descartes-Euler, even if it is a reconstruction of the *process* leading to its mathematical proof and not of its mathematical proof itself. However, in the footnotes there happens to be another book, which tells us the story of the development of the same theorem. The result is somewhat schizophrenic: the theorem is an historical product but it can still be discussed "from anywhere".

The situation did not change in the following decades, with some exceptions, such as the collection edited by Tymoczko (Ed.) (1986; rev. ed. 1998), which contained articles criticizing the contemporary philosophy of mathematics, or the one edited by Aspray and Kitcher (Eds.) (1988), which had a strong interdisciplinary line of attack to modern mathematics. In the 1990s, van Bendegem (1993) writes as follows: "if science is what scientists do, as it has become fashionable to claim, should it then not be the case that mathematics is what mathematicians do?" (p. 263). Kuhn himself would be reluctant to reply unhesitatingly 'yes' to such a question, since in his view mathematics has a special status that seems to elude the typical complications having to do with change and scientific development; by contrast, Lakatos' answer would undeniably be positive. However, as van Bendegem points out, differently from philosophy of science, in the philosophy of mathematics of the 1990s there still existed something like a "received view": mathematics tended to be considered always – and wrongly, according to him – as "the exact science".

A sharp distinction between the research in philosophy of science and in philosophy of mathematics is however very surprising: on the one hand, mathematics is a scientific discipline; on the other hand, science makes use of many mathematical tools.[5] However, because of its supposed special status compared to other scientific disciplines, mathematics has also built strong connections with formal logic, a discipline that did not exist before the work of Frege. In fact, a drastic change took place even before Frege: if up to the 19th century mathematics was conceived as a discipline that, in one way or another, aimed at describing the real world, it later became an independent corpus of ideas that are proper to mathematics and mathematics only. This conception brought to new questions relative to the nature of mathematics that are very familiar today: if mathematics does not speak about the real world, what is its object of study? It also led to the search for stable foundations for mathematics. For these reasons, at the turn of the 20th century, philosophy of mathematics was specifically concerned with topics such as justification and formal mathematics, leaving

[4] This is even more striking for philosophy of logic, but for reasons of space I won't expand on this issue here.

[5] The argument of the indispensability of mathematics for the sciences has also brought about reasons to believe in the existence of mathematical entities. For a recent debate on the subject, see for reference Panza and Sereni (2013, chapters 6 and 7).

aside more epistemological issues.[6] Recently however, an interest toward philosophical questions about the practice of mathematics has finally risen, undoubtedly in reaction to this neglect. Mancosu, in his introduction to a collection of essays published less than 10 years ago entitled *The Philosophy of Mathematical Practice*, claims that the works in the book represent "the first steps in a very difficult area and we hope that our efforts might stimulate others to do better" (p. 20). In 2009, the *Association for the Philosophy of Mathematical Practice* (APMP) was created.[7]

The picture of the philosophy of mathematical practice emerging today is very varied and heterogeneous. In Section 2, I will present some possible ways of sorting out the different approaches that have been proposed so far in this domain. In Section 3, I will briefly go back to the problem of how to define a mathematical practice. Finally, in Section 4, I will sketch a picture of the philosophy of mathematical practice today and draw some tentative conclusions about the most crucial issues that remain open.

2. Not Just One Philosophy of Mathematical Practice (and more)

One evident feature of the philosophy of mathematical practice in 2017 is that it is characterized by a variety of different proposals and different methodologies, which only in some cases overlap or happen to be complementary. Recently, some authors have tried to identify the distinct orientations. I will focus in particular on two proposals.

Van Bendegem (2014) presents a list of eight disciplines that look at mathematics from the point of view of its practice: (1) the *Lakatosian* approach, namely the "maverick" tradition; (2) the *descriptive analytical naturalizing* approach; (3) the *normative analytical naturalizing* approach; (4) the *sociology of mathematics* approach; (5) the *mathematics educationalist* approach; (6) the *ethno-mathematical* approach; (7) the *evolutionary biology* of mathematics; and (8) the *cognitive psychology* of mathematics. For him, only the first three perspectives have a distinct philosophical nature, which means that philosophy is not alone in the enterprise; this establishes already a difference with the received view of mathematics as the exact science. The Lakatosian approach goes back to *Proofs and Refutations*: to make sense of the development of mathematics, it is essential to understand discovery processes. This is in tension with a second tradition that mainly refers to Kitcher's work.[8] According to this tradition, which van Bendegem labels "analytical and naturalizing", the object of research is the final version of the proof and the *desideratum* is to find out what is needed to justify the claim that such a final version is indeed a proof. Differently from the first approach, such a justification is here totally independent of the process that has brought to the proof. In fact, the analytical naturalizing approach is related to the more general agenda in analytic philosophy and in particular to Quine's program of naturalizing epistemology, and breaks down into the *descriptive* and the *normative* analytic naturalizing approaches. For the descriptive approach, the methodology is simply to relate what mathematicians think about their everyday work, for example about what counts for them as a proof; for the normative approach, the methodology is to examine the nature of the proofs that are put forward by mathematicians and to establish whether they are genuine proofs, no matter what the mathematicians' beliefs are. Proposals from (4) to (8)

[6] In Agazzi and Heinzmann (2015)'s reconstruction, it is precisely to overcome the foundational crisis that affected the exact sciences, mathematics and physics, at the end of the 19th and at the beginning of the 20th century, that new trends and ideas in philosophy were produced and philosophy of science in its contemporary sense was born. Such a crisis challenged "the pervasive positivist view that had attributed to science the monopole of secure knowledge and the role of being the ground of human progress" (p. 8).

[7] In 2017, the APMP will celebrate its fourth international meeting. See for information http://institucional.us.es/apmp/.

[8] See for reference Kitcher (1984). We will go back to Kitcher's views in the remainder of the article.

are related to disciplines other than philosophy that are however interested in the practice of mathematics.[9]

From the emerging picture, it is evident that a unity of these distinct approaches is questionable and as a consequence the possible unity of the study of mathematical practice with traditional philosophy of mathematics is even more problematic.[10] Van Bendegem (2014)'s charitable suggestion is to keep an open mind and accept "to work in different 'registers' where reading texts in the field (or that I, at least, consider to be relevant)" (p. 221).

In a forthcoming paper, Carter (forthcoming) identifies three different, in some cases overlapping, "strands" in the philosophy of mathematical practice: (1) the *agent based* strand, (2) the *historical* strand, and (3) the *epistemological* strand. Differently from van Bendegem's proposal, all these strands mainly ask philosophical questions, but in a clear interdisciplinary fashion. A crucial point is that according to Carter the philosophy of mathematical practice has changed in the most recent years: in fact, it does not develop anymore in contrast with a more traditional philosophy of mathematics, but on the contrary is aimed to complement it.

The agent based strand emerges from the belief that philosophy has to take into account the human beings who are doing mathematics. In her reconstruction, this strand develops along two lines, the first having strong interconnection with sociology – mathematics is a *social* activity – and the second following the views of philosophers such as Peirce, Dewey and Putnam – the so-called *pragmatic orientation*. The historical strand focuses on the *products* of the activity of doing mathematics and on how such products shape across time. History of mathematics is of course an old discipline that has already provided interesting results; however, as Carter points out, the possible relationship between philosophy and history is still matter of discussion, ranging from positions considering history as philosophically laden to others defending the independence of history from philosophy. The third strand that Carter calls epistemological "for lack of a better term" is closer to traditional philosophy of mathematics. However, differently from it, it does not consider epistemology as a view from nowhere but demands that new topics emerging from everyday mathematics be considered as philosophically relevant. Other possible names for this strand could be the *extension-of-topics* strand, the *phenomenological* strand or the *philosophy of real mathematics* strand.

Van Bendegem claims that if the aim is to consider mathematical practice, then philosophy should collaborate with other disciplines; Carter seems instead to argue that philosophy itself should be open to change its nature in view of contributions coming from other disciplines. Moreover, for both of them, the listed approaches are not exclusive, as some scholars happen to endorse more than one. The picture of the philosophy of mathematical practice that results from these two proposals is evidently complex and in progress.

Instead of presenting a new possible catalog of the available views, I will introduce here briefly three categories of new questions about mathematics that have clearly emerged. As Soler and Jullien (2014) claim,

> *dynamic*, *genetic*, and *heuristic* aspects were largely ignored from the pre-practice turn philosophy of science (including mathematics), and the simple fact to take them as an object of study has often worked as a sufficient reason to classify an author as an actor of the practice turn (p. 232).

[9] Löwe (2016) explores the interplay between philosophy and other disciplines and its effect on the further development of the field.

[10] See the comment to van Bendegem's contribution by Soler and Jullien (2014).

I will follow their hint and focus on these three concerns in turn.

1. Dynamic aspects. One question formulated by Lakatos himself has not been answered yet: can we talk about 'progress' in mathematics? Does mathematics evolve? And if this is the case, are there constraints in its evolution? These issues were among the first to be considered by pioneering works going beyond traditional worries about mathematics. In particular, two collections focused on "revolutions" in mathematics – Gillies (Ed.), 1992 – and on the "growth" of mathematical knowledge – Grosholz & Breger (Eds.), 2000. However, in more recent years, the attention has moved away from questions about change in mathematics to new emerging topics in this category. One relevant subject is the consideration of *values* in mathematics that might influence the direction of research, for example when a certain framework is recognized as interesting, fruitful, providing explanations, or containing promises of solution. What do all these expressions mean? On this subject, it is not clear yet if and how sociological elements may be taken into account.

2. Genetic aspects. When it comes to the genetic aspects of mathematics, history of mathematics on the one hand and cognitive science on the other become relevant. Many authors have acknowledged the importance of an historical perspective to investigate mathematical practice. Corfield (2003) points out that in the course of the 20th century, philosophy moved its attention away from the real mathematical progresses because of the "foundationalist filter" that is the "unhappy" idea behind all forms of neo-logicism. In his view, the job of the philosopher is to dismantle such filter: mathematics is a human activity, and therefore it is situated in time. An interdisciplinary investigation may be of help, "in the process demonstrating that philosophers, historians and sociologists working on pre-1900 mathematics are contributing to our understanding of mathematical thought, rather than acting as chroniclers of proto-rigorous mathematics" (p. 8). The interest in looking at the research in cognitive science for a philosophy of mathematical practice is instead more controversial. Some scholars have explicitly discussed cognitive science research, for example Giaquinto (2006) and Ferreiros (2015). However, a shared intuition is that much work still need to be done to understand what exactly the views about the cognitive foundations of mathematics (Butterworth, 1999; Dehaene, 2007) or the role of conceptual metaphors and conceptual blending in mathematics (Lakoff & Nunez, 2001) might bring to the consideration in particular of the practice of *advanced* mathematics.[11]

3. Heuristic aspects. This category of problems goes back to typical themes from the work of Polya (1945) about the methods to put in place to solve mathematical problems. To give an example, one subject that has been very extensively addressed in the most recent years is the use of diagrams in mathematics.[12] However, thinking in terms of heuristics might be misleading, since heuristics pertains traditionally to the context of discovery but is not part of the context of justification, where proofs are "syntactic objects consisting only of sentences arranged in a finite and inspectable way"[13]. However, a crucial point about the consideration of the role of representations, notations and other kinds of cognitive tools is the evaluation of the influence that they might have on understanding and even on creating mathematics.

[11] Some of these issues are discussed in Schlimm (2013) and Giardino (2014).

[12] For a survey of the studies about diagrammatic reasoning in mathematics, see Giardino (2017).

[13] This passage is quoted from Tennant in Barwise and Etchemendy (1996, p. 3) as expressing the "dogma" of logocentricity that they want to challenge.

It is possible to argue that some proofs, because of their format, have both a heuristic and a justificatory role. For these reasons, the distinction between a context of discovery and a context of justification has become more and more precarious.

In the next section, I will discuss how this heterogeneity of topics for the philosophy of the mathematical practice is reflected in the heterogeneity of possible answers that are given to the question of what a mathematical practice is.

3. Not Just One (Mathematical) Practice

As for philosophy of science, a major difficulty for the philosophy of mathematical practice is how to intend 'mathematical practice'.[14] What do we talk about when we talk about a mathematical practice? Under which conditions is an agent recognized as a practitioner? Shall we talk about one practice or *more* practices? In other words, do several practices exist within one same practice?

On the website of the APMP, the philosophy of mathematical practice is defined as "a broad outward-looking approach" to the study of mathematics "which engages with mathematics in practice (including issues in history of mathematics, the applications of mathematics, cognitive science, etc.)". This definition is indeed far-reaching, most likely to the aim of being as inclusive as possible given the current state of the domain.

Famously, Kitcher (1984) defined a mathematical practice as the quintuple *<L, M, S, R, Q>* composed by *L*anguage, *M*etamathematical views, accepted *S*tatements, *R*easoning methods and *Q*uestions (chapters 7 and 8). Some authors thought of extending or reconsidering this quintuple. For example, Ferreiros (2015, chapter 3) has recently pointed out that Kitcher's quintuple is misleading because it is still based on an analysis of the production of scientific knowledge that depends mainly on linguistic knowledge. If the approach is instead intended to be agent-based and practice-oriented, non-linguistic elements, such as for example some forms of tacit knowledge, become relevant. For this reason, Ferreiros argues that it is necessary to think in terms of the couple Framework *plus* Agent, to whom the metamathematical views belong. Moreover, frameworks are of two kinds: *theoretical* and *symbolic*. However, the Framework-Agent pair is not identified with mathematical practice but is at the core of practice and of the production and reproduction of knowledge. As it is evident, the picture gets more and more complex.[15]

Of course, scholars who are interested in the philosophy of mathematical practice seem to share some notion of practice, but this happens on the surface, while the devil is in the details. In fact, the variety of philosophies of mathematical practice described in the previous section is indeed reflected in the variety of possible views about mathematical practice. For this reason, I propose here to identify (at least) four replies that philosophers of the mathematical practice might give to the question about mathematical practice.

I will call the first reply (1) the *situated* reply. Mathematical practice is a historically situated human activity, and therefore the aim of the philosophy of mathematical practice is to reconstruct the history of the different practices. If mathematics is truly what mathematicians do, then this target has considerably varied over times and places; as a consequence, mathematical practice must be understood in a way that would include this variance. A static view of mathematics considering it as merely a collection of theories,

[14] To be true, also the very definition of a practice *in general* is problematic, and therefore it is not clear how to intend the claim that mathematics is analogous to *other* human practices.

[15] Another extension of Kitcher's model was proposed by van Kerkhove & van Bendegem (2004), who generalized it and arrived at a seventuple *<M, P, F, PM, C, AM, PS>*, containing a mathematical *community M* of individual mathematicians, a *research program P*, a *formal language F*, a set *PM* of *proof methods*, a set *C* of *concepts*, a set *AM* of *argumentative methods* and a set *PS* of *proof strategies*. Also in this case, the model gets more convoluted.

independent of human activities, is misleading, and it is necessary to move towards a *dynamic* view of mathematics. Of course, Lakatos' lesson is always in the background. The target of the philosophical investigation is the subject matter of mathematics at each time: mathematical theories, theorems, and proofs. Of course, these objects may be presented in different formats and media, and there is an interest in considering their development.

The second reply is (2) the *semiotic* reply. Mathematical practice is a human activity implying the use of many different tools, more importantly several kinds of texts, which are the target of the philosophical research. Mathematicians in their everyday work write drafts, inscriptions, publications; they draw objects, calculate on paper, and write demonstrations. These are all elements of the practice of mathematics, and they can be analyzed one by one without leaving aside their mutual relations. The mathematical practice has then to do with the *traces* of mathematics that are left in sketchbooks, textbooks, essays, and proofs.

The third reply is (3) the *epistemological* reply. Mathematical practice is the construction of theories, but this does not imply endorsing more dogmatic points of view, for example the claim that such theories are necessarily formal systems. As Mancosu (2008) claims,

> the epistemology of mathematics needs to be extended well beyond its present confines to address epistemological issues having to do with fruitfulness, evidence, visualization, diagrammatic reasoning, understanding, explanation and other aspects of mathematical epistemology which are orthogonal to the problem of access to 'abstract objects' (pp. 1-2).

Despite the fact that case studies are necessary precisely because certain areas of mathematics can provide useful tools for addressing important philosophical problems, such approach is not meant to be simply a description of the mathematical theories and of their growth.[16]

The fourth reply is (4) the *pragmatist* reply. For example, for Ferreiros (2015), mathematical practice is what the community of mathematicians does when they employ resources such as frameworks and other tools to the aim of solving problems, proving theorems, and in some cases elaborating new theories and frameworks. Moreover, the choice of these tools is constrained by their cognitive abilities. The study of mathematical practice broadens the scope of philosophical and historical studies by considering carefully the contexts from which mathematical theories and proofs emerge, and issues such as understanding beyond mere logical reconstruction become crucial. Ferreiros' approach is agent-based: practice can be understood only by focusing on practitioners; moreover, it is pragmatist and historically oriented.

As for the approaches in the previous section, these groups of replies are of course not disconnected from each other. For example, the semiotic reply has clearly epistemological interests as well;[17] or, for both the situated and the epistemological reply, the construction of theories and the notion of proof are still two crucial issues in the investigation of mathematical practice.

[16] I will come back to this issue in the last section.

[17] For example Chemla (2009), discussing the importance of learning to read mathematical texts, explicitly refers to the notion of "epistemological cultures" as introduced by Fox Keller (2002) to define the primordial character of the specific epistemological choices that are made and shared by the agents of cultures that are far from ours.

4. Conclusions: A Pluralism of Approaches

The emerging picture of the philosophy of mathematical practice seems thus to describe a domain of research, mathematics, which has lost its unity because it is now partitioned in many different case studies. More importantly, there is neither unity in the philosophy of mathematical practice itself, since each of the distinct approaches may endorse different points of view on each of these case studies. With this worry in mind, in this last section I will discuss some open problems.

First, the relation between the philosophy of mathematical practice and traditional philosophy of mathematics is not easy to evaluate. Many scholars believe that there is a true need for an extension of the theoretical inquiry that would address topics ignored by the foundationalist tradition (because of what Corfield called the "foundationalist filter"). As Mancosu (2008) highlights, the philosophical literature has extensively pursued the Benecerraf's ontological and epistemological problems (are there abstract objects? and if there are, how can we access them?) and without this extension, it risks being drastically impoverished. However, this does not mean that the work in traditional philosophy of mathematics has to be forgotten or considered as irrelevant. On the contrary, the tools that it has provided can be extended as well to new areas of research that have been previously largely neglected. As he sums up, philosophers today are less ambitious and at the same time more ambitious than before. They are less ambitious because differently from scholars such as Lakatos or Kitcher, they are not concerned anymore with metaphilosophical issues; however, they are more ambitious because they want to cover "a broad spectrum of case studies arising from mathematical practice" that are subject to analytic investigation (p. 14). In a motto: less metaphysical questions, more topics addressed.

Second, it is not clear what the purpose of philosophy should be in considering the practice of mathematics. Some approaches aim at maintaining a normative role for philosophy while others consider that the research in philosophy should provide an attentive description of the situated practice. This might create some tension. In fact, a potential risk is that too much focus on practice will end up dispelling philosophy. As Maddy (1997) already argued in *Naturalism in Mathematics*:

> if our philosophical account of mathematics comes into conflict with successful mathematical practice, it is the philosophy that must give. [...] Similar sentiments appear in the writings of many philosophers of mathematics who hold that the goal of philosophy of mathematics is to account for mathematics as it is practiced, not to recommend reform (p. 161).

However, many philosophers of the mathematical practice today would not subscribe to Maddy's naturalistic claim (in her specified meaning of this term).

Third, another issue not settled yet is the autonomy of mathematics from the natural sciences. In his book, Ferreiros (2015) emphasizes the interplay between mathematics and other kinds of practices. In his view, the problem of the "applicability" of mathematics should not be considered as external to mathematical knowledge but on the contrary as *internal* to its analysis. There is no opposition between "pure" and "applied" mathematics, since to some extent all frameworks are designed to be applicable.

Fourth, for many of the different views that have been described so far, practice has to do with some form of *action*. However, an appropriate analysis of this feature in the practice is still lacking.[18] In this spirit, a promising direction of research will be to explore the view of mathematical knowledge as a *knowing-how*, as practical and/or tacit knowledge, in contrast

[18] One attempt in this direction is made in Salanskis (2014).

with or more modestly in addition to the standard view of mathematical knowledge as a knowing-that.

To quote again van Bandegem (2014),

> if all of this looks rather sketchy, it is important to realize, [...] that we are looking, in comparison with developments in the philosophy of science, at a *very young discipline*. Nevertheless, I do think it is important, right from the start, to look for collaborations and not exclusions. Exclusions are only to be accepted when everything else fails, and this is definitely not the case at the present moment (p. 224, emphasis added).

Moreover, I would argue that this lack of unity in the philosophy of mathematical practice is to some extent what is really revolutionary about it: after the practical turn, the territory of philosophical inquiry has radically changed. Enterprises such as the identification of criteria of validity for what counts as a mathematical proof have become local enterprises, which may vary in their methodology and in their results depending on the particular practice and on the particular case study that are taken at each time into account. It would then seem that mathematics, which was considered as a stable, static, certain, exact science not subject to change or development, has finally exploded into pieces and it will be impossible for philosophy ever again to provide a unitary account for it. Alternatively, an improved philosophy of mathematics will consider this as an occasion to specify new questions and take really into account the actual richness of its domain of interest in all its complexity. Will the philosophy of mathematical practice ever become an adult discipline? More time is needed to reply to this question.

REFERENCES

Agazzi, E. & Heinzmann, G. (Eds.) (2015). *The practical turn in philosophy of science*. Milan: Franco Angeli.

Aspray, W. & Kitcher, P. (Eds.) (1988). *History and Philosophy of Modern Mathematics.* Minneapolis: University of Minnesota Press.

Barwise, J. & Etchemendy, J. (1996), Visual Information and Valid Reasoning. In G. Allwein & J. Barwise (Eds.) (1996). *Logical Reasoning with Diagrams*. Oxford: Oxford University Press, 3-25.

Butterworth, B. (1999). *The Mathematical Brain*. London: Macmillan.

Carter, J. (forthcoming). What is Philosophy of Mathematical Practice - motivation, themes and prospects. *Philosophia Mathematica.*

Chemla, K. (2009). Apprendre à lire: La démonstration comme élément de pratique mathématique. *Communications*, 84, 85-101.

Corfield, D. (2003). *Towards a Philosophy of Real Mathematics*. Cambridge: Cambridge University Press.

Dehaene, S. (1997). *The Number Sense*. Oxford: Oxford University Press.

Ferreiros, J. (2015). *Mathematical Knowledge and the Interplay of Practices.* Princeton: Princeton University Press.

Fox Keller, E. (2002). *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines*. Cambridge, MA: Harvard University Press.

Giardino, V. (2017). Diagrammatic reasoning in mathematics. In L. Magnani & T. Bertolotti (Eds.), *Spinger Handbook of Model-Based Science*. Dordrecht; Heidelberg; London; New York: Springer, 499-522.

Giardino, V. (2014). Matematica e cognizione. In A. C. Varzi & C. Fontanari (Eds.) *Matematica e filosofia*, Special issue of *La matematica nella società e nella cultura. Rivista della Unione Matematica Italiana*, VII(3), 397-415.

Gillies, D. (Ed.) (1992). *Revolutions in Mathematics*. Oxford: Clarendon Press.

Grosholz, E. & Breger, H. (Eds.) (2000), *The Growth of Mathematical Knowledge.* Dordrecht; Boston; London: Kluwer Academic Publishers.
Hersh, R. (Ed.) (2005). *18 Unconventional Essays on the Nature of Mathematics*. Berlin: Springer-Verlag.
Giaquinto, M. (2007). *Visual Thinking in Mathematics*. Oxford: Oxford University Press.
Jullien, C. & Soler, L. (2014). Commentary to "The Impact of the Philosophy of Mathematical Practice to the Philosophy of Mathematics" by Jean Paul van Bendegem. In Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (Eds.), Science after the Practice Turn in the Philosophy, History, and Social Studies of Science. New York-London: Routledge, 227-237.
Kitcher, P. (1984). *The Nature of Mathematical Knowledge*. Oxford: Oxford University Press.
Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago-London: The University of Chicago Press.
Lakatos, I. (1976). *Proof and Refutations: the Logic of Mathematical Discovery*. Cambridge: Cambridge University Press.
Lakoff, G. & Nunez, R. (2001). *Where mathematics comes from: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books.
Löwe, B. (2016). Philosophy or not? The study of cultures and practices of mathematics. In Ju, S., Löwe, B., Müller, T. & Xie Y. (Eds.). *Cultures of Mathematics and Logic*. Basel: Birkhäuser, 23-42.
Maddy, P. (1997). *Naturalism in Mathematics*. Oxford: Oxford University Press.
Panza, M. & Sereni, A. (2013). *Plato's Problem. An Introduction to Mathematical Platonism*. Palgrave Macmillan.
Polya, G. (1945). *How to solve it*. Princeton: Princeton University Press.
Salanskis, J. M. (2014). Some notions of action. In Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (Eds.), Science after the Practice Turn in the Philosophy, History, and Social Studies of Science. New York-London: Routledge, 44-57.
Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (Eds.) (2014). *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science*. New York-London: Routledge.
Tymoczko, T. (1998). *New Directions in the Philosophy of Mathematics*. Revised and extended version. Princeton: Princeton University Press.
Schlimm, D. (2013). Mathematical practice and conceptual metaphors: On cognitive studies of historical developments in mathematics. *Topics in Cognitive Science*, 5, 283-298.
van Bendegem, J. P. (1993). Real-life Mathematics versus Ideal Mathematics: the Ugly Truth. In E.C.W. Krabbe, R.J. Dalitz, & P.A. Smit (Eds.), *Empirical Logic and Public Debate, Essays in Honour of Else M. Barth*. Lanham: Rowman & Littlefield, 263-272.
van Bendegem, J.P. (2014). The Impact of the Philosophy of Mathematical Practice to the Philosophy of Mathematics. In Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (Eds.), Science after the Practice Turn in the Philosophy, History, and Social Studies of Science. New York-London: Routledge, 215-226.
van Kerkhove, B., & van Bendegem, J.P. (2004). The unreasonable richness of mathematics. *Journal of Cognition and Culture*, 4(3), 525-549.

BEATRICE MAGNI
University of Milan
beatrice.magni@unimi.it

NEW WINE IN OLD BOTTLES: THE KIND OF POLITICAL PHILOSOPHY WE NEED

abstract

There isn't an overall consensus on the aim, meaning and role(s) of contemporary political philosophy. The relationship between philosophy and politics has been addressed and sharpened – not just today but in different ways and from various, separate and sometimes conflicting perspectives (Leopold & Stears, 2008). Regardless, the main aims, meaning and role of a field of study are key issues, and the quality and credibility of the research will most likely depend on our capacity to draw a path through this conflicting background. The purpose of this paper is to contribute to drafting elements of a new road map that could lead contemporary political philosophy out of this crippling impasse. It builds on a specific version of political theory – Walzer's interpretation path reviewed (Walzer, 1985) – and addresses a kind of political practice able to reconcile political philosophy's normative commitments – as is the case with the Rawls' four roles of political philosophy (Rawls, 2007) – with its actual ambitions and conditions of achievability (Hall, 2015; Galston, 2010).

keywords

political philosophy, normativity, interpretation, judgement, injustice

Phenomenology and Mind, n. 12 - 2017, pp. 30-39
DOI: 10.13128/Phe_Mi-21103
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

NEW WINE IN OLD BOTTLES: THE KIND OF POLITICAL PHILOSOPHY WE NEED

Introductory Remarks

In her work on the differences between political science, political theory, and politics, R. Grant (2002) identifies – starting from Berlin's seemingly critical estimation of the scientific project of political philosophy[1] – what she calls the "practical and theoretical problem" (p. 578) inherent in the humanities: political theory would never become a science because of the character of the concerns it addresses: normative concerns, which indicate how political agents and political institutions *should* act in the domain of politics. When one makes a normative claim, one expresses an evaluation of something; when one evaluates something, it is assessed relative to some standards, ideals or possible alternatives. In other words, something is, in some respect, better, worse or on a par with some standard, ideal, or alternative. Normative questions and concerns contain an element of evaluation and ultimately remain – in political theory in general – obstinately philosophical, and consequently, their claims cannot be either validated or falsified definitively through any scientific method. Grant sees three possible answers to this sort of characterisation: the first is simply to accept it because the main aim in the humanities should be not so much about acquiring scientific knowledge but to provide a type of educational experience that can be inspirational, revelatory, and transformative of our common world. The second response posits that some elements of uncertainty are inevitable, even in the most formal sciences; therefore, the distance between the so-called "hard" sciences and the "social" sciences is smaller than its followers (on both sides) are willing to admit. Both lines of argument, Grant continues, have some merit, but they are not sufficient to define the character of political theory and its importance for the study of politics. The third possibility, then, is to acknowledge that humanities research requires a special defence, a defence on its own terms:

[1] "Nevertheless, attempts made by the *philosophes* of the eighteenth century to turn philosophy, and particularly moral and political philosophy, into an empirical science, into individual and social psychology did not succeed. They failed over politics because our political notions are part of our conception of what is to be human, and this is not solely a question of fact, as facts are conceived by the natural sciences; nor the product of conscious reflection upon the specific discoveries of anthropology or sociology or psychology, although all these are relevant and indeed indispensable to an adequate notion of the nature of man in general, or of particular groups of men in particular circumstances. Our conscious idea of man – of how men differ from other entities, of what is human and what is not human or inhuman – involves the use of some among the basic categories in terms of which we perceive and order and interpret data. To analyze the concept of man is to recognize these categories for what they are. To do this is to realize that they are categories, that is, that they are not themselves subjects for scientific hypothesis about the data which they order" (Berlin, 1999, pp. 162-163).

the distinctiveness of humanities research, to which political philosophy belongs, has its own particular characteristics and should be defended as integrally related to the aims and the limits of humanistic inquiry, i.e., the means of interpretation and judgement (or, following Grant's vocabulary, the historical understanding):

> There is nothing arbitrary about the methodological approach [of political theory, e.d.]. You cannot discover either what something means or why it matters without both interpretation and historical understanding. The characteristic uncertainty, disagreement, and lack of closure found in the discourse of humanities are not arbitrary either. These characteristics reflect both historical and epistemological realities [...]. I would suggest that, whereas the sciences are primarily concerned with knowledge of cause and effect, the humanities are primarily concerned with understanding of meaning and judgment of significance (Grant, 2002, pp. 581-582).

Along the lines of this argument on the political theory's stance, my aim in this essay is to clarify the extent to which reflecting on the relationship between philosophy and politics enables us to highlight the unique character of political philosophy. In the first part, I will attempt to isolate two main concerns surrounding political philosophy and its issues – descriptive and normative concerns – that are covered by three different and sometimes conflicting levels of analysis: epistemic, moral, and political. In the second part, I will consider the main lines of one of the most compelling efforts to gather concerns and analytical levels: Walzer's attempt to find a *connected* criticism and to identify what (political philosophers as) social critics do and how they go about doing it. Linked to that attempt, my provisional conclusion attempts to suggest that the reasons and arguments one can use to blame political philosophy are the same that make its unfinished work so necessary today.

The Old Questions (and Socrates' Cold Case)

To depict the directions in which political philosophy is heading, the first concern may even appear to be merely a matter of definition: what is political philosophy, and why does it matter?
It is difficult to answer even this question univocally. In one sense, one could say that political philosophy is simply a branch, or what we call a subfield, of the field of political science. It exists alongside of other areas of political inquiry such as policy studies, comparative politics, and international relations. In another sense, political philosophy is something much more different than simply a subfield; it appears to be the oldest and most fundamental element of political theory. Its purpose is to address, as it were, the fundamental problems, concepts and categories that frame and identify the study of politics. In terms of content, political philosophy is primarily concerned with questions of freedom, equality, justice and political authority. Matters of political authority concern why and to what extent political authority has legitimate power over individuals and groups. Do governments derive their authority from the consent of the governed? If so, what does that consent look like? Can the state do anything it wants to the governed, or are there limits? If there are limits, where do those limits come from? In this general sense, we can state that political philosophy investigates whether, on what grounds, and to what extent politics and power, or political authority, can be justified. Political philosophy, in this regard, will focus on the examination of a series of basic and central questions:

- What is the nature of justice, freedom, and equality?
- What is the justification for the authority of a state?
- How should we envision the relationship between ethics and politics?
- What is a just society?

- What constitutes a good citizen?
- What is the relationship between order, authority and freedom?

These are a few such questions. Political philosophy can explore these questions, for example, through the careful study of classic and contemporary texts in the field and will take the form of a broad inquiry of some of these most fundamental topics. Classic philosophical works accordingly provide us with the most basic questions that continue to guide the field. We keep asking the same questions that were asked by Plato, Machiavelli, John Locke, and others. It can be argued that we do not accept their answers, and it is likely that we ultimately do not, but their questions are often posed with a type of unrivalled clarity and insight, and their doctrines have not simply been refuted, replaced, or historically superseded; they remain, in many ways, constitutive of our most basic perspectives and attitudes about the world. However, when these old and classic questions – as Rawls specifies in his *Lectures on the History of Political Philosophy* (Rawls, 2007) – are raised in different historical contexts, they can be taken in different ways and have been approached by different scholars from different points of view according to their political and social worlds, their circumstances and problems as they saw them. It is the *fact* of pluralism that implies that, regardless how impartial and altruistic people are, they still disagree in their factual judgements and in religious, philosophical and moral doctrines (Freeman, 2014). To understand their works, then, we must identify these points of view and how they shape the way the writer's questions are *interpreted* and *discussed*. If we go one step further, engaging in political philosophy will therefore mean answering questions to which we often do not have safe and sure answers, and we can say that political philosophy works as a critical approach in terms of being:

- a commitment to make distinctions between states of the world;
- a commitment to identify criteria for evaluating states of the world;
- a commitment to order the possible states of the world according to some preferred principles.

This critical effort drives political philosophy from a first descriptive level to another one: political philosophy becomes foremost, then, a normative discipline[2] – that is, one concerned less with questions about how political life is or was and more with how it should be. The primary aim of political philosophy becomes helping those who address it to think more deeply about important theoretical concepts and crucial political problems.

All this is in accordance with the Socratic method. In the *Apology,* for example, Plato has Socrates explaining and justifying himself, his way of life and thinking before a jury of his peers: Socrates speaks in a public forum when defending the utility of philosophy for political life. At the same time, the *Apology* demonstrates the vulnerability of political philosophy in relation to the city and political power. From its beginnings, philosophy and the city, as well as philosophy and political life, have existed in a sort of tension with one another. Socrates is charged, as we know, by the city for corrupting the youth and for impiety towards the gods (in short, treason), and the *Apology* puts not merely an individual but, we might say, the idea of political philosophy on trial. For the philosopher – as in the case of Socrates and Delphi – it is not enough simply to hold a belief on faith but to be able to give a rational and reasoned account for one's belief: its goal, again, is to replace civic faith with rational knowledge.

[2] Even if boundaries are not always clear between descriptive and normative, i.e., between description and prediction. See Sen (1980).

Therefore, philosophy is necessarily at odds with belief and this kind of civic faith. The citizen may accept certain beliefs about faith because he or she is attached to a particular kind of political order, regime or ideology. However, for the philosopher, this is never enough. The philosopher seeks to judge these beliefs by true standards, i.e., what is always and everywhere true as a quest for knowledge.

Thus, there is a necessary and inevitable tension between philosophy and belief, or to put it another way, between philosophy and the civic compromises that hold the city together. However, even though one might say that Socrates appears to be engaged in a sort of highly personal quest for self-perfection (he maintains throughout the entire trial that the unexamined life is not worth living), there is also something deeply *political* about the *Apology* and his teachings that one cannot avoid. At the heart of the dialogue and this speech is a dispute with his accusers over the question, which is never stated directly, of who has the right to guide the future citizens and statesmen of the city of Athens. Socrates' defense speech, like every Platonic dialogue, is ultimately a dialogue about education: who has the right to teach and who has the right to educate the city? This is in many ways the fundamental political and philosophical question of all time for Socrates. This is essentially a question of who governs or, said otherwise, who should manage, i.e., who should manage disagreements that represent the main feature of political life. Socrates intends to put the democracy of Athens itself on trial: not only does the *Apology* force Socrates to defend himself before the city of Athens, but Socrates, with his strong critique of democratic practices, puts the city of Athens on trial and makes it defend itself before the high court of (his political) philosophy.

Keep Your Position

Thus, if we decide to enter the debate on some of the most basic and fundamental merits and limits of the study of today's political philosophy through the *Apology*, the shifty Plato's reference to Achilles (Plato, 1991, 28c)[3] could probably be significantly more revealing than the most famous *gadflying* (Colaiaco, 2001). The latter case is the reference by which the philosopher explains his benefaction to the *polis* as analogous to the good done by a gadfly to "a large and well-bred horse, a horse grown sluggish because of its size and in need of being roused" (Plato, *Ap.* 30e-31a). With the Achilles example, Socrates maintains, as he states near the end of the defense speech, that the examined life is alone worth living and only those engaged in the continual struggle to clarify their thinking and remove sources of contradiction and incoherence can be said to live worthwhile lives. The Socratic paradigm of political philosophy may reveal some features in common with the older Homeric warrior: Socrates and Achilles are paradigms of the tradition – philosophical and heroic in the order – and are two connected critical voices within the tradition. In a significant and scarcely examined passage, Plato re-reads Homer when Socrates invokes Achilles as an exemplar of the courage he himself must display in pursuing his mission: ultimately, he wants to replace military combat with a new type of epistemic fight, in which the person with the best argument – the best justified argument – is declared victorious. The principle is for the best argument to prevail while maintaining one's position – as Achilles did to protect his friends and comrades – to show who one was (at descriptive level), who one is (at descriptive level), and who one should be (at normative level). Here, Achille's *aretè* – specifically, the *soldier*'s virtue and courage – becomes the most peculiar character of the Socratic methodology and philosophy. The *Apology* then shows Socrates offering a new model of citizenship and a new kind of citizen. As was the case

[3] "This is the way it is, men of Athens, in truth. Wherever someone stations himself, holding that it is best, or wherever he is stationed by a ruler, there he must stay and run the risk, as it seems to me, and not take into account death or anything else compared to what is shameful" (Plato, *Ap.* 28c).

with Achilles, so is with Socrates: fear of death, or any other punishment, will never induce a philosopher to abandon her stance.

Now and Around Here: Political Philosophy and the Present

My position concerning where political philosophy must begin and what it ought to take into account when so doing therefore considers two distinct and equally much-needed commitments underlying the meaning and the aims of political philosophy: as we know from the Socratic cold case and the constitutive ambiguity of political philosophy – the Socratic rational inquiry that even calls the Oracle into question – the analytical ambition must always come to terms with the contingency constraint. We saw that we can engage in political philosophy in a descriptive and normative way. Political philosophy should therefore descriptively remain at a certain distance from political events and contingencies to normatively develop appropriate criteria and categories that can make some specific difference in politics. The problem, though, is that by placing itself at a certain distance from its object in relation to politics, political philosophy may, of course, respond adequately to its philosophical commitments and meet the demands of theoretical rigor but appears to be less able to honor another commitment, which is also indispensable. Political philosophy appears to be less committed to putting categories and criteria into place that make a difference and can be useful in politics.

What I pose here is, therefore, not only a problem of distance, the recurrent trouble of the correct distance at which the philosophy must be placed to fruitfully examine politics and its contingencies. The problem I am raising concerns *how* political philosophy, once the gap has been exploited with respect to policy and contingency, can claim its usefulness, how it can make categories and criteria that it processes relevant while staying *away* from contingencies relevant to politics. Thus, the problem I raise appears to be an epistemic problem and concerns how political theory matters today, i.e., the way in which political philosophy should justify itself by conceiving or reconstructing the link between its principles that are theoretically elaborated, on the one hand, and the politics and its contingencies on the other. The tension between philosophy and politics is reformulated in a question of guidance: where must one go to engage in political philosophy? I suggest it is at this point where the political philosopher is faced with a choice.

From Rawls' Epistemic Account to Walzer's Connected Criticism

In the Socratic turn, philosophy is a kind of "public service" that constantly demands dialogue, which is never a mere theoretical exercise but always a mutual crossing in the context of the political exchange that compels the interlocutor to become involved and to be a moral agent who has to always give (good) reasons for her positions. In the epistemological account of Rawls (Rawls, 2007), political philosophy is tasked – perhaps more modestly – with explaining how we know and apply political philosophy's principles and categories.

According to Rawls, we can distinguish four roles of political philosophy as part of the public culture of every society: practical, guiding, reconciling, and realistically utopian. The practical one – the first – aims to find a common and rational ground for political dealings in political conflict and disagreement. It focuses on some controversial issues, and – against all odds – considers whether it is possible to find some basis for a philosophical and moral agreement or at least whether it is possible to limit the existing political divisions to save social cooperation based on mutual respect. The idea behind the second role is that reason and thinking (both theoretical and practical) should orient individuals and institutions in the conceptual space of every possible end. Political philosophy can further help us to reconcile with our comprehensive views (see also Rawls, 2005, pp. 10, 40, 144), showing us the reason of the fact of pluralism, its benefits, and some political advantages; finally, political philosophy can be realistically utopian, i.e., it can attempt to create a decent political order and a reasonably just democratic regime.

To perform such tasks, political philosophy needs statements of value because of its prescriptive or normative attitude. The premise is that political life and institutions are not regarded as unchanging and part of the natural order but as potentially open to change and therefore as recurrent stances in need of philosophical justification. However, political philosophy for Rawls is relevant, especially in times of crisis, where it becomes imperative to find and implement some new shared criteria of judgement. It then calls for critical clarification of and reflection on the most fundamental terms of our political life and suggests new possibilities for the future. Political philosophy exists and only exists in that "zone of indeterminacy" between the "is" and the "ought", between the actual and the ideal,[4] which is why political philosophy is always and necessarily a potentially disturbing endeavour. What is distinctive is its prescriptive or evaluative concern – in short, its concern with how political societies should be, how policies and institutions can be justified, and how we and our political leaders ought to behave in our public lives. This tension between the best and the actual is the only way in which a Rawlsian perspective makes political philosophy possible: in an ideal situation, political philosophy would be unnecessary or redundant; it would wither away. At the same time, however, the actual cannot prevent philosophy from assessing truths, answering questions, and settling disagreements.

In the wake of Rawls and his four roles of political philosophy, one could be led to believe that the difficulty political philosophy faces in expressing its object – politics – comes from its tendency to represent it through categories that hide or remove its prevalent content, i.e., conflict (of interest, power, and values). To be clear, such a difficulty is not only connected to the choice of certain authors; it is rather inherently connected to the functioning and status of the Platonic model of political philosophy, where political philosophy is structurally unable to consider conflict because it is originally oriented towards the question of order. However, if we consider that conflict is not a slag to be eliminated but is rather the irreducible core, basis and substance of politics, we should admit that no attempt at giving shape or order to politics can dismiss it unless it is possible to completely revamp the political itself. To help frame this issue, it is useful to reconsider Michael Walzer's proposal of a *connected criticism* in political theory. Walzer, like Rawls, sees no way in which the pluralism in politics might be avoided and no definitive way to ending the disagreement. However, his reflection on the possible positioning of political philosophy starts with this question: where do we have to start criticizing? He explicitly refers to the Platonic allegory: the philosophy has to dwell in a cave. However, he cannot maintain or claim an external or superior position: "We have to start from where we are: I do not mean to deny the reality of the experience of stepping back, though I doubt that we can ever step back all the way to nowhere; even when we look at the world from *somewhere else*, we are still looking at the world. We are looking, in fact, at a particular world; we may see it with special clarity, but we will not discover anything that is not already there" (Walzer, 1987, p. 16).[5] This approach suggests that people critically examine their own practices, or better, it wants to chronicle and extend patterns of critical arguments that already exist. Walzer's social criticism, in this respect, requires critical distance, but this new kind of criticism "does not require us to step back from society as a whole but only to step away from certain sorts of power relationships within society. It is not connection but authority and domination from which we must distance ourselves" (Walzer, 1987, p. 52).

[4] I owe Salvatore Veca this idea of *actuality.*

[5] On that point, see also Galston (2010, p. 396): "we must begin from where a given political community is" and Hall (2015, p. 6): "The basic thought, then, is that we cannot clarify the nature of various political values in any meaningful manner before we consider the historical and political question of what their elaboration requires 'now and around here' [...]. I will refer to this idea as the 'realism constraint'".

Walzer conceptualizes the activity of social criticism in a way that puts special emphasis on the connection of the critic with the society in which she operates. In *Interpretation and Social Criticism* (1987), Walzer distinguishes between three paths in moral philosophy: discovery, invention and interpretation. The first path, discovery, is one where the moral philosopher receives her ideas from the outside the communities at which these ideas are directed. The classical example of this path is the revelation of the commandments to Moses. The second path, invention, starts from the assumption that the rules of human interaction cannot simply be discovered – they need to be constructed to guide collectives. The difference in the path of discovery is that the moral philosopher makes use of some hypothetical device or thought experiment so as to generate principles of justice. Here, Walzer is clearly referencing Rawls: the underlying intuition is that no external creator is needed in order to produce rules of human interaction; instead, the application of the appropriate method alone will lead to the right kind of results. However – and this is a crucial facet of many such hypothetical devices and thought experiments – the individuality of those who contribute to the constructive process becomes effaced. Individual standpoints disappear in the course of inventing principles of justice. Such is the purpose of impartial procedures. The path of invention thus accentuates human agency but only to the extent to which it can help generate principles of justice that abstract from the individuality of those involved. In the second path, the end is given with the morality we hope to invent. Walzer identifies a third path of moral philosophy that breaks with both discovery and invention: if discovery and invention are efforts at escape in the hope of finding some external and universal standards with which to judge moral and political existence, the effort may well be commendable but unnecessary; this is the path of interpretation. Moral philosophy as an interpretation conceives of the activity of social criticism as embedded in and dependent on society. In the third path, we do not have to discover the moral world because we have always lived there. We do not have to invent it because it has already been invented. A moral argument in such a setting is interpretive in character, closely resembling the work of a lawyer or judge who struggles to find meaning in a combination of conflicting laws and precedents. This emphasis on the connectedness of social criticism naturally invites doubt about distance: how far should the critic be from society if absolute detachment is, in fact, detrimental to her activity? If shared understanding of what is valuable in a society is a precondition for effective social criticism, how much commitment to these communal values is necessary? Social criticism is an immanent activity and is typically considered to be the practice of one who can be detached enough to examine a particular society from a vantage point that is “no place in particular” (pp. 5, 16).

Walzer is correct when he posits that a moral and political world already exists, as a historical product, that gives structure to our lives but whose ordinances are always uncertain and in need of scrutiny, argument, and commentary. This perspective turns out to be particularly useful to proposing a theoretical proposal to fitting “hard and dark times” in politics, to thinking about and evaluating answers, solutions, and to finding a way out.

The New Questions (Political Philosophy and the Problem of Injustice)

The most important issues of political philosophy so far address the controversial question of what justice requires. If now we attempt to take a step in the direction suggested by Walzer, at a more normative and political level, we will probably find a new philosophical black list, where the stakes are as follows:

- What is injustice?
- What are the goals of a decent society?
- What constitutes the basis of human dignity?
- What does this imply for our obligations as human beings and citizens?

- What relationships should we establish among our passions, our subjectivities, our main interests, and the rules of public life?
- How much inequality can we live with?

If one looks at our many injustices, what becomes clear is how ordinary and pervasive they are. They do not involve only acts of obvious misconduct but also failures of both governments and citizens to act when they could. The political philosopher as a connected critic is not separate from or outside the society that he or she interrogates and challenges but is rather "connected" to it, engaged in its central concerns and passionately, if complicatedly, involved in the struggles of the common people. What does this mean for the meaning and the role(s) of political philosophy? Can political theory meet the challenges of the present? We now know that we need a political philosophy that engages political science without attempting to become a science. The best contribution that political philosophy can make to the study of its main issues – i.e., political issues, as injustice – depends on its loyalty and commitment to philosophical questions as they arise in our political and everyday life. Political philosophy as it stands is an imminently practical discipline and field, where the purpose is not simply contemplation or reflection alone: it is advice giving. The fact is that the work of political philosophy is irreducibly plural and multidimensional, and although we are most familiar with the character of a modern democratic regime such as ours, a consistent and distinctive conception of political philosophy is, in many ways, a type of immersion into what we might today call comparative politics. Regarding this attempt to find a road map to political philosophy, it is not justice that brings us to politics but injustice – the avoidance of evil rather than the pursuit of good. Heading off evil, not the attempt to realize that an ideal condition of justice and fairness, should be the central focus of political thought and action. It is also important to realize that philosophy is not without a history; philosophy is a historical movement that tackles social and political questions as well as more technical problems of logic and epistemology.

The Vulnerability and the Usefulness of Political Philosophy. Conclusive Remarks

In this brief essay, I did not want so much to propose a theory as to explore and expose difficulties in the ways we characteristically think and act when we currently discuss political theory in general and political philosophy in particular. I believe that Isaiah Berlin was right: political theory will never be a science due to the presence of pluralism and disagreement. However, its main weakness could coincide with its primary constructive power, and if we refer to Grant's argument, we might perhaps agree that the fact of disagreement does not imply that nothing can be known, only that everything cannot be "between ignorance and knowledge, in the realm of judgment, [is] where the humanities reside" (Grant, 2002, p. 585).[6] Therefore, thinking realistically about the audience, the authority, and the position of political philosophy could mean attempting to present it as a viable and fruitful method for interpreting the political events of our time without removing its philosophical commitments. If Nagel is right (Nagel, 2005), the path of justice is a consequence of correctly finding

[6] See also Grant (2002, pp. 589-590): "Every good causal explanation of political phenomena cannot exclude the questions of interpretation and judgment that drive political philosophy. Political theory as an enterprise assumes that interpretations, conceptual regimes, judgments of significance, and ideas of all kinds are themselves both causes and effects. Ideas have significant consequences [...]. In other words, the study of politics needs both to seek general laws to explain the causes of political behavior and to develop interpretations of the meaning and significance of political events and conceptual regimes to inform evaluative judgments of them. Political studies have both scientific and humanistic aims. These are distinct but complementary enterprises; the 'permeability' does not efface the distinction".

injustices. The normative constraint of political philosophy involves questions of value, what we should do, or what we ought to do when we face a political dilemma. To be concerned with finding reasons and justifications to eliminate or reduce injustice are still normative concerns. If Walzer is right, we need distinguish the epistemic problem of knowledge and how we come to know moral distinctions from the problem of motivation and what moves us to act based on moral distinctions. In this sense, political philosophy should take on the responsibility over the long term to understand politics and meet the contingencies, ask the right questions, find possible and reasonable answers, and contribute to reducing injustices. The "possible-accessible" is what we call for in political philosophy today: the priority of the *actual* over the *possible*. The priority of actuality is the only path to any form of possibility in political theory. Rooms can be rearranged, as Walzer suggested, and old bottles can be refreshed with new wine. Between desirability and feasibility, the purpose could only be to provide some elements for a political philosophy that is *achievable*, i.e., an accessible normative theory starting from our actual world. This is not a definition, as Rawls would have said, but just an indication.

REFERENCES

Berlin, I. (1999). *Concepts and Categories*. Princeton, NJ: Princeton University Press.
Cath, Y. (2016). Reflective Equilibrium. In H. Cappelen, T. Gendler, & J. Hawthorne (Eds.), *Oxford Handbook of Philosophical Methodology*. Oxford: Oxford University Press, 213-230.
Cohen, G.A. (2008). *Rescuing Justice and Equality*. Harvard: Harvard University Press.
Colaiaco, J. (2001). *Socrates against Athens. Philosophy on Trial*. New York and London: Routledge.
Dunn, J. (1980). *Political Obligation in its Historical Context: Essays in Political Theory*. Cambridge: Cambridge University Press.
Estlund, D. (2014). Utopophobia. *Philosophy & Public Affairs*, 42(2), 113-134.
Freeman, S. (2014). Original Position, in E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/original-position/.
Galston, W.E. (2010). Realism in Political Theory. *European Journal of Political Theory*, 9(4), 385-411.
Grant, R. (2002). Political Theory, Political Science, and Politics. *Political Theory*, 30(4), 577-595.
Hall, E. (2015). How to do realistic political theory (and why you might want to). *European Journal of Political Theory*, 1-21.
Leopold, D., Stears, M. (2008). *Political Theory. Methods and Approaches*. New York: Cambridge University Press.
Miller, D. (2013). *Justice for earthlings: Essays in Political Philosophy*. Cambridge: Cambridge University Press.
Nagel, T. (2005). The Problem of Global Justice. *Philosophy & Public Affairs*, 33(2), 113-147.
Pettit, P. (2007). Analytical Philosophy, in R. Goodin, P. Pettit, T. Pogge (Eds.), *A Companion to Contemporary Political Philosophy*. Oxford: Blackwells, 7-38.
Plato (1991). *The Apology of Socrates*, Oxford: Clarendon Press.
O'Shea, J. (2000). Sources of Pluralism in William James, in M. Baghramian, I. Attracta (Eds.), *Pluralism: The Philosophy and Politics of Diversity*. London: Routledge, 17-43.
Rawls, J. (2005). *Political Liberalism*. New York: Columbia University Press.
Rawls, J. (2007). *Lectures on the History of Political Philosophy*. Boston: Harvard University Press.
Sen, A. (1980). Description as Choice, *Oxford Economic Papers*, 32(3), 353-369.
Walzer, M. (1987). *Interpretation and Social Criticism*. Cambridge, London: Harvard University Press.

TERESA MARQUES
LOGOS Group, University of Barcelona
teresamatosferreira@ub.edu

WHAT METALINGUISTIC NEGOTIATIONS CAN'T DO[1]

abstract

Philosophers of language and metaethicists are concerned with persistent normative and evaluative disagreements – how can we explain persistent intelligible disagreements in spite of agreement over the described facts? Tim Sundell recently argued that evaluative aesthetic and personal taste disputes could be explained as metalinguistic negotiations – conversations where interlocutors negotiate how best to use a word relative to a context. I argue here that metalinguistic negotiations are neither necessary nor sufficient for genuine evaluative and normative disputes to occur. A comprehensive account of value talk requires stronger metanormative commitments than metalinguistic negotiations afford.

keywords

metalinguistic negotiation, evaluative disagreement, verbal disputes, metanormative commitments

Phenomenology and Mind, n. 12 - 2017, pp. 40-48
DOI: 10.13128/Phe_Mi-21104
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

In recent work, David Plunkett and Tim Sundell have explored a new answer to the problem of persistent normative and evaluative disagreement – how to explain the persistence of intelligible disagreements in spite of agreement over descriptive facts? This is a problem that both philosophers of language and metaethicists are concerned with. In this paper, I raise doubts about the explanatory range of metalinguistic negotiations, and argue that they are neither sufficient nor necessary for evaluative and normative disputes of the right kind to occur. An account of value talk should accommodate the possibility of discourse that is literally evaluative or normative.

The explanation of evaluative and normative disagreement raises questions that metanormative theories must address. As Mike Ridge says,

> [W]e must now ask the question "What is a moral judgment?" On the one hand, we assess people's moral judgments as true or false, we subject them to epistemic norms, and they can figure in rational inferences. These features suggest that moral judgments are beliefs. On the other hand, moral judgments can guide action without the help of an independently existing desire. Furthermore, *intelligible moral disagreement can persist beyond agreement on the relevant facts*. These considerations suggest that moral judgments are desires (Ridge 2006, p. 304, my emphasis).

Ridge argues that hybrid expressivist theories are an attempt to offer a cohesive answer to this set of questions. Metaethical hybrid expressivist theories are theories about the nature of ethical discourse and judgment that combine cognitivism about ethical thought and talk, and expressivism about such judgments and claims. Cognitivist theories state that ethical claims express beliefs that can be true or false, whereas expressivist theories state that ethical claims

[1] I'd like to thank Nicholas Allott, Hrafn Asgeirsson, Bartłomiej Czajka, Bianca Cepollaro, Esa Díaz-Leon, Dan López de Sa, Ryan Doerfler, Luís Duarte d'Almeida, Manuel García-Carpintero, Max Kölbel, Euan MacDonald, Josep Joan Moreso, Marie Newhouse, Genoveva Martí, Michele Palmira, David Plunkett, Sven Rosenkranz, Diogo Santos, Alex Sarch, Isidora Stojanovic, Tim Sundell, Kevin Toh, Pepa Toribio, Chiara Valentini, Lisa Voigt, and Frederick Wilmot-Smith for comments and discussion of material in this paper. Work funded by the European Commission, for projects CAND and DIAPHORA [Grant agreements PIEF-GA-2012-622114; H2020-MSCA-ITN-2015-675415]; Ministerio de Economía y Innovación for projects ABOUT OURSELVES and CONCEDIS [Grant agreements: FFI2013-47948-P; FFI2015-73767-JIN], and AGAUR de la Generalitat de Catalunya for Grup Consolidat LOGOS.

express attitudes that are not beliefs, but desires, preferences, intentions to act, etc., i.e., they express attitudes that are action-guiding.

Other areas of discourse raise similar questions, and have also invited expressivist approaches. Aesthetic and personal taste statements are normally assessed as true or false, can figure in rational inferences, can guide action and experience, and intelligible aesthetic disagreement can persist beyond agreement on the relevant descriptive facts. Legal statements can also be assessed as true or false, are subject to epistemic norms and to other norms of rationality, they guide action, and there are important persistent intelligible legal disagreements.

Legal theorists, after H.L.A. Hart (2012), have assumed that legal statements present a sort of ambiguity: they can be internal, i.e. normative or prescriptive statements made from the point of view of the participants in a legal system, or they can be external, i.e. uncommitted descriptions of what the law is at jurisdiction. Any legal sentence can in principle be interpreted either way, e.g. 'The Fourteenth Amendment allows states to regulate bakery employees' work hours'. Theorists like Kevin Toh (2005, 2011) have argued for expressivism about legal discourse. In his 2005 paper, for instance, Toh offers an expressivist interpretation of internal legal statements where, by making a statement, a legal official expresses her acceptance of a fundamental rule of recognition of the legal system of her community, and presupposes that other members of the community accept the same fundamental rule. This expressivist approach would help explain how judges and lawyers may sometimes agree on all factual questions but still disagree on the fundamental rules of their legal system.

Expressivism has been defended in other domains. For instance, several authors have defended that knowledge is a normative concept, and that knowledge claims are expressive of the speaker's acceptance of certain norms. Blackburn (1998) claimed that the primary role of knowledge talk is to indicate that a judgement is beyond revision. Gibbard (2003) held that knowledge attributions indicate that the knower's judgments are reliable, and Field (1998) held that epistemic claims express commitments to sets of norms for belief formation. More recently, Chrisman (2007) has made the case that disputes over knowledge attributions suggest that knowledge attributions, e.g., 'Sally knows that the bank will be open on Saturday', display similar features to those that Ridge identified in the moral domain. Chrisman argued that disagreement can persist in spite of agreement on the fact that Sally has a (justified) true belief that the bank will be open tomorrow. He suggests that the disagreements in question concern which standard should be accepted.

A common feature in these various domains – the moral, the aesthetic, the legal, or the epistemic – is that the general description applies; there is an intelligible disagreement of kind *X* that can persist beyond agreement on the relevant descriptive facts surrounding *X*.

In recent individual and joint work, Plunkett and Sundell have offered a new explanation of some of the disagreements that meet this very condition. Their explanation starts by pointing out that some phrases or expressions can be used metalinguistically. After Sundell (2011), they state that metalinguistic uses can also convey disagreements. They then distinguish between *descriptive* metalinguistic disagreements, and *normative* metalinguistic disagreements – disagreements about what the linguistic facts should be, which they label *metalinguistic negotiations*. Sundell (2016) further argues that aesthetic and personal taste *evaluative* disputes, for example, evaluative disputes about whether some food is tasty or about whether some piece of music is lyrical, could be entirely explained as metalinguistic negotiations.

There are three desiderata that an account of evaluative or normative disagreements should satisfy. First, it should characterize evaluative or normative disputes as *normative*, and not, for instance, as *factual* disputes about whether the interlocutors share the same standards. Second, the explanation of an evaluative dispute in domain X must be based on considerations pertaining to that very domain. For instance, we should be able to characterize an aesthetic dispute based on issues that are, by their nature, intrinsic to aesthetic aspects of the topic in focus, and not, say, to comedic aspects. Third, an evaluative disagreement should be a *disagreement* and not an attempt to coerce or manipulate others into sharing one's point of view.[2]

Metalinguistic negotiations concern what the linguistic facts should be. This, Plunkett and Sundell claim, is a "distinctive normative" question. But this normative question is not evaluative in any sense, and, *qua* normative question, it can be of the wrong kind. It depends on the interests, intentions, and goals of the interlocutors of the context. If speakers are moved by specific concerns over values, metalinguistic uses can introduce those concerns in the negotiation over how to use a word. But nothing about metalinguistic negotiations requires such concerns to be taken into account, or to exist, when nothing about the meaning of the words used mandates that speakers should be concerned with values, as the examples discussed in the following sections illustrate. The main problem with Sundell's generalized suggestion, as I'll argue, is that metalinguistic negotiations are neither necessary nor sufficient for genuine evaluative and normative disputes of the right kind to occur.

2. Metalinguistic Uses and Negotiations

The notion of a metalinguistic negotiation was developed from the notion of *metalinguistic* or *context sharpening* uses of gradable adjectives. Barker (2002) noted that a speaker may assert a sentence like 'Feynman is tall' either to give information about Feynman's height, or to give information about the threshold for height in a context. The latter is a metalinguistic use of 'tall'. Sundell (2011) in turn suggested that if information can be communicated through metalinguistic uses, then the information thus conveyed could also be the focus of disputes. So-called metalinguistic uses are common. For instance, there are apparent metalinguistic uses in sentences like (1). The material in the scope of the negation is not used literally, but echoically or metarepresentationally. There is a wide range of foci of metarepresentational disputes, including the right pronunciation of 'tomatoes' in (2):

(1) It's not a car. It's a Volkswagen.
(2) You like tom[eiDouz] and I like tom[a:touz].[3]

Plunkett and Sundell make their case for metalinguistic negotiations on the basis of examples like (3) and (4) below. Normally, (3) would be used in order to add to the common ground new information concerning Feynman's height. But Feynman's height may be common knowledge, and (3) could instead be uttered to provide information about the threshold of 'tall' in the context; it is a "context-sharpening use" (Barker, 2002, p. 1). A different example is given by Ludlow (2008). He describes a debate he heard on sports radio about the greatest athletes of the 20th century, and the question under discussion was whether that list should include the racehorse Secretariat.

(3) Feynman is tall.
(4) Secretariat is an athlete.

Secretariat's case does not concern the sharpening of gradable adjectives, and 'athlete' is not a context-dependent word. However, the dispute over whether Secretariat is an athlete

[2] Kevin Toh highlighted the importance of these three desiderata for accounts of fundamental legal disputes in a conference presentation of 2016.
[3] The examples are based on cases discussed in Horn (1989).

does not concern factual matters about the horse's race performance, over which people agree. Plunkett and Sundell say that a dispute over whether Secretariat is an athlete may be a dispute between two people with different conceptions of what 'athlete' means, and disputants can disagree about which competing concept 'athlete' should express (one that includes non-human animals in its extension, versus another that does not) (Plunkett and Sundell, 2013a, p. 16).[4]

Plunkett and Sundell discriminate between two kinds of metalinguistic disagreements: the descriptive and the normative. Descriptive metalinguistic disputes are disagreements about what the linguistic facts actually are. Normative metalinguistic disputes, or negotiations, are negotiations over what the linguistic facts *should be* in a given context. This, as they say, is a distinctive normative question.

According to Plunkett and Sundell, metalinguistic negotiations have two fundamental components. First, they are disputes not directly about the truth of the semantic content conveyed by a used sentence. Second, metalinguistic negotiations involve "conceptual ethics",[5] a normative activity that "concern(s) a distinctive normative question – how best to use a word relative to a context". Possible metalinguistic disputes over the cases above fall under the range of disputes about how (if at all) to use a word in a context. These processes may be sensitive to moral considerations and reasoning, although various concerns may push towards one or another way of making a given word more precise, or of fixing the meaning of a word that was previously undetermined.

In a recent article, Sundell (2016) takes the extra step of arguing for a more radical thesis about aesthetic and taste predicates. He argues that semantic theories for ordinary relative gradable adjectives (e.g. 'tall' or 'large') can fully account for the semantic properties of apparent value words, like personal taste or aesthetic predicates. On his newest proposal, the context-sensitivity of these adjectives is of the same kind as that of relative gradable adjectives. Aesthetic and personal taste adjectives are not semantically evaluative and do not require context to determine an aesthetic, personal taste, or experiencer parameter.

Kennedy (2007) identified three features of relative gradable adjectives: to exhibit contextual variation in truth-conditions, to have borderline cases, and to give rise to Sorites paradoxes. Sundell claims that taste and aesthetic adjectives have all of these features. He suggests that aesthetic and taste adjectives are relative gradable adjectives, and a theory of aesthetic and taste adjectives should give an explanation of why we tend to associate them with evaluations. The application of semantically gradable adjectives like 'tall' is complicated, he notes. Speakers may dispute which scale or comparison class should be applied in context; they can dispute how much weight to give to different scales or comparison classes; they can dispute the threshold of application of the adjective; or speakers may agree on the relevant scale or comparison class, agree on what weight to give to different dimensions, agree on the threshold of application of the predicate, and use tallness claims to dispute someone's height (pp. 16 ff.). Sundell concludes, thus, that there is a fairly complex range of possible *foci* of disputes involving gradable adjectives. Given the case for recognizing that aesthetic and taste adjectives are also relative and gradable, all the complexity of uses of taste and aesthetics adjectives could be explained with the resources available to explain the context-sensitivity of relative gradable adjectives. The account would anyway allow speakers to agree on the relevant scale, comparison class, etc., and use taste (or beauty) claims to dispute *directly*

[4] I have expressed doubts about this interpretation of the case before (Marques, 2017), but those doubts are not crucial for the present paper.

[5] See Burgess and Plunkett (2013) for further development of the notion.

something's taste (or something's beauty). Strange as it may seem, these latter disputes would be *descriptive*, and not *normative or evaluative*.

Sundell further proposes that metalinguistic negotiations over how to use aesthetic and taste adjectives account for their apparent normativity by allowing speakers to disagree about what the threshold, the weighing of scales, or the comparison class should be. A metalinguistic negotiation over how to use a word in a context is guided by the fact that the word in question plays a specific functional role in our lives and interactions. This role is bound by our preferences, motivations, and goals.[6] Since metalinguistic negotiations concern a "distinctively normative" question, the seeming normativity of personal taste or aesthetic discourse would be inherited from the normativity of the metalinguistic negotiation itself, or from the normative concerns that motivate and guide speakers in the conversational context. An advantage of this account is that the resulting semantic theory would be metaphysically neutral and compatible with a range of different metanormative views.

In summary, Sundell claims that

(i) Aesthetic and personal taste adjectives are not semantically evaluative, they are semantically descriptive – there's no implicit parameter for an aesthetic/taste standard, and there's no experiencer parameter;

(ii) The meaning of aesthetic and personal taste adjectives is not relativized to an experiencer or standard.

(iii) Evaluative aesthetic and personal taste disagreements are metalinguistic.

Presumably, these central claims could be generalized to any other evaluative and normative domains, for instance all those that have invited hybrid expressivist solutions, like those discussed earlier, e.g. knowledge attributions or internal legal statements. It would be a further strength of Sundell's proposal if it generalized to other normative or evaluative domains of discourse.

3. The Value Neutrality of Metalinguistic Negotiations

I think that metalinguistic negotiations are neither necessary nor sufficient to account for core evaluative disagreements. I will argue for this claim in the remainder of the paper.

The normative dimension of metalinguistic negotiations – conceptual ethics – concerns what the *linguistic facts should be* in a conversational context. In this respect, 'conceptual ethics' is a misleading title. First, because conceptual ethics is not necessarily conceptual; it is one thing to disagree about how to pronounce 'tomatoes', another about whether Volkswagens are great cars. Neither hinges on strictly conceptual issues, nor do context-sharpening disputes about 'tall'. When I say that neither hinges on conceptual issues, what I mean is that if we are disagreeing about how to pronounce 'tomatoes' (either as tom[eiDouz] or as tom[a:touz]), we are not disagreeing about which concept 'tomato' expresses. Likewise, when someone states that a vehicle is not a car, but a Volkswagen, she does not question which concept we should deploy in uttering the word 'car'.

Second, because conceptual ethics is not ethics. All the normative questions involved in negotiating how to use a word at a context can respond to mere practical reasons, i.e. to prudential or procedural reasoning, given the interlocutors' current goals and interests. Nothing about the meaning of the words used mandates that speakers should be concerned with *value*, and nothing about the nature of metalinguistic negotiations (as described) requires such concerns to be taken into account. Thus, whether or not a child counts as tall may

[6] See for instance Sundell (2016, p. 21).

depend on whether speakers are deciding whether she's tall enough to go on a roller coaster, or whether she is healthy for her age group. It is true that 'tall' can be used normatively, for instance, to assess if a child is developing correctly. But nothing about the semantics of 'tall' requires such a use.

For the same semantic treatment to be relevant for 'tasty' or 'beautiful', it would have to be the case that claims of taste or beauty are entirely dependent and determined by the occurrent interests, goals, and intentions of speakers in a conversational context. That is what follows from the assumption that taste and aesthetic adjectives are semantically context-sensitive in the same way as relative gradable adjectives.

Sundell's proposal could be extended to other presumably evaluative or normative domains, as I mentioned in the introduction, e.g. to knowledge attributions, moral statements, or legal statements. If nothing requires speakers to introduce an intrinsically evaluative or normative dimension that is not required by the semantics, then, for instance, the question of how best to use 'know' may not address the motivation to treat knowledge attributions as normative. Yet, epistemic expressivists are concerned with the essential normative status of knowledge attributions: that the primary role of knowledge talk is to indicate that a judgment is beyond revision (Blackburn), that knowledge attributions indicate that the knower's judgments are reliable (Gibbard), or that epistemic claims express commitments to sets of norms for belief formation (Field). There may be resilient *epistemic* disputes about whether *S* knows that *p*, and an explanation of these resilient disputes as normative ought to capture their intrinsically epistemic character.

The main problem with Sundell's claim about aesthetic and taste adjectives, and with its possible application in other domains, is that metalinguistic negotiations are neither necessary nor sufficient for genuine evaluative and normative disputes *of the right kind* to occur. For instance, in a certain context there may be prudential or procedural reasons to say 'That is a beautiful painting' that respond to the interests and goals of the speakers in the conversation, but do not respond to any genuine aesthetic reasons or values.

Imagine an art critic *A* that has always refused to give positive reviews of artist *B*'s work. We can further imagine that *B*'s work is exceptional. Her paintings are beautiful, and that is the view of the majority of the art world and the public. In fact, the reason why *A* does not give positive reviews of *B*'s paintings is that *A* is not a very good art critic: he is a superficial critic, he is ignorant of art history and current trends, he is a sexist, and he is in the business motivated merely by the power, fame, and money that surrounds the art world. Now, imagine that the artist, *B,* has seduced the critic, *A*. *A* is finally willing to write a positive review of *B*'s work. *A*'s infatuation with *B* has led him to see everything she does with a positive valence:

> (5) The author's painting is beautiful.

In writing (5), the art critic is responding to the wrong kind of reason.[7] His use of 'beautiful' responds to reasons that, in the conversational context, do not respond to aesthetic features of the painting. Yet, the reasons that motivate him to write (5) do respond to the interests and goals of the speakers in the conversation. If nothing about the meaning of 'beautiful' as applied to an artwork requires an evaluative aesthetic standard, then a metalinguistic negotiation about how 'beautiful' should be used in the context need not require any such standard. It follows that metalinguistic negotiations are not sufficient for disputes to be about aesthetic value.

[7] On the wrong kind of reasons problem, see for instance Hieronymi (2005).

Metalinguistic negotiations are also not necessary for evaluative aesthetic disputes. It is part of Sundell's argument that aesthetic adjectives are not semantically evaluative, and that their context-sensitivity is like that of relative gradable adjectives. However, there are reasons not to regard aesthetic adjectives as relative gradable adjectives. In recent work, Liao *et al.* (2016) show that the standard relativity of aesthetic adjectives is not dependent on the immediate situational context, unlike that of 'large', 'long', or 'tall'. In their paper, they report recent experimental work that found that aesthetic adjectives behave neither like relative nor like absolute gradable adjectives. Their results suggest that aesthetic adjectives depend on more stable and general aesthetic standards, in particular, they allow for judgments of beauty or elegance that do not require a standard of application to be supplied by the immediate conversational context.[8]

Sundell's radical proposal is that aesthetic and personal taste adjectives are not semantically evaluative and that there is no need for an aesthetic standard parameter to be part of the semantics of aesthetic adjectives. All context-sensitivity would be that of regular relative gradable adjectives. But Liao *et al.* (2016)'s results indicate a) that aesthetic adjectives' context-sensitivity depends on aesthetic standards, and b) that the selection of an aesthetic standard is not determined by the immediate conversational context.

Since it is possible to apply an aesthetic adjective like 'beautiful' to a statue or a painting without requiring further specific information to be provided by the conversational context, the following is also possible. First, two people may agree on all the formal features of a painting, and second, they may still disagree about whether the painting is beautiful. Furthermore, they may disagree about whether the painting is beautiful even though they don't share a conversational context. And they may do so in spite of not engaging in any negotiation about how to use 'beautiful', and without sharing a common goal towards which they can coordinate their use of the word.

In other words, two people can disagree about whether a painting is beautiful while not disagreeing about what the linguistic facts concerning 'beautiful' should be. And their disagreement is not a mere *factual* disagreement either. Hence, metalinguistic negotiations are neither sufficient nor necessary for evaluative aesthetic disagreement. I conjecture that this conclusion generalizes to other evaluative domains.

There is no doubt that pragmatic mechanisms like metalinguistic negotiations exist, are important, and capture a way in which people may agree about some descriptive facts, while disagreeing about how language is to be used. But the idea that metalinguistic negotiations can account for all value disputes is, as I have argued, mistaken. Metalinguistic negotiations do not explain all cases of evaluative or normative disputes. Any account of value talk should accommodate the possibility of discourse that is literally evaluative or normative, and not only metalinguistically so.

Part of the reason that metalinguistic negotiations do not suffice, by themselves, to explain all persistent evaluative disagreements is that there are central claims that are semantically evaluative. A theory about the semantics of evaluative adjectives, for instance aesthetic or personal taste predicates, that leaves no room for value to be semantically encoded cannot explain what unifies a set of adjectives as *aesthetic* or of *taste*. Some of the core distinctive roles that evaluative discourse is presumed to play, besides its *cognitive* role, are that it normally *expresses speakers' conative attitudes,* that it is (normally) motivational, and that it serves a *connection building role*. These are roles that talk of height or distance, for instance, does not uniformly play. This manifests a disparity between the role of evaluative discourse and of

[8] They also found that aesthetic adjectives do not behave exactly like absolute gradable adjectives.

uses of non-evaluative relative gradable adjectives. These distinctive features require, in my view, a comprehensive explanation of how value discourse serves to communicate values, and stronger metanormative commitments than Sundell seems willing to undertake.

REFERENCES

Barker, C. (2002). The dynamics of vagueness. *Linguistics and Philosophy*, 25(1), 1-36.

Blackburn, S. (1998). *Ruling Passions: A Theory of Practical Reasoning.* New York: Oxford University Press.

Burgess, A., & Plunkett, D. (2013). Conceptual Ethics I. *Philosophy Compass*, 8(12), 1091-1101.

Chrisman, M. (2007). From Epistemic Contextualism to Epistemic Expressivism. *Philosophical Studies*, 135(2), 225-254.

Field, H. (1998). Epistemological Nonfactualism and the A Prioricity of Logic, *Philosophical Studies*, 92, 1-24.

Gibbard, A. (2003). Thinking How to Live. Cambridge, MA: Harvard University Press.

Hart, H.L.A. (2012). *The Concept of Law*. Oxford University Press, 3rd edition.

Hieronymi, P. (2005). The Wrong Kind of Reason. *Journal of Philosophy*, 102(9), 437-457.

Horn, L. (1989). *A natural history of negation.* Chicago: Chicago University Press.

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison.* New York: Garland (1997 UCSC Ph.D thesis).

Liao, S., McNally, L., & Meskin, A. (2016). Aesthetic adjectives lack uniform behavior. *Inquiry*, 59(6), 616-631.

Marques, T. (2017). Can metalinguistic negotiations and 'conceptual ethics' rescue legal positivism?. In A. Capone and F. Poggi (Eds.), *Pragmatics and Law: Practical and Theoretical Perspectives.* Cham, SZ: Springer, 223-241.

Plunkett, D. & T. Sundell (2013a). Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint*, 13(23), 1-37.

Plunkett, D. (2015). Which Concepts Should We Use: Metalinguistic Negotiations and the Methodology of Philosophy. *Inquiry*, 58(7-8), 828-874.

Ridge, M. (2006). Ecumenical expressivism: Finessing Frege. *Ethics*, 116(2), 302-336.

Sundell, T. (2011). Disagreements about taste. *Philosophical Studies*, 155(2), 267-288.

Sundell, T. (2016). The tasty, the bold, and the beautiful. *Inquiry*, 59(6), 793-818.

Toh, K. (2011). Legal judgments as plural acceptance of norms. In L. Green and B. Leiter (Eds.), *Oxford Studies in Philosophy of Law.* Oxford: Oxford University Press, DOI: 10.1093/acprof:oso/9780199606443.001.0001.

Toh, K. (2016). Collectivity and the Law, manuscript presented in the conference *Collective Action and the Law* that took place at University Pompeu Fabra in September 2016.

GIULIANO TORRENGO
Centre for Philosophy of Time, Department of Philosophy,
University of Milan
giuliano.torrengo@unimi.it

THE MYTH OF PRESENTISM'S INTUITIVE APPEAL[1]

abstract

Presentism, the view that only what's present exists, seems to be intuitively very appealing. The intuitive appeal of presentism constitutes a main reason for treating the view as a serious option and worthy of consideration. In this paper, I argue that the appearance of presentism's intuitiveness is based upon a series of misconceptions.

keywords

presentism, eternalism, common sense, intuition

Phenomenology and Mind, n. 12 - 2017, pp. 50-56
DOI: 10.13128/Phe_Mi-21105
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

THE MYTH OF PRESENTISM'S INTUITIVE APPEAL

1. The Received View

Presentism is the view that only what's present exists. On this view, merely past entities (such as the dinosaurs or Julius Caesar) and merely future entities (such as human outposts on Mars) do not exist at all. This seems to be a natural thing to think, which makes presentism very appealing at an intuitive level. At least, presentism appears more natural and intuitive, prima facie, than its main rival: eternalism. Eternalism is the view that past, present, and future entities exist. On this view, dinosaurs, Roman Emperors, and Mars outposts all exist, just as you and I and the Trevi Fountain exist.

It seems that the intuitive appeal of presentism constitutes a main reason for treating the view as a serious option and worthy of consideration. And, although not all presentists appeal to the apparent intuitiveness of presentism in order to motivate or defend their position, the idea that presentism is more intuitive than eternalism – or closer to "common-sense" than eternalism – has, to my knowledge, never been seriously challenged.

Let's review a few brief quotes to corroborate my claim that the intuitive appeal of presentism is the received view for presentists and non-presentists alike to dispute. That is:

> Though I think presentism ultimately must be rejected, its guiding intuition is compelling: the past is no more, while the future is yet to be (Sider, 2001, p. 11).

> I endorse Presentism, which, it seems to me, is the 'common sense' view, i.e. the one that the average person on the street would accept (Markosian, 2004, p. 48).

> The natural, intuitive, view is that the past is not a part of what exists. Indeed, presentism, the view that *only* the present exists, is taken to be our intuitive view of time (Tallant, 2009, p. 425).

The idea that underlies the "received view" (of presentism's intuitiveness) seems simple and straightforward. We naturally tend to think that what's past has existed (it was present), but exists no more, and what's future will exist (it will be present), but doesn't exist yet. And

[1] For useful comments, thanks to Dave Ingram and Bianca Cepollaro. Thanks to the project *2015-0746* (15-5-3007000-601) of Fondazione Cariplo and Regione Lombardia, and project 15-6-3007000-2021 of Università degli Studi di Milano for financial supports.

presentism fits with this way of thinking: only what's present exists. Thus, we have a natural tendency to incorporate presentism into our ordinary way of thinking.
In what follows, I argue that the received view should be abandoned, since it's unjustified, appearances notwithstanding. Even if our intuitions or "common-sense" support presentism in part, it's not true that common-sense favors presentism over eternalism. It strikes me that, like in other cases, common-sense doesn't coherently (or emphatically) pull in one direction rather than another.[2]

2. Common-Sense and One Argument from Intuitiveness

Here, I understand 'common-sense' to be a set of beliefs that are widely shared among a given group of people. And I will refer to the common-sense beliefs that are, in some sense, suggested to us by our ordinary experience as 'intuitions'. These notions of common-sense and intuitions are neither uncontroversial nor unproblematic. But I take it that such notions should be acceptable to both parties in the debate between presentists and eternalists. Further, I make the following three provisos:

(a) Cultural and temporal variability. Although there's evidence that common-sense beliefs do vary through cultures and across historical periods, such variation is irrelevant for my purposes here.
(b) The coherence of common-sense. Intuitions can be incompatible with each other. This implies that, globally, common-sense is likely to be incoherent. Even so, it's possible to individuate and partition elements of common-sense that are coherent.
(c) The structure of common-sense. We can individuate beliefs that are more or less central to common-sense. Roughly, the more peripheral beliefs (i.e. those that are less central) are those that are easier to reject or else accommodate against evidence to the contrary.

If this framework is accepted, I take it to be uncontroversial that the following claim ("Intuition 1") is part of common-sense:

> **Intuition 1.** What has existed (and exists no more) is not what we meet in the present.

Now, both presentists and non-presentists alike should agree on **Intuition 1**, since even eternalists believe that we *do not* (*cannot*) meet in the present things that are wholly located in the past. However, since presentism and eternalism are intended to be distinct and competing positions in temporal ontology, they must (substantively) disagree on some claims concerning what exists – in the sense relevant for ontology. Let's suppose that presentists and the rest disagree about what exists *simpliciter* (cf. Sider, 2006; Torrengo, 2012). More precisely, presentists claim that merely past entities (dinosaurs, Caesar, etc.) don't exist *simpliciter*, whereas non-presentists disagree. Non-presentists, such as eternalists and growing block theorists, claim that merely past entities do exist *simpliciter*.

One who defends the intuitive appeal of presentism may then argue as follows. Even if the eternalist agrees that we cannot meet merely past things (in the present), there's a very natural way to read **Intuition 1** that's incompatible with the denial of presentism. What's

[2] I remain neutral (at least, until §5) on whether support from common-sense is a theoretical virtue and, if so, how it should be measured against other virtues such as ontological (and ideological) parsimony, simplicity, and other respects of explanatory power. It may well be that intuitive judgement should be disregarded (or else not appealed to) in metaphysics, as Jiri Benovsky (2013) has argued.

required is to construe talk of existence in **Intuition 1** in terms of the more perspicuous (and disputed) notion of existence *simpliciter*. Thus, the presentist's view of temporal ontology can be seen as a transposition in those terms of the ordinary intuition that what's past is no more (exists no longer). In slightly more precise terms, we have the following argument:

(1) **Intuition 1** should be construed in terms of existence *simpliciter*.
(2) Thus construed, **Intuition 1** states that what *existed* doesn't exist *simpliciter*.
(3) **Intuition 1** is incompatible with eternalism.

3. The Eternalist's Rejoinder

Unfortunately for presentists, the argument introduced at the end of the last section isn't very good at all. Eternalists may doubt (1) quite naturally. It seems assured that an eternalist can maintain **Intuition 1** and, thus, agree with the presentist that we don't meet (in the present) entities that don't exist in the present. What our eternalist needs to add, however, is that what's expressed by past-tensed locutions such as 'existed' shouldn't be understood in terms of existence *simpliciter*. This seems reasonable. After all, why should we think that an ordinary intuition about what exists no longer in the present (about what has ceased to be present) has anything to do with an intuition about what exists *simpliciter*? We might well ask if we have any ordinary intuitions about what exists *simpliciter* at all. As I see it, the most straightforward reading of **Intuition 1** is neutral with respect to how talk of (tensed) existence is to be construed within a theoretical framework. And, if this is so, then the idea that presentism entails ordinary intuitions that eternalism rejects seems to be the result of a simple misunderstanding about how to construe properly those common-sense beliefs. More importantly, and apart from the above point, it looks here as if common-sense, as in other cases, comprises intuitions that support contrasting and competing views. For instance, the ordinary notion of existence (such as it is) seems to be captured by common-sense beliefs of the following sort:

> **Intuition 2.** What exists possesses causal powers and is located in space and time (cf. Berto, 2014; Reicher, 2016).

Tacit or unreflective endorsement of **Intuition 2** is why we tend to say that common-sense supports the idea that fictional and abstract entities don't exist. Sherlock Holmes, the fictional detective, lacks causal powers (although a thought about Sherlock Holmes can have causal powers) and isn't located in space or time. By contrast, real detectives have powers and locations. I cannot meet Holmes, much as I would like to. But I can meet any number of real detectives. Something similar is true for paradigmatically abstract entities, such as the number 5 or the empty set. The question arises: can we follow the same line of thought for merely past entities? Our ordinary experience seems to testify to the contrary. It's not only that we think about the past and talk about it with simple ease, we are also directly and indirectly affected by it. We all act on the assumption that what has happened (in the past) led to what's happening now (even if we are not determinists of any stripe). Of course, presentists deny existence to past entities, and have some reason to think that this thought is mistaken, but then this ordinary thought supports eternalism over presentism. There's a tension with the received view when we reflect on the intuitive causal efficacy of the past.

What about location in space and time? Do past entities have it or not? It seems that they do have location in time: they are located *in the past*. Presentists may well claim that it's obvious that we cannot meet past things (in the present) as we meet present things, as **Intuition 1** has it. But the claim that we think of past things as *outside* time seems manifestly false. At any rate, it's far from clear that the ordinary thought that we cannot meet past things supports the thesis that the past things lack existence in any sense. Indeed, once we reflect on it, it seems to support the claim that past things are in space and time, but located away from us (in a sense). But this is merely a matter of indexicality. The past is never here and now. We never meet past things.

4. Further Debunking the 'Myth' of Intuitiveness

Continuing and developing the point from the previous section, it strikes me as nearly analytic that past things are within time. Although it's perhaps divisive to assert confidently that past things *are* spatiotemporally located, it seems far less controversial to rule out the claim that past things *aren't* so located. If we understand 'abstract' in terms of what isn't located in space and time, the following claim ("Intuition 3") seems to be part of common-sense:

Intuition 3. What is past *isn't* abstract.

Of course, presentists don't need to deny **Intuition 3**, since it's compatible with the idea that what's past doesn't exist at all either. Put differently, presentists can combine **Intuition 3** with the claim that what exists in space and time is restricted to a constantly changing present. Thus, past entities lack spatiotemporal location and lack abstract existence. But again this means that ordinary intuitions fall far short of confirming or supporting the presentist's thesis that only what's present exists (*simpliciter*). **Intuition 3**, by itself, appears to support the claim that past things are concrete (not abstract) and located in time. This claim is distinctly non-presentist. It's compatible with eternalism or growing block.

Common-sense doesn't seem to provide a theoretical advantage for presentism in this case. On one hand, we ordinarily assume that what's past has ceased to be (i.e. has left reality for good), and presentism takes this intuition very seriously. Yet, on the other hand, we accept that the past has an effect on us. And it's more than a little strange to think of something that has an effect on us as utterly outside existence.

5. A Pyrrhic Victory

From the discussion above, it follows that presentism is more intuitive than eternalism *only if* the ordinary idea that what's past has gone out of existence is more central to common-sense than the idea that the past has a causal effect upon the present. Thus, if it turns out that we are considerably less keen to revise **Intuition 1**, rather than **Intuition 2** or **3**, when confronted with theoretical beliefs that seem to be incompatible with it, then the received view would be vindicated. I take the issue of establishing which intuition is more central, at least in this case, to be a tricky one. Besides, even if we establish which intuition is "more central", whether the centrality of a "supporting" intuition constitutes a theoretical advantage for a metaphysical position still depends on what are the theoretical options at our disposal.

For instance, suppose that the intuition that the present is causally effected by the past is more central (more *fundamental*) than the intuition that the past is gone for good (nowhere and "nowhen"). On this supposition, would eternalism have an advantage? Well, it depends on how well a presentist can explain causal influence from the past without assuming the existence of the past causes (cf. Crisp, 2005). I don't find the presentist account of causation (and the metaphysics of the causal relation) entirely convincing, along with others (cf. Davidson, 2003). However, even if we grant that an explanation can be formulated, it seems more correct to say that presentism and eternalism would be in a stalemate.

Similarly, suppose that the belief that the past lacks existence (*simpliciter*) is more basic in our naive "conceptual scheme" than any "eternalism friendly" intuition to the contrary (and that it can be shown it to be the case). The kind of advantage over eternalism that presentism gains from **Intuition 1** being "core" common-sense would be extremely shallow. The eternalist can accept this fact and explain it away, i.e. explain why it *seems* to us that the past lacks the relevant kind of existence. We don't need to argue for the claim that elements of our best (well-established) science, e.g. the special theory of relativity, confirm eternalism at the expense of presentism in order to deploy this empirically well-confirmed theory in serve of that aim. If one asks *why* we ordinarily think of the past as non-existent, elementary knowledge of physics allows us to explain that it's because we cannot physically effect what's outside the scope of causal relations of our future light-cone (cf. Butterfield, 1984; Callender, 2008).[3] Thus, even if presentism *is* more intuitive than eternalism, in one sense, it remains doubtful that this intuitiveness can be regarded as a theoretical advantage at all, because we have a scientifically acceptable explanation of why this intuition can lead us astray. Again, the situation looks suspiciously close to a stalemate.

6. Conclusions

In evaluating positions in metaphysics, philosophers sometimes appeal to intuitiveness as a theoretical virtue of some kind or other. Overall, such a stance may be sound. If two theories are roughly on a par with respect to their explanatory capacity (and other theoretical virtues, such as parsimony), then the theory that allows us to retain more from everyday thought and common-sense is preferable.

However, when we make claims about the intuitive appeal of a metaphysical thesis, there should be a clear sense in which the theory at issue is definitively more intuitive than its rivals. In the case of presentism, I've argued that there's no such clear sense. There are indeed certain intuitions that seem to be supportive of presentism, but (i) on the most straightforward reading, they actually support a trivial thesis that an eternalist could accept, and (ii) on the reading in which they are incompatible with eternalism, they are counterbalanced by contrary intuitions which are incompatible with presentism. The intuitive appeal of presentism is a myth. At least, it doesn't confer the motivation typically advertised.

REFERENCES

Benovsky, J. (2013). From experience to metaphysics: on experience-based intuitions and their role in metaphysics. *Noûs*, 49, 684-697.

Berto, F. (2014). Modal noneism: transworld identity, identification, and individuation. *The Australasian Journal of Logic*, 11, 61-89.

Butterfield, J. (1984). Seeing the present. *Mind*, 93, 161-176.

Callender, C. (2008). The common now. *Philosophical Issues*, 18, 339-361.

Dyke, H. & Maclaurin, J. (2002). 'Thank goodness that's over': the evolutionary story. *Ratio*, 15, 276-292.

Le Poidevin, R. (2007). *The Images of Time*, Oxford: Oxford University Press.

Markosian, N. (2004). A defense of presentism. In D. Zimmerman (Ed.), *Oxford Studies in Metaphysics*, 1, Oxford: Oxford University Press, 47-82.

Paul, L. (2010). Temporal experience. *The Journal of Philosophy*, 107, 333-359.

Reicher, M. (2016). Nonexistent objects. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016), https://plato.stanford.edu/archives/win2016/entries/nonexistent-objects/.

[3] Eternalists often provide error-theoretic explanations of the intuitively less appealing aspects of their theories. See Dyke & Maclaurin's (2002) evolutionary explanation of why the same event elicits different emotional responses from us depending on whether it's past, present, or future; and, see Le Poidevin (2007), Paul (2010) or Torrengo (2017), amongst others, for an explanation in terms of psychology and philosophy of perception of why we have the "feeling" that time flows.

Sider, T. (2001). *Four-Dimensionalism*. Oxford: Oxford University Press.
Sider, T. (2006). Quantifiers and temporal ontology. *Mind*, 115, 75-97.
Tallant, J. (2009). Ontological cheats might just prosper. *Analysis*, 69, 422-430.
Torrengo, G. (2012). Time and simple existence. *Metaphysica*, 13, 125-130.
Torrengo, G. (2017). Feeling the passing of time. *The Journal of Philosophy*, 114(4).

SUBMITTED CONTRIBUTIONS

SUBMITTED CONTRIBUTIONS

Dan Zeman
Contextualist Answers to the Challenge from Disagreement

Diogo Santos
How to Dispel the Asymmetry Concerning Retraction

Simone Carrus
Slurs: At-issueness and Semantic Normativity

Andrés Soria Ruiz
Thomason (Un)conditionals

Paolo Labinaz
Assertion and the Varieties of Norms

Enrico Cipriani
Chomsky on Analytic and Necessary Propositions

Hashem Ramadan
The Two-Way Relationship Between Language Acquisition and Simulation Theory

Marco Fenici
Rebuilding the Landscape of Psychological Understanding After the Mindreading War

Alessandra Buccella
Naturalizing Qualia

Marco Viola
Carving Mind at Brain’s Joints. The Debate on Cognitive Ontology

Joana Rigato
Looking for Emergence in Physics

DAN ZEMAN
Institute of Philosophy
University of Vienna
danczeman@gmail.com

CONTEXTUALIST ANSWERS TO THE CHALLENGE FROM DISAGREEMENT[1]

abstract

In this short paper I survey recent contextualist answers to the challenge from disagreement raised by contemporary relativists. After making the challenge vivid by means of a working example, I specify the notion of disagreement lying at the heart of the challenge. The answers are grouped in three categories, the first characterized by rejecting the intuition of disagreement in certain cases, the second by conceiving disagreement as a clash of non-cognitive attitudes and the third by relegating disagreement at the pragmatic level. For each category I present several important variants and raise some (general) criticisms. The paper is meant to offer a quick introduction to the current contextualist literature on disagreement and thus a useful tool for further research.

keywords

contextualism, disagreement, disagreement in attitude, pragmatic disagreement, predicates of taste

Phenomenology and Mind, n. 12 - 2017, pp. 62-73
DOI: 10.13128/Phe_Mi-21106
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Suppose I utter the sentence

(1) Marmitako[2] is delicious,

and my lifelong friend and partner in culinary endeavors utters its negation. On the face of it, we disagree. After all, at minimum, I claim that a certain food (marmitako) has a certain property (being delicious), while my friend claims that the same food doesn't have that property. This looks like a textbook case of disagreement.

Such (admittedly raw) data has played a great role in the contemporary debate between various semantic views about predicates of taste and similar subjective – or, as I will refer to them throughout this paper, *perspectival expressions*. The main characteristic of perspectival expressions is that appeal to a subject's perspective is needed for their semantic interpretation. Besides predicates of taste like 'delicious', aesthetic adjectives like 'beautiful', moral terms like 'good' or 'ought', epistemic modals like 'might' and 'must', gradable adjectives like 'tall', epistemic vocabulary like 'know' have been thought of as perspectival. What counts as a "perspective" in each case is, of course, different, but here I'm using the term in a broad sense to refer to whatever element captures the subjective character of the relevant expressions.

The point of drawing attention to exchanges like the one between my friend and I about marmitako was to show that certain views in the debate cannot accommodate the intuition of disagreement that seems to be present in the exchanges in question. For example, according to the view in the debate focused on in this paper – contextualism – when speakers utter sentences like (1) or their negations, the propositions they express are *perspective-specific* – that is, perspectives are part of the expressed propositions. In contrast, for contextualism's main rival, relativism, when speakers utter sentences like (1) or their negations, the semantic contents they express are *perspective-neutral*, with perspectives being relegated to the "circumstances of evaluation" (a technical term familiar from Kaplan (1989), comprising possible and actual

[1] I would like to thank the editors of this issue and two anonymous referees of the paper for their comments. Work on this paper has been financially supported by a MINECO "Juan de la Cierva" postdoctoral grant, the *Foundations and Methods of Natural Language Semantics* project at the University of Barcelona and the *Lexical Meaning and Concepts* project at the University of the Basque Country.

[2] https://en.wikipedia.org/wiki/Tuna_pot.

situations in which an utterance is evaluated for truth).[3] And while the perspectives relevant for evaluating each utterance can be the same, they need not be: in the exchange above, for example, the perspective relevant for the interpretation of my utterance is mine, while that relevant for the interpretation of my friend's utterance is hers. In this case, disagreement – at least in an intuitive understanding of it – is not accounted for by the contextualist.

This, in a nutshell, is the challenge from disagreement that present day relativists have leveled against contextualism. Recently, however, several ways to meet the challenge have surfaced in the literature. While some of those ways have been sporadically engaged with, and while a number of papers describe various strategies available to contextualists (e.g., Stojanovic, 2007; Marques & Garcia-Carpintero, 2014; Silk, 2016; Khoo, 2017), a detailed systematization of the answers and their variants has not been done. The present paper aims to do precisely that. This will provide both a quick introduction to the latest dialectical moves of the debate focused on and a (hopefully) useful resource for future research. And while I focus on predicates of taste for illustration, the same strategies are possible (and some have been proposed) for other perspectival expressions as well.

Before getting to the strategies mentioned, it would be useful to spell out what the notion of disagreement that the proponents of the challenge have relied on is. Although the case for disagreement is well supported by intuitions,[4] relativists have relied on a specific way of understanding disagreement: as involving certain types of *cognitive attitudes* (belief, judgment, acceptance etc.) directed towards propositions. What exactly is the nature of the cognitive attitude involved in disagreement varies among relativists; here I rest content with describing the attitudes in question as "doxastic", while leaving open their exact nature.[5] The following characterization is I think in line with what most relativists have taken disagreement to be:[6]

> **Doxastic Disagreement** (**DD**)
> Two interlocutors disagree if they have opposite doxastic attitudes towards the same proposition.[7]

The contextualist conundrum can be understood by attending to (**DD**) in the following way: in order for the exchange between my friend and I about marmitako to count as disagreement,

[3] The terminology here is not ideally clear. As a referee points out, historically, views that have been called by their proponents "relativism" turn out to be, in the more recent understanding of the term, contextualist views (e.g. Harman, 1996; Dreier, 1990). Things are further complicated by the fact that authors like MacFarlane (2014) reserve the term 'relativism' to a certain view about the absoluteness of utterance/propositional truth. Here I take relativism to be the general view, exemplified above, that the perspectival character of the relevant expressions is to be accounted for by postulating a corresponding parameter in the circumstances of evaluation.

[4] Or so I will assume for the purposes of this paper. Most contextualists agree that such intuitions exist, and they attempt to solve the challenge from disagreement by taking them at face value. See, however, the discussion in section 1.

[5] I also leave it open whether disagreement can happen with other types of cognitive but non-doxastic attitudes such as assuming, enquiring, doubting etc. as well.

[6] See, for example, Kölbel (2004), Lasersohn (2005, 2016) etc. Famously, MacFarlane (2007, 2014) argues against (**DD**); Marques (2014) does too. However, the condition specified by (**DD**) remains a part of both author's proposed replacements: it is part of MacFarlane's improved definitions of disagreement (see "CAN'T BE BOTH ACCURATE (RELATIVE TO C)" in MacFarlane (2007, p. 26) and the various notions discussed in MacFarlane (2014, chapter 6)) and of Marques revised notion ("Doxastic Disagreement" in Marques (2014, p. 132)). Marques also notes that, strictly speaking, it is enough for disagreement to have opposite attitudes towards a proposition (on one hand) and a proposition entailed by the negation of the first proposition (on the other hand). The point is well taken; however, the fix is easily available (Marques provides one herself).

[7] A relevant distinction must be mentioned here: that between disagreement *in act* and disagreement *in state* (Cappelen & Hawthorne, 2009). Given that I deal here with exchanges, the former notion seems more suitable; however, this shouldn't be taken to mean that the latter notion is not more fundamental.

we have to hold opposite doxastic attitudes towards the same proposition. But if the propositions expressed by each of us is perspective-specific – so that the proposition expressed by my utterance of (1) is that marmitako is delicious to me (or to a group I belong to), while the proposition expressed by my friend uttering the negation of (1) is that marmitako is not delicious to her (or to a group she belongs to) – then, according to (**DD**), the exchange doesn't count as disagreement. Yet, the intuition is that my friend and I disagree; hence, the challenge. On the other hand, relativism gets the data right: given that the propositions expressed by my friend and I in the exchange above are perspective-neutral, there is a proposition we have opposite attitudes towards: namely, the proposition that marmitako is delicious, full stop.[8]

1. Rejecting the Intuition of Disagreement (in Certain Cases)

One immediate reaction to the challenge from contextualist quarters has been to question the disagreement data. This has not been done by flat-out denying that the intuition of disagreement in exchanges like the one between my friend and I about marmitako exists, but more indirectly by questioning the dialectical import of such exchanges and by claiming that, when they are suitably fleshed out, the contextualist can yield disagreement, even if conceived along the lines of (**DD**). Thus, contextualists have complained that the scenarios provided by relativists are too skeletal to support solid intuitions about disagreement. For example, Schaffer claims that "the case for relativism relies on a misrepresentative sample of underdeveloped cases" (Schaffer 2011, p. 211), and other authors (e.g., Glanzberg, 2007; Stojanovic, 2007; Cappelen & Hawthorne, 2009) have expressed similar opinions. Such authors then proceed to flesh out the said scenarios by employing uses of perspectival expressions that, if not strictly speaking neglected by relativists, have not been their main focus in mounting the challenge.

So, what are these uses contextualists have appealed to in the case of predicates of taste? First, predicates of taste and other perspectival expressions can be used *exocentrically* (that is, from another person's perspective), as opposed to *egocentrically* (from one's own perspective). The exchange between my friend and I regarding marmitako can be interpreted in a way in which we both use 'delicious' from another person's perspective (we are trying to decide where to take a common friend out for lunch on her birthday, say, and thus we both speak from *her* perspective), or as my friend trying to correct me about marmitako, given my previous unpleasant experiences with the food (thus speaking from *my* perspective).[9] Second, the expressions at stake can be used *collectively* (that is, from the perspective of a group). The exchange between my friend and I can thus be interpreted in a way in which we both use 'delicious' collectively (we are trying to decide where to organize the next department lunch, say, and thus we speak from the perspective of the entire group). Finally, the predicates in question can be used *generically* (that is, they convey how things stand from the perspective of the majority, or how things usually are seen). The exchange between my friend and I can thus be interpreted in a way in which we both use 'delicious' generically (we are discussing whether marmitako is generally considered delicious). In each of these cases the contextualist has an easy time explaining disagreement: given that in each case the relevant perspectives are the same, disagreement can be cashed out as two interlocutors having opposite attitudes towards *the same* proposition (albeit different in each case), which is exactly as (**DD**) requires.

Now, it is obvious that this much won't get the contextualist too far. And although it is hard to deny that predicates of taste and other perspectival expressions do have the uses

[8] There is a further issue whether relativists do indeed capture disagreement; I ignore this issue here.

[9] Each of these interpretations becomes more natural when embedded in larger chunks of discourse. For lack of space, I leave this to the imagination of the reader.

mentioned above, they also have other uses for which disagreement is not accounted for (these are the same ones that proponents of the challenge, either explicitly or implicitly, have focused on – see Kölbel, 2004; Lasersohn, 2005 etc.). To give just one obvious example, the exchange between my friend and I can be interpreted in a way in which we both use 'delicious' egocentrically. This seems to be the main use of such predicates underlying the challenge, and in such a case the contextualist doesn't get disagreement, since there is no single proposition the interlocutors can have opposite attitudes towards.

However – and this is the catch –, the trend under scrutiny accompanies the showcase of examples in which disagreement can be accounted for with a denial that the intuition of disagreement is present in the case described above. Coupling the claim that disagreements in cases where predicates of taste are used exocentrically, collectively and generically can be accounted for with the claim that there are *no other cases* of disagreement is certainly one way to solve the issue. But, crucially, much hinges on the reasons given for this latter claim. Sometimes the reason given is merely lack of the relevant intuition (e.g., Glanzberg, 2007). It might be impossible to argue about intuitions, but simply leaving things at that is deeply unsatisfactory because it raises the methodological issue of which intuitions to rely on, while flat-out rejecting those who are problematic for one's theory is most probably not a valid methodological practice. In other cases, however, certain considerations are brought to support the claim. For example, both Stojanovic (2007) and Moltmann (2010) raise the issue of what the topic of disagreement could be when interlocutors use predicates of taste egocentrically, and the issue of the point of engaging in such disputes. Others, like Cappelen & Hawthorne (2009), argue by analogy with cases in which the intuition of disagreement is lacking. This is not the place to take up such arguments;[10] what I want to point out here is that this strategy has the potential to meet the challenge from disagreement *if* the arguments to the effect that in the relevant cases disagreement is impossible are sound. Whether this is so, and thus, whether the present contextualist strategy is successful, still remains to be seen.

2. Disagreement as Clash of Non-cognitive Attitudes

Disagreement has been conceived by the relativist as a clash of cognitive (i.e., doxastic) attitudes. But, intuitively, at least, there are other ways of disagreeing. For example, when one person likes something (say, a certain food), while the other doesn't like that same thing, we can felicitously say that they disagree. If so, disagreement may not involve doxastic attitudes or propositions. This suggests that the disagreement in exchanges like the one between my friend and I about marmitako could be said to involve attitudes that are not cognitive in nature towards mere objects, thus explicitly rejecting (**DD**).

This intuitive idea finds support in philosophical literature. In tackling the issue of disagreement in relation to expressivism, Stevenson (1944) draws a distinction between "disagreement in attitude" and "disagreement in belief",[11] the former comprising ways of disagreeing like the one mentioned above. Recent contextualists have borrowed this distinction and have made it part of their answer to the challenge from disagreement by interpreting disagreements like the one between my friend and I about marmitako as a clash of opposite non-cognitive attitudes (Huvenes, 2012, 2014; Lopez de Sa, 2015; Marques, 2015, 2016; Marques & García-Carpintero, 2014; Stojanovic, 2012; Sundell, 2011). How exactly to cash out this disagreement is a choice point for contextualists. Huvenes (2012), for example, is uncommitted with respect to the exact nature of the attitudes involved in disagreement,

[10] I do that in Zeman (2016); my conclusions are negative.

[11] See Ridge (2012) for a detailed discussion of Stevenson's distinction. See also Huvenes (2017) for expressing skepticism that the distinction is ideal.

remaining content with following the expressivist orthodoxy that they are attitudes of *approval* or *disapproval* towards the object the relevant expression is predicated of. More detailed versions are available too. Marques & García-Carpintero (2014), for example, claim that the attitude involved is a special type of *desire*: what they call "desire *de nobis*" – that is, a desire about a collective course of action, based on our evolutionarily developed need to find solutions to coordination problems. This view also brings to the fore the idea that disagreements often arise in cases in which the people involved face a *practical* decision (similar points have been made by Stojanovic, 2012 and Marques, 2015). In another version of this strategy, that of Marques, the attitudes involved in disagreement are *second-order desires*: desires that the interlocutors desire to desire the object that the relevant predicate is predicated of (2016, p. 23).

Appeal to "disagreement in attitude" also marks a point of convergence between contextualism and expressivism – in particular, "hybrid" versions of the latter (for a representative sample of papers, see Fletcher & Ridge, 2014). The distinctive feature of hybrid expressivism is that in uttering sentences containing perspectival expressions a speaker expresses a non-cognitive attitude of sorts, but also asserts a proposition. The non-cognitive attitude expressed and the proposition asserted can be thought of as different levels of meaning/content. And while some authors claim that "[o]ne can think of disagreement [as clash of non-cognitive attitudes] without endorsing expressivism" (Huvenes, 2011, p. 13), many expressivist views are contextualist in holding that the propositions asserted by uttering sentences containing a perspectival expression are perspective-specific.[12] To give only two examples from recent literature on predicates of taste: Buekens (2011) postulates a level of meaning in addition to the perspective-specific proposition asserted, a level of meaning he calls "affective-evaluative" and which consists in the speaker's attitude of approval towards the object a predicate of taste is predicated of. In a similar vein, Gutzmann (2016) distinguishes between a truth-conditional level (the perspective-specific proposition) and a "use-conditional" level, the latter consisting in a "deontic attitude towards what shall count as [P] in the utterance context" (Gutzmann, 2016, p. 45), where "P" is a predicate of taste. In addition to postulating two levels of meaning/content, Gutzmann's view is also explicitly *normative*. It is important to note that part of the motivation for such contextualist-cum-hybrid-expressivist views is precisely answering to the challenge from disagreement.

This contextualist strategy to answer the challenge from disagreement has a lot in its favor. For one thing, it captures the intuitive idea that disagreements need not involve propositions, and that a mere clash of non-cognitive attitudes is sufficient for disagreement. Second, it also captures the equally intuitive ideas that a normative component is sometimes present in uttering sentences like (1) and that many of our disagreements take place against a background in which a practical issue needs solving. However, interpreting *all* disagreements featuring perspectival expressions as practical or as normative might go one step too far. For example, there seem to be scenarios in which no practical issue is at stake: suppose, for instance, that my friend and I are not pressed to find a place to eat, we are not planning to organize a lunch etc., but merely ponder over the culinary virtues of marmitako – perhaps in comparison with other foods.[13] On the other hand, it seems very intuitive that in ordinary

[12] Of course, being a broader view with many variants, hybrid expressivism comprises also views according to which the propositions asserted by uttering sentences containing a perspectival expression are perspective-neutral (e.g., Boisvert, 2008), or even propositions that are semantically incomplete (e.g., Clapp, 2015).

[13] In a certain sense, *all* issues are practical – namely, in the sense of solving theoretical issues for the sake of knowledge itself. I take it though that this sense of 'practical' is not what the proponents of the view discussed have in mind.

scenarios people don't mean to make normative claims, but they simply utter sentences like (1) to express their preferences. When faced with contradictory sentences that signal disagreement, they might just retreat to qualifications like 'delicious to me'. If uttering sentences like (1) would always have a normative component, it is not clear what the point of such a retreat would be.[14] Such cases put pressure on the corresponding variants of the strategy scrutinized. As for the very idea of clash of non-cognitive attitudes itself, note first (as a few contextualists have themselves argued) that it is not entirely clear under which conditions a mere clash of attitudes amounts to a full-fledged disagreement (e.g., Huvenes, 2011, 2017; Marques, 2015). But even assuming this issue is solved, a claim could be made that there are disagreements that are best not interpreted as a clash of non-cognitive attitudes. For example, in scenarios in which both interlocutors use predicates of taste exocentrically (like the one in which my friend and I are planning to take our common friend to lunch for her birthday) and disagreement ensues, the disagreement is arguably doxastic (in the case of my friend and I about the perspective-specific proposition that marmitako is delicious from our common friend's perspective). This points to the need for the contextualist to appeal to doxastic disagreement as well, in addition to appeal to disagreement as clash of non-cognitive attitudes. This, in turn, immediately raises the question of theoretical parsimony: a view that needs to appeal to two notions of disagreement is clearly costlier than a view that appeals to only one such notion.[15] But while things might not be as clear-cut as the contextualist wishes in this respect, the strategy of appealing to "disagreement in attitude" offers the contextualist enough leeway to approach the challenge from disagreement and thus cannot be ignored in further discussions of the issue.

3. Pragmatic Disagreement

The third contextualist strategy to answer the challenge from disagreement tackled in this paper consists in "going pragmatic": that is, to construe the disagreement in exchanges like the one between my friend and I about marmitako as a pragmatic, rather than semantic phenomenon. This in itself doesn't require abandoning (**DD**), but merely relegating it at the pragmatic level. And since there are quite a few phenomena that are traditionally considered to be pragmatic, the variants of this strategy are numerous.

To start with, one way to relegate disagreement at the pragmatic level is to claim that it involves presuppositions. First, it is easy to note that disagreement can target not what has been asserted, but what is presupposed in a given context (see Sundell, 2011; Plunkett & Sundell, 2013 for convincing examples). This suggests that the disagreement between my friend and I could also be interpreted as disagreement over what is presupposed, and not about what is asserted. But which presupposition is it that the two of us disagree about?

A recent view that situates disagreement at the level of presuppositions is Silk's (2016). Silk's basic idea is that, in order to retrieve the semantic content of an utterance, we make certain presuppositions about the values of the required contextual parameters, values that are provided by context. In the case of predicates of taste, for example, when one utters a sentence like (1) one presupposes a certain value of the contextually-given perspective parameter that

[14] One way to go is to say that in retreating to the qualified statement one is limiting the range of individuals the normative component of the utterance is supposed to apply to. Insofar as this makes sense (isn't normativity supposed to be universal?), it opens up the question of what that range of individuals is to begin with.

[15] One way out for the contextualist would be to treat disagreement involving exocentric uses in the same way as that involving egocentric uses: as clash of non-cognitive attitudes. However, two considerations militate against this solution: first, it is difficult to argue that in using a predicate of taste exocentrically, one is *expressing* an attitude in the genuine sense of the term (Buekens (2011) forcefully argues against this claim); second, if the contextualist allows *any* use of a predicate of taste to express an attitude, then the view fully collapses into hybrid expressivism.

makes the semantic content expressed by the sentence to be what it is. In understanding what has been said, the hearer retrieves that value of the perspective parameter and, if she doesn't object, the presupposition is accommodated and that value is added to the common ground. However, the hearer may not agree with the speaker: in this case, the presupposition is *not* accommodated and the relevant value is not added to the common ground. Thus, disagreement is explained by Silk as a refusal from the hearer's part to accommodate the presupposition that the required value of the contextual parameter is as the speaker intends it to be. It is an explanation that involves a standard semantics and ordinary discursive maneuvers, such as presupposition accommodation, and thus distinctively conservative.[16]

An earlier presuppositional view of disagreement – that belonging to López de Sa (2007, 2008, 2015) – has a different take on the issue. While the previous view has construed disagreement as involving opposite presuppositions (or certain discursive maneuvers associated with them), for López de Sa disagreement *becomes possible* when a "presupposition of commonality" is in place. A presupposition of commonality being in place simply means that the interlocutors take themselves to be alike with respect to the relevant contextual parameter – in the case of (1), alike in taste. Disagreement arises, according to López de Sa, precisely when such a presupposition is in place, and is about whether the marmitako is delicious or not *from the common perspective that is presupposed*. However, when such a presupposition is *not* in place, López de Sa denies that disagreement arises and that usually the interlocutors retreat to claims made from their own individual perspectives.[17]

Other presuppositional views trade on different ideas of what the presuppositions that fuel disagreement are. For instance, Zakkou (2015) claims that disagreement becomes possible not when a presupposition of commonality is in place, as López de Sa has it, but rather when a "presupposition of superiority" is. A presupposition of superiority being in place means that the interlocutors take one of them to be better situated than the other with respect to the relevant contextual parameter – in the case of (1), better in taste. Disagreement arises because each interlocutor holds a different presupposition: namely, that *she herself* is the one with the superior taste.[18] Another presuppositional view is that of Parsons (2013). According to Parsons, when one utters a sentence like (1), the presupposition is not that the interlocutors are alike in taste and hence that all interlocutors find marmitako delicious, but that *they aren't*, and hence that one of them finds it delicious and the other doesn't (he calls this "antisupposition"). Disagreement arises because, under the antisupposition mentioned, if what one of the interlocutors says is true, then what the other says is false (given the rules for negation laid out by Parsons, 2013, p. 166).

Another way disagreement can be thought of as pragmatic is to see it as arising at the level of implicatures. Detailed discussion of the suggestion to construe disagreements like the one between my friend and I about marmitako as arising at the level of implicatures can be found in many places: Huvenes (2011), Sundell (2011), Plunkett & Sundell (2013) etc. Here, however, I want to point towards a further point of convergence between contextualism and hybrid expressivism. Earlier hybrid expressivist views such as Barker's (2000) or Finaly's (2005) have

[16] For a similar view about accounting for disagreement in terms of discursive maneuvers, but which doesn't rely on presuppositions, see Björnsson (2015).

[17] More recently, López de Sa (2015) has clarified his position by holding that in such cases disagreement arises, but cannot be expressed by using the sentences at stake. In addition, he also holds that the existing disagreement should be cashed out as clash of non-cognitive attitudes.

[18] Zakkou's view is in fact more general, in that she doesn't put weight on the distinction between presupposition and implicature (she talks about "pragmatically conveyed" content). She can thus figure in the next variant of the pragmatic strategy as well (see below).

it that the evaluative component of a sentence containing moral terms is expressed by way of an implicature. As before, differences in such views come from the type of the implicature postulated (merely evaluative or normative etc.), but in principle most of these views can be applied to a wider range of perspectival expressions. Marques (2016), for example, claims that her hybrid expressivist view about aesthetic predicates can be cast in terms of implicatures.[19]
Finally, it has been proposed that pragmatic disagreement be cashed out in terms of disagreement about the meaning of words ("metalinguistic disagreement") or about the context interlocutors are in ("metacontextual disagreement"). To name only a few works, Sundell (2011), Plunkett & Sundell (2013) and Plunkett (2015) contain a significant number of exchanges in which the disagreement the interlocutors have can be rightfully construed in such terms. As regarding the first type of disagreement, one way in which it can be cashed out is by having the interlocutors argue about the (Kaplanian) *character* of the relevant expressions (e.g., about the term 'athlete' - see Sundell, 2011). As regarding the second, the standard case here is taken to be Barker's (2002) example involving the gradable adjective 'tall': there are situations, Barker claims, in which what the interlocutors do is argue either about the comparison class or about the threshold that determines whether a person counts as tall – that is, about features of the context and not about, say, the actual height of the person. The suggestion then is that disagreement in exchanges like the one between my friend and I about marmitako can be seen as metalinguistic or metacontextual. A normative version of this strategy is possible too: instead of claiming that the disagreement is about what words mean or what context the interlocutors are in, the authors cited claim that the disagreement is about what words *should* mean or (perhaps more controversially) about what context the interlocutors *should* be in.
All the pragmatic strategies mentioned point towards interesting aspects of our usage of perspectival expressions like predicates of taste. Given the pervasive presence of pragmatic effects in our language use, it would be quite surprising if disagreement were limited only to semantics. However, the pragmatic strategy faces several challenges. For one, it is notoriously difficult to pry apart purely semantic phenomena from pragmatic ones. Both in the case of presuppositions and implicatures there is still a vivid debate surrounding the viability of the classical tests for such phenomena.[20] Further, it is questionable whether rendering disagreement in such a way does justice to all the cases of intuitive disagreement (one relevant question being what happens when the presuppositions or implicatures postulated are *not* in place[21]). As for metalinguistic and metacontextual disagreements, while it is hard to deny that such disagreements exist, claiming that *all* disagreements involving perspectival expressions are of this kind might go one step too far. For example, it seems very intuitive that in ordinary scenarios people disagree not about words or contexts, but about the very topic of their discussion; in the exchange between my friend and I, it is very intuitive to think that the disagreement is over whether marmitako is, in fact, delicious,[22] and not over the word 'delicious' itself. Second, as several authors have pointed out, talk and belief about language and talk and belief about the world can *coexist*. This has important methodological implications: as Lassiter (2011) writes about what he calls "mixed uses" of predicates of taste (both ordinary and metalinguistic), "[t]his aspect is important because beliefs about the world and beliefs about language obviously do interact: we would not want a theory that

[19] To be more precise, Marques claims that her view can be cast either in terms of presuppositions or in terms of generalized conversational implicatures. I thank a referee for drawing my attention to this.

[20] See, for example, Åkerman (2015) for a discussion of the cancellation test for conversational implicatures.

[21] See, for example, Marques & García-Carpintero (2014) for making this point against López de Sa.

[22] Whatever that fact might amount to. I leave metaphysical considerations about deliciousness aside in this paper.

separates them completely" (132, fn. 1). Applied to the case of disagreement, this observation amounts to the claim that disagreements about the world and disagreement about language can coexist. This, in turn, means that treating disagreement involving "mixed uses" as solely metalinguistic or metacontextual would leave one crucial aspect of their use (that is, the one concerning the world) unexplained. Unless an argument is given to rule out the possibility of "mixed uses" being part of ordinary exchanges like the one between my friend and I about mamitako,[23] this strategy is incomplete. Thus, both the existence of disagreements that are intuitively non-metalinguistic or non-metacontextual and that of "mixed" cases of disagreement make a strong case against the claim that the strategy under scrutiny is a satisfactory answer to the challenge from disagreement.[24] That being said, this strategy, as the one before, points to important aspects of our use of predicates of taste and thus has to be carefully considered in the next phase of the debate.

***

This completes the overview of recent contextualist strategies responding to the challenge from disagreement and their main variants. Needless to say, other strategies/variants are possible; also, some authors appeal to more than one strategy to explain the whole range of data (e.g., López de Sa, 2015). Whether or not the strategies presented, in themselves or in combination with others, are ultimately successful in dealing with the challenge from disagreement is something that needs to be further inquired into. In any case, they have all significantly advanced the debate surrounding disagreement in semantics and are thus well worth engaging with in the future.

REFERENCES

Åkerman, J. (2015). Infelicitous cancellation: the explicit cancellability test for conversational implicature revisited. *Australasian Journal of Philosophy*, 93, 465-474.

Barker, S.J. (2000). Is Value Content a Component of Conventional Implicature?. *Analysis*, 60(267), 268-279.

Barker, C. (2002). The Dynamics of Vagueness. *Linguistics and Philosophy*, 25, 1-36.

Barker, C. (2013). Negotiating Taste. *Inquiry*, 56(2-3), 240-257.

Björnsson, G. (2015). Disagreement, correctness, and the evidence for metaethical absolutism. *Oxford Studies in Metaethics*, 10, 160-187.

Boisvert, D. (2008). Expressive-Assertivism. *Pacific Philosophical Quarterly*, 89, 169-203.

Buekens. F. (2011). Faultless Disagreement, Assertions and the Affective-Expressive Dimension of Judgments of Taste. *Philosophia*, 39, 637-655.

Cappelen, H., & Hawthorne, J. (2009). *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Clapp, L. (2015). A Non-Alethic Approach to Faultless Disagreement. *Dialectica*, 69(4), 517-550.

Dreier, J. (1990). Internalism and Speaker Relativism. *Ethics*, 101, 6-26.

Finlay, S. (2005). Value and Implicature. *Philosophers' Imprint*, 5(4), 1-20.

Fletcher, G., & Ridge, M. (Eds.) (2014). *Having It Both Ways. Hybrid Theories and Modern Metaethics*. Oxford: Oxford University Press.

[23] Barker himself claims that such uses of vague terms (which, according to him, include predicates of taste) "typically, perhaps normally" inform "both about the facts in world, and about the prevailing standards" (2013, p. 243).

[24] To be fair, the proponents of the metalingusitc/metacontextual strategy are very cautious and refrain from making completely general claims – see, for example, the hedges used when stating the strategy in Plunkett & Sundell (2013, pp. 4, 25); Barker (2013, p. 242); Ludlow (2014, p. 62) etc.

Glanzberg, M. (2007). Context, Content, and Relativism. *Philosophical Studies*, 136(1), 1-29.

Gutzmann, D. (2016). If Expressivism is Fun, Go for It! In C. Meier & J. van Wijnbergen-Huitink (Eds.), *Subjective Meaning: Alternatives to Relativism*. Berlin/Boston: Mouton de Gruyter, 21-46.

Harman, G. (1996). Moral Relativism. In G. Harman & J.J. Thomson, *Moral Relativism and Moral Objectivity*. Oxford: Blackwell, 1-64.

Huvenes, T.T. (2012). Varieties of Disagreement and Predicates of Taste. *Australasian Journal of Philosophy*, 90(1), 167-181.

Huvenes, T.T. (2014). Disagreement without Error. *Erkenntnis*, 79 (1, Supplement), 143-154.

Huvenes, T.T. (2017). On Disagreement. In J. Ichikawa (Ed.), *Routledge Handbook of Epistemic Contextualism*. New York: Routledge, 272-281.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry & H. Wettstein (Eds.), *Themes from Kaplan*. Oxford: Oxford University Press, 481-563.

Khoo, J. (2017). The disagreement challenge to contextualism. In J. Ichikawa (Ed.), *Routledge Handbook of Epistemic Contextualism*. New York: Routledge, 257-271.

Kölbel, M. (2004). Indexical Relativism vs Genuine Relativism. *International Journal of Philosophical Studies*, 12(3), 297-313.

Lasersohn, P. (2005). Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy*, 28(6), 643-686.

Lasersohn, P. (2016). *Subjectivity and Perspective in Truth-Theoretic Semantics*. Oxford: Oxford University Press.

Lassiter, D. (2011). Vagueness as Probabilistic Linguistic Knowledge. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *Vagueness in Communication*. Berlin: Springer, 127-150.

López de Sa, D. (2007). The many relativisms and the question of disagreement. *International Journal of Philosophical Studies*, 15(2), 269-279.

López de Sa, D. (2008). Presuppositions of Commonality: An Indexical Relativist Account of Disagreement. In M. García-Carpintero & M. Kölbel (Eds.), *Relative Truth*. Oxford: Oxford University Press, 279-310.

López de Sa, D. (2015). Expressing Disagreement. *Erkenntnis*, 80, 153-165.

Ludlow, P. (2014). *Living Words. Meaning Underdetermination and the Dynamic Lexicon*. Oxford: Oxford University Press.

MacFarlane, J. (2007). Relativism and Disagreement. *Philosophical Studies*, 132(1), 17-31.

MacFarlane, J. (2014). *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.

Marques, T. (2014). Doxastic Disagreement. *Erkenntnis*, 79(1, Supplement), 121-142.

Marques, T. (2015). Disagreeing in Context. *Frontiers in Psychology*, 6, 1-12.

Marques, T. (2016). Aesthetic predicates: a hybrid dispositional account. *Inquiry*, 59, 723-751;

Marques, T., & García-Carpintero, M. (2014). Disagreement about Taste: Commonality Presuppositions and Coordination. *Australasian Journal of Philosophy*, 92(4), 701-723.

Moltmann, F. (2010). Relative Truth and the First Person. *Philosophical Studies*, 150(2), 187-220.

Parsons, J. (2013). Presupposition, Disagreement, and Predicates of Taste. *Proceedings of the Aristotelian Society*, 113, 163-173.

Plunkett, D. (2015). Which Concepts Should We Use?: Metalinguistic Negotiations and The Methodology of Philosophy. *Inquiry*, 58(7-8), 828-874.

Plunkett, D., & Sundell, T. (2013). Disagreement and the Semantics of Normative and Evaluative Terms. *Philosophers' Imprint*, 13(23), 1-37.

Ridge, M. (2012). Disagreement. *Philosophy and Phenomenological Research*, LXXXVI (1), 41-63.

Schaffer, J. (2011). Perspective in taste predicates and epistemic modals. In A. Egan & B. Weatherson (Eds.), *Epistemic Modality*. Oxford: Oxford University Press, 179-226.

Silk, A. (2016). *Discourse Contextualism. A Framework for Contextualist Semantics and Pragmatics*. Oxford: Oxford University Press.
Stevenson, C.L. (1944). *Ethics and Language*. New Haven: Yale University Press.
Stojanovic, I. (2007). Talking about Taste: Disagreement, Implicit Arguments, and Relative Truth. *Linguistics and Philosophy*, 30(6), 691-706.
Stojanovic, I. (2012). Emotional Disagreement. *Dialogue*, 51(1), 99-117.
Sundell, T. (2011). Disagreements about Taste. *Philosophical Studies*, 155(2), 267-288.
Zakkou, J. (2015). *Tasty Contextualism. A Superiority Approach to the Phenomenon of Faultless Disagreement*. Ph.D. Thesis, Humboldt University of Berlin.
Zeman, D. (2016). The Many Uses of Predicates of Taste and the Challenge from Disagreement. *Studies in Logic, Grammar and Rhetoric*, 46(1), 79-101.

DIOGO SANTOS
LanCog, University of Lisbon
diogoferreirasantos@campus.ul.pt

HOW TO DISPEL THE ASYMMETRY CONCERNING RETRACTION[1]

abstract

MacFarlane (2014) advocates a radical form of semantic relativism. He argues that his proposal complies with the norms governing our assertion practices in various areas of discourse. These practices also include norms regarding the conditions in which it is inappropriate not to retract an assertion. Ferrari & Zeman (2014) identify an asymmetry concerning retractions in two relevant areas of discourse and argue that assessment-sensitivity needs to be supplemented with further theoretical tools to explain it. I dispel the asymmetry and conclude that assessment-sensitivity needs no supplementation to account for it.

keywords

retraction, assessment-sensitivity, asymmetry concerning retraction

Phenomenology and Mind, n. 12 - 2017, pp. 74-82
DOI: 10.13128/Phe_Mi-21107
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

HOW TO DISPEL THE ASYMMETRY CONCERNING RETRACTION

1. Introduction

In his book *Assessment Sensitivity* (2014), John MacFarlane advocates a radical form of semantic relativism. In defence of his proposal, he argues that assessment-sensitivity (henceforth AS) complies with the norms governing our assertion practices in various areas of discourse, such as taste, morality, epistemic modality or deontic discourse. These practices include norms regarding the conditions in which it is appropriate to make an assertion, but also regarding the conditions in which it is appropriate to *retract*, or take back, an assertion.

Ferrari & Zeman (2014) identify an asymmetry concerning moral and personal taste retractions. They argue that there are data supporting the existence of the asymmetry, and that AS needs to be supplemented with further theoretical tools to explain it. I argue that there is no relevant asymmetry even if we accept that the data provide reason in favor of a disparity between retracting moral judgements and retracting personal taste judgements. I conclude, thus, that AS needs not be supplemented with anything else because of it.

The plan is as follows: in Section 2, I briefly describe AS's framework and the constitutive norms of assertion that the view endorses. In Section 3, I outline MacFarlane's characterization of retraction and the practical consequences of endorsing the norm of retraction for assertions. Section 4 describes the apparent asymmetry and argues that supplementing AS with more theoretical tools does not do the needed work. In section 5, I argue that there is no relevant asymmetry related with moral and personal taste retractions and that the discrepancy that the data show does not affect AS. I conclude by summarizing my view and saying something more about the debate concerning retraction.

2. Assessment-Sensitivity

AS's central claim is that certain sentences are assessment-sensitive. Candidates for sentences that possess this feature can be found in discourse about matters of personal taste, moral discourse, aesthetic discourse, and so on. To say that a sentence is assessment-sensitive is to say that its truth depends not only on features of the context of utterance but also on features of the context from which its use is assessed (MacFarlane 2005; 2014). According to AS, propositions are assessed from contexts of assessment, which are not fixed by any feature of

[1] Acknowledgements: Thanks to Adriana Silva Graça, Elia Zardini, Hugo Luzio, José Mestre, Laura Delgado, Luís Veríssimo, Max Kölbel, Michele Palmira, Pedro Dinis, Ricardo Miguel, Ricardo Santos, Sanna Hirvonen and Teresa Marques for helpful comments and discussion. I am also thankful to two anonymous referees for comments and suggestions that greatly improved this paper.
Work for this paper was supported by a doctoral fellowship (UL/BD/453/2015) awarded by the University of Lisbon.

the context of utterance, but by the assessor. Since the context of assessment is the context of the assessor, the proposal is not about relativizing truth to an additional parameter provided by some feature of the context of utterance.[2]

MacFarlane believes that we can make sense of assessment-sensitive truth by considering its role in the norms that govern our assertion practices. Part of his defence of AS therefore consists in spelling out these norms, showing that they do presuppose a relativistic framework like the one he proposes. According to AS, assertions are (partially) constituted by a truth norm. The norm is formulated as follows:

> **Reflexive Truth Rule (RTR).** *An agent is permitted to assert that p at context c_1 only if p is true as used at c_1 and assessed from c_1.*[3]

RTR forbids the performance of assertions that are false as used and assessed from the context in and from which they are performed. Nonetheless, the rule does not commit one to claiming that it is wrong to assert a false proposition. If other norms are in play, the rule can be overridden.[4]

Asserting is also "[p]utting a sentence forward in the public arena as true"; this implies that the sentence becomes "available for others to use in making further assertions" (Brandom, 1998, p. 170). If this is so, then there should also be a rule whereby, given certain conditions, the agent is required to remove the sentence she put forward from the "public arena". The rule may be stated as follows:

> **Retraction Rule (RR).** *An agent in context c_2 is required to retract an (unretracted) assertion of p made at c_1 if p is not true as used at c_1 and assessed from c_2.*[5]

According to RR, an agent is only required to retract her assertion that *p* when, as assessed from the context of retraction, *p* as used at c_1 is not true. RR captures the agent's responsibility for putting forth a sentence as true by requiring that she should retract once the sentence is assessed as false.[6]

[2] "It is important that the context of assessment is not fixed in any way by facts about the context of use, including the speaker's intentions; there is no 'correct' context from which to assess a particular speech act" (MacFarlane, 2014, pp. 61-62).

[3] MacFarlane (2014, p. 102).

[4] For a discussion of what it means for constitutive norms to be overridden, see e.g. García-Carpintero (2015).

[5] MacFarlane (2014, p. 108).

[6] RR is paramount for AS, for it is what pragmatically differentiates it from contextualist theories. As MacFarlane states:

> The basic thought is that the pragmatic difference between R[elativism] and C[ontextualism] manifests itself in norms for the retraction of assertions rather than norms for the making of assertions. R[elativism] predicts that an assertion of p at c1 ought to be retracted by the asserter in c3, while C[ontextualism] predicts that it need not be retracted (2014, p. 108).

To clearly see this difference consider the following example by MacFarlane:

> Let c1 be a context centered on ten-year-old Joey, who loves fish sticks. According to both R and C, the proposition that fish sticks are tasty is true as used at and assessed from c1. So the Reflexive Truth Rule tells us that Joey is permitted to assert that fish sticks are tasty. Let us suppose that he does. Now consider another context c2 centered on Joey, ten years later. As a twenty-year old, Joey no longer likes the taste of fish sticks. Here R and C diverge. According to R, the proposition that fish sticks are tasty is false as used at c1 and assessed from c2, so by the Retraction Rule, Joey is now required to retract his earlier assertion. According to C, by contrast, the proposition that fish sticks are tasty is true as used at c1 and assessed from c2, and Joey need not retract (2014, p. 109).

3. Retracting Assertions

MacFarlane (2014, p. 108) characterizes retraction as the speech act of taking another speech act back. We can make sense of this rough characterization by making it clear what each component of the phrase "a speech act of taking another speech act back" is intended to convey. First, retraction is a speech act – that is, retraction is performed by saying that one is doing so. The usual English expressions for retracting involve the following: 'I retract that', 'Scratch that', 'Delete/Erase that', 'I was mistaken'. Second, "to take another speech act back" is best understood as targeting the commitments the agent made while performing the speech act she seeks to retract.

This paper is especially concerned with retraction of assertions. When an assertion is retracted, the retractor is no longer expected to stand by it. The commitments undertook when asserting – for instance, to asserting truthfully or putting forward a sentence as true in the "public arena" – are no longer in play. The retractor is no longer subject to the norms governing the retracted assertion. Consider the example below.

Example 1
Paula$_9$: My mother's fish soup is not tasty.
Paula$_{19}$: My mother's fish soup is so tasty.

Paula$_{19}$ is obliged to retract her previous assertion. If she were not to retract, she would be violating RR. To do so would be incorrect, because Paula$_{19}$ would allow for a sentence that is false – relatively to her present standard of taste – to be available as true for others to use. Given that the contexts of assessment in RTR and in RR differ, it is possible that speakers have a duty to retract an assertion that (until then) violated no norm. "To say that one was wrong in claiming that *p* is not to say that one was wrong to claim that *p*. Sometimes it is right to make a claim that turns out to have been wrong (false)" (MacFarlane, 2011, p. 148). Keep in mind that MacFarlane regards the act of assertion as involving the speaker's commitment to the truth of the propositional content expressed. Thus, it may be the case that one must retract a correct assertion, i.e. one that violated no norm. The idea here is that the agent ought to update what she has put forth if it turns out to be false.

4. The (Apparent) Asymmetry Concerning Retraction

Ferrari & Zeman (2014) identify a (purported) asymmetry regarding retraction in two distinct areas of discourse. On the one hand, the asymmetry has to do with the fact that we generally do not expect personal taste retractors to blame themselves for the gustatory standard they held at the time of utterance. On the other hand, it concerns the fact that we generally *do* expect moral retractors to blame themselves for the moral standard they previously held. To put it more plainly, personal taste retractors usually do not implicitly nor explicitly convey that they regard the standard they held in the past as *the wrong one to have*; by contrast, moral retractors usually implicitly or explicitly convey that they take the moral standard they adopted as *the wrong one to have*.

In order to identify the phenomenon at issue, an example may help.

Example 2
Paula$_{t1}$: Bullfighting is not wrong.
Albert$_{t2}$: You said that bullfighting was not wrong.
Paula$_{t2}$: I was mistaken. Bullfighting is despicable.

According to AS, in retracting her previous assertion at t_2, Paula is not necessarily admitting that she was somehow blameworthy for making that assertion. Still, it is a perfectly fine retraction if she retracts while conveying that she was in some sense wrong to assert what she did – AS allows for such a possibility, since RR is perfectly compatible with it. However, what may be damaging for AS is that there is something more than compliance with RR in the above example: agents generally think that Paula$_{t2}$ should implicitly or explicitly blame herself for having held the standard she did at t_1. Arguably, we are inclined to consider that the utterance "Bullfighting is despicable" implicitly conveys that idea i.e. that Paula$_{t2}$ finds any moral standard from which the proposition *Bullfighting is not morally incorrect* is true to be the wrong one to have. If "in many cases we would expect [Paula$_{t2}$] to feel ashamed for having held such a judgement" (Ferrari & Zeman, 2014, p. 92), then it seems natural to claim that something else is happening besides the retraction of the assertion itself, namely a disapproval of the standard Paula held at t_1.

MacFarlane (2011, p. 148) distinguishes between making an assertion that violates RTR and retracting an assertion that is no longer true, as assessed from the context of retraction. In a way, retraction signals that the retractor recognizes that "what she said was wrong". But, this fails as an explanation for what seems to be happening in the bullfighting example. The retractor is expected not just to recognize that "what she said was wrong", but also that *she should not have said it in the first place*. Thus, the theory appears to lack the appropriate tools to account for what is happening with (presumably) most moral retractions.

Notice, however, that the same does not seem to happen for retractions about matters of personal taste.

Example 3
Hillary$_{t1}$: Barnacles are tasty.
Hillary$_{t2}$: I was mistaken. Barnacles are awful.

Hillary$_{t2}$'s retraction does not implicitly (or explicitly) convey that she finds each gustatory standard from which the proposition *Barnacles are tasty* is true to be the wrong one to have. Arguably, the follow-up "Barnacles are awful" is not used to convey it. More importantly, we would not expect Hillary to convey such a thing.

AS neatly accounts for the case of personal taste retractions. The retractor is only expected to recognize that "what she said was wrong", and not that *she should not have said it in the first place*. Thus, nothing other than the compliance with RR is going on.

A natural solution to the mystery is to propose some new theoretical device compatible with AS's framework that allows the theory to explain how the asymmetry occurs.[7] For example, one may claim that in some cases retractors "will access the circumstances in which the assertion was made in evaluating a previous assertion and thus will attempt to retrieve the specific value of the relevant parameter and assess it" (Ferrari & Zeman, 2014, p. 93). This proposal accounts for the possibility that the relevant parameter is assessed as the wrong one. It therefore captures how moral retractors generally assess the assertion ("and the specific value of the relevant parameter") they are retracting. This also reduces the asymmetry to a simple difference between how frequently the *wrong standard to have* comes about from moral assessments and how frequently it comes about from personal taste assessments.

Additionally, supplementing AS in a way that allows agents not only to assess propositions from their context of assessment, but also to assess parameters of the contexts of assessment they previously held, explains why we would expect some retractions to be accompanied by retractors being overly critical of their previous view on the matter.

None of the data implies that RR does not hold for both personal taste and moral discourses.

[7] This is what Ferrari & Zeman (2014) do.

Agents are still expected to retract under the conditions that are determined by the rule. No reason was presented to maintain that, if an agent's moral standards change from t_1 to t_2, and relative to the moral standard held at t_2 what was asserted at t_1 is false, then she is under no obligation to retract.[8] Quite the contrary, in all the examples provided, agents uphold the rule. Thus, the data presented impinge in no way on RR.

Ferrari and Zeman do not assume that the asymmetry is incompatible with RR. Nonetheless, they feel that AS should explain it. In fact, assuming that the asymmetry as described occurs and that RR is correct, that is the only reasonable worry. However, given the theoretical improvement described above, the worry is yet to be met, for it can be restated as follows: (1) why would we expect that only some retractions (confined to a specific area of discourse) should be accompanied by a criticism of the retractor's previous view on the matter? Answering (1) will shed light on how to interpret the data – and, as we will see, supplementing contexts of assessment with the feature introduced by Ferrari & Zeman provides no help.

5. The Asymmetry Dispelled

In this section, I will directly address (1). The asymmetry is somewhat related with distinct areas of discourse. Thus, it is reasonable to think that, only by investigating the relevant distinct features in each area of discourse, one can formulate an appropriate explanation for the asymmetry. To do so, consider the following examples that should explicitly illustrate the asymmetry.

Example 4
Bill$_{t1}$: Abortion is murder.
Bill$_{t2}$: I was wrong. No one should think that abortion is murder.

Bill$_{t2}$ retracts the assertion made at *t1* while being critical of his previous stance on abortion. In such a case, the retractor not only recognizes that his previous assertion is, from the new context of assessment, false, but he also indicates that his previous stance on the matter is the wrong one to have.

Example 5
Gwen$_{t1}$: Monkfish isn't tasty.
Gwen$_{t2}$: I was wrong. But it's fine if you think that monkfish isn't tasty.

Gwen$_{t2}$ retracted her previous assertion about monkfish's taste. However, and contrary to Bill's case, she is willing to retract and indicate that the previous stance she held about monkfish is perfectly fine. If you find Gwen's retraction unnatural, imagine that she is talking to her six year old daughter. So, she would be conceding that it is alright for a six year old to hold that monkfish is not a tasty fish.

Examples 4 and 5 purport to provide good reasons for us to endorse the claim that there is an asymmetry related with moral and personal taste retractions. But this is yet insufficient data to support the conclusion that the asymmetry is as we have been describing it thus far.

For instance, one can come up with examples where the moral retractor is not very critical of her previous moral standards and where the personal taste retractor is highly critical of her previous gustatory standards. These cases are interesting, because they provide us with some

[8] Marques (2015) argues that RR makes the wrong predictions: "to retract a past assertion after a change of perspective (standard of taste) is neither irrational nor insincere" (p. 12).

insights about what the relevant features of highly critical retractions are and thus can help us understand what the apparent asymmetry is all about.

Example 6
Chris$_{t1}$: It is wrong to lie.
Ruth$_{t2}$: You said that it was wrong to lie.
Chris$_{t2}$: I was mistaken. Lying is morally permissible. But there is nothing wrong with thinking that lying is morally wrong.

Chris$_{t1}$ asserted that it was wrong to lie. Chris$_{t2}$ thinks that lying is morally permissible but, since (arguably) no terrible consequence would come about from holding the contradictory view, it is perfectly reasonable for Chris$_{t2}$ to state that his former self did nothing wrong in asserting what he did.

Example 7
Lucy$_{t1}$: Gordon Ramsay's Beef Wellington is not tasty.
Peter$_{t2}$: You said that Gordon Ramsay's dish was not tasty.
Lucy$_{t2}$: I was mistaken. The Wellington is delicious. Whoever thinks that Gordon Ramsay's Beef Wellington isn't tasty is wrong.

In personal taste discourse, there are cases where retraction is accompanied by the speaker being very critical of her former self. Lucy$_{t2}$ is very critical about the gustatory standards she previously held – she cannot believe that she doubted Gordon Ramsay's abilities as a chef. The cases where the retractor is critical of her former self display a common feature: they are clear-cut cases. An agent finds abortion (or bullfighting) either wrong or permissible, Ramsay's Beef Wellington is either tasty or not – one is not usually conflicted about such issues. Less clear-cut cases do not usually involve the retractor being critical of her former self. Distinguishing clear-cut cases from less clear-cut cases allows us to distinguish the examples where we would expect retractors to be critical of their previous views from examples where we would not expect them to be so. Moreover, there is no longer an asymmetry concerning retraction in two distinct areas of discourse. Retraction works in the same way in both areas of discourse – it is just that clear-cut cases are less common in personal taste discourse when compared with moral discourse. To account for this discrepancy one only needs to tell some sociological story about what makes clear-cut cases more uncommon in discourse about taste. Here is an attempt.

Arguably, clear-cut cases are more common in moral discourse because moral issues are usually interpreted as more universalizable than issues concerning personal taste, which seem to depend much more on the idiosyncrasies of individuals. This is an empirical claim about how speakers intuitively view these two areas of discourse. The claim explains that speakers' opinions about taste do not usually involve strong views on gustatory standards – agents do not feel compelled to do so because such matters usually lack universalizability. On the other hand, given that agents think that moral properties are usually universalizable, opinions on such matters usually are accompanied by strong views about moral standards. Also, people seem to give more importance to moral matters than to matters about personal taste – which is consistent with the idea that we have stronger moral opinions than opinions about personal taste.

Whatever the precise explanation is, it is largely irrelevant to our purposes. What is important is that the asymmetry is dispelled. Even granting that the empirical data support the idea that it is more common for moral retractors to be critical of their former selves than it is for personal taste retractors, it does not follow that we get an asymmetry between moral and personal taste retractions. There actually is none. There is a discrepancy concerning the amount of clear-cut cases in personal taste issues and in moral issues. If so, then it is not the case that AS needs to account for it. The reason is very straightforward: speakers can disapprove their previous standards of assessment in both areas of discourse. This implies that there is no asymmetry and no explanatory gap to be filled. There is quite probably a difference in the amount of clear-cut cases in the moral area of discourse and in the taste area of discourse. But this is to be explained sociologically (i.e. by distinguishing between claims that people find universalizable and claims they do not) and that is beyond our present purposes. The asymmetry is now dispelled. We can thus claim that, because of it, AS requires no theoretical improvements.

6. Conclusion

The Retraction Rule is a crucial component for Assessment Sensitivity. Recently, new data have come to light that undermine the connection between retraction and the falsity of what was originally asserted (e.g. Knobe & Yalcin, 2014). This entails that the rule makes the wrong predictions about when one is required to retract an assertion. In this paper, I assumed that RR makes the correct predictions. Nonetheless, one might claim that such data undermining RR also undermine Ferrari and Zeman's description of the asymmetry. If retractions are not as closely connected with the falsity of what was originally asserted, then, in some of the examples they provide to support the asymmetry, retracting is arguably not mandatory. If so, the claim that there is an asymmetry between mandatory retractions across distinct areas of evaluative discourse is undermined. Given that we have operated on the assumption that RR holds, these considerations will have to be discussed elsewhere.

Still, it is indisputable that retractions *do* sometimes happen, even for the skeptic about RR. The claim that agents behave differently when moral retractions and personal taste retractions are concerned is then of interest, because it tells us that retraction may be more than just admitting the falsity of one's previous assertion; it may involve taking a stance on the values adopted when the assertion was performed – e.g. their holding universally or merely expressing idiosyncrasies, as suggested by the sociological explanation I have proposed.

Note, however, that the skeptic cannot question RR by appealing to the putative asymmetry pointed out by Ferrari and Zeman. For, even granting that there is enough data to support the claim that moral retractions are generally accompanied by retractors being overly critical about their own previous view on the matter while personal taste retractions are not, this does not pose explanatory demands on RR. The assumption (on which we have operated so far) that agents are required to retract under the conditions that the rule determines is not questioned by the putative asymmetry.

Hopefully, I have shown that the asymmetry concerning retraction across two distinct areas of discourse is nothing but apparent. The most we get from the data is a discrepancy on the amount of clear-cut issues in moral discourse and in personal taste discourse, but that is not relevantly related with retraction and, thus, it does not affect AS. The discrepancy can be accounted for by providing a sociological explanation for it, independently of the semantic and pragmatic theories we choose to endorse. In the light of such considerations, supplementing AS with new theoretical tools would be an overreaction to the empirical data on the matter.

REFERENCES

Brandom, R. (1998). *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.

Ferrari, F., & Zeman, D. (2014). Radical relativism, retraction and "being at fault". In S. Caputo, M. Dell'Utri, & F. Bacchini (Eds.), *New frontiers in truth*. Newcastle: Cambridge Scholar, 80-102.

García-Carpintero, M. (2015). Contexts as shared commitments. *Frontiers in Psychology*, 6. DOI: 10.3389/fpsyg.2015.01932.
Knobe, J. & Yalcin, S. (2014). Epistemic modals and context: Experimental data. *Semantics and Pragmatics*, 7(10), 1-21.
MacFarlane, J. (2005). Making sense of relative truth. *Proceedings of the Aristotelian Society*, 105(3), 321-339.
MacFarlane, J. (2011). What is assertion? In J. Brown & H. Cappelen (Eds.), *Assertion: New philosophical essays*. Oxford: Oxford University Press, 79-96.
MacFarlane, J. (2014). *Assessment sensitivity: Relative truth and its applications*. Oxford: Oxford University Press.
Marques, T. (2015). Retractions. *Synthese*, 1-25.

SIMONE CARRUS
Vita-Salute San Raffaele University
carrussimone@gmail.com

SLURS: AT-ISSUENESS AND SEMANTIC NORMATIVITY

abstract

In the first part of the article, we present the main approaches to analyze slurs' content and we investigate the interaction between an assertion containing a slur and a denial ('It's not true that P' / P is false') showing to what extent a "neutral counterpart account" works better than a "dual account". Additionally, the analysis offers the opportunity to discuss the usefulness of the notion of "at-issueness" for a debate on the lexical semantics of slurs. In the second part, we use our apparatus to analyze a real case of non-standard use of 'frocio' ('faggot'). Our conclusion is that even if a family resemblance conception of category membership could account for these uses, it cannot account for the related semantic normativity problem.

keywords

ambiguity, at-issueness, lexical semantics, semantic normativity, slurs

Phenomenology and Mind, n. 12 - 2017, pp. 84-97
DOI: 10.13128/Phe_Mi-21108
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction: The Derogatory Content of Slurs

One of the main points of interest about slurs (‘nigger’, ‘faggot’, ‘wop’, ‘kike’, etc.) is the linguistic nature of their derogatory content. At the state of the art, we cannot find a completely uncontroversial account: among other things, we are still wondering if their Derogatory Force (Hom, 2008) plays a truth-conditional role. In the literature, we can find at least two families of theories. On the one hand, we have the truth-conditional accounts: the proposal by Hom (2008) is probably the most representative of this field. It has been followed by the account proposed in Hom & May (2013) and, with some significant variations, by the work of Croom (2011, 2014, 2015). On the other hand, we have the non-truth-conditional accounts: starting from Macià (2002), then with Potts (2007a, 2007b), Schlenker (2007), Williamson (2009), Jeshion (2013b), etc., these scholars defend the idea according to which the derogation has nothing to do with the truth-conditional and/or at-issue content.[1]

Most researchers shares the idea according to which every slur has a neutral counterpart (NC): ‘nigger’ - ‘Afro-American’, ‘faggot’ - ‘homosexual’, ‘kike’ - ‘Jew’, ‘wop’ - ‘Italian’, etc. However, on this point we find the first substantial difference between a truth-conditional approach and a non-truth-conditional one: according to the former, slurs and the corresponding NCs would denote different sets of entities; according to the latter, slurs and NCs would be equivalent concerning the denotation.

Let us consider an example of assertion containing a slur.

[1a] Antony: “Mark is a faggot”.

According to Hom & May (2013), the derogatory content of ‘faggot’ has a double articulation: it characterizes the truth-conditional component, and then is conveyed by a conversational implicature. For this reason, henceforth I will talk of a *dual account*. According to the non-truth-conditional accounts, the truth-conditional content of ‘faggot’ is the same as that of the

[1] According to Potts (2015, p. 169), the “at-issue” content corresponds to what Frege (1892/1980) calls the ‘sense’ and what Grice (1975) calls ‘what is said’. Potts suggests that talking about “truth-conditional content” is confusing, because even non-at-issue (traditionally “pragmatic”) contents like presuppositions and implicatures can generally affect the truth-conditions of an utterance. However, given that several scholars do not adopt Potts’s taxonomy of meanings, in this paper we will talk about truth-conditional contents and truth-conditional theories. Furthermore, in paragraph §2.3, we suggest that if the notion of “at-issueness” concerns the relation between a proposition and the “Question Under Discussion” in a discourse, its relevance at the lexical level may seem controversial.

NC (‘homosexual’) and the derogatory content is *pragmatic* and/or not *at-issue*. Henceforth, I will talk of a *neutral-counterpart account* (NC account).[2]
In table 1, the analysis of [1a], according to the two approaches:

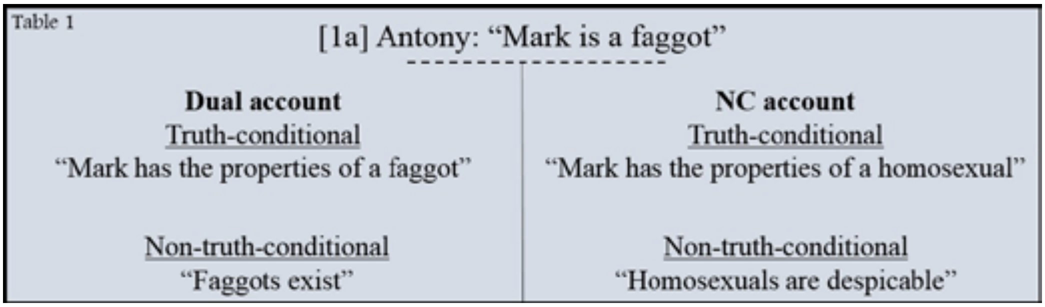

Table 1

[1a] Antony: “Mark is a faggot”

Dual account	**NC account**
Truth-conditional	Truth-conditional
“Mark has the properties of a faggot”	“Mark has the properties of a homosexual”
Non-truth-conditional	Non-truth-conditional
“Faggots exist”	“Homosexuals are despicable”

Two reading notes:

- concerning the dual account: the lexical content of ‘faggot’ is assumed to be something like ‘despicable because of being homosexual’.
- concerning the NC accounts: according to expressivism, the derogatory non-truth-conditional content is not propositional. It would correspond to a negative affective state.

2. At-issueness and Projective Behavior

2.1. The Persistence of Offensiveness in Non-assertive Structures

In the literature, the projection of the derogatory content of slurs has been strongly highlighted. According to many scholars, it would be evident that (at least a part of) the content of a slur embedded in a complex structure (negation, quote, question, conditional) shows a different behavior respect to the content of generic predicates and even pejoratives (‘fucker’, ‘asshole’, ‘idiot’, etc.). Williamson (2009), for instance, notes that concerning the occurrences of ‘boche’ (slur for ‘Germans’), “the xenophobic abuse is preserved in the negations” (p. 146).
The phenomenon has been clearly defined, but we find in the literature several ways to refer to it. In Potts (2007a), among the features of expressives, we find the “nondisplaceability”. Hedger (2012) talks about “scoping-out” whereas Camp (2013) talks about “projecting-out”. Finally, in Hom & May (2013), a paragraph on the topic is generically entitled “The persistence of offensiveness”.
As an example, let us consider the negation of [1a]:

[2] Antony: “Mark is not a faggot”.

The common intuition is that this utterance, although it is the negation of [1a], would continue to convey a derogatory content against homosexuals.[3] Cepollaro (2015) suggests that, given this view, the persistence of offensiveness in non-assertive structures seems to benefit the NC account. However, as we can infer from the next table, the dual account also has an explanation for this phenomenon.

[2] The author would like to thank the anonymous reviewer that suggested this label.
[3] See Panzeri & Carrus (2016) for some interesting empirical findings.

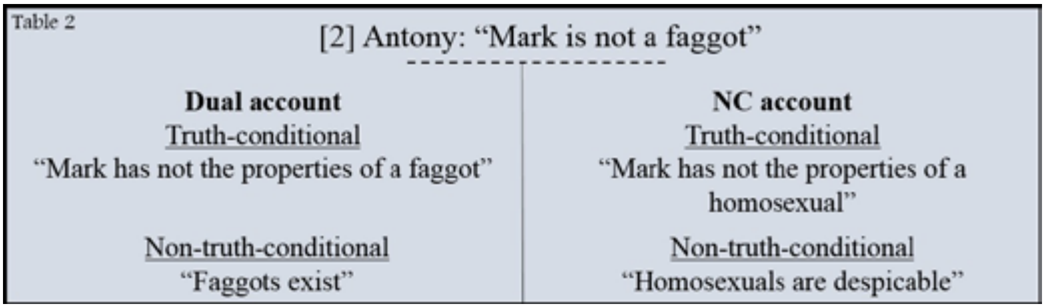

Table 2

[2] Antony: “Mark is not a faggot”

Dual account	**NC account**
Truth-conditional	Truth-conditional
“Mark has not the properties of a faggot”	“Mark has not the properties of a homosexual”
Non-truth-conditional	Non-truth-conditional
“Faggots exist”	“Homosexuals are despicable”

The explanations are slightly different:

- as we said, a dual account recognizes two kinds of derogation. The non-truth-conditional kind would depend on the fact that, using a slur, the speaker conversationally implicates his commitment to the non-null extension of the slur. In [2], just like in [1a], the speaker would implicate (that he is assuming) the existence of individuals despicable because of being homosexual;
- the NC account proposes that occurrences of slurs make derogatory the utterances in which they are embedded because they convey a derogatory content via presupposition (Schlenker, 2007; Cepollaro, 2015) or implicature (Potts, 2007a, 2007b; Williamson, 2009; McCready, 2010).

2.2. The Case of Denial

Following the presentation of the projection phenomenon, Camp (2013, p. 1) writes: “if we avoid repeating the offensive term by responding to [1] with something like - That’s not true. / That’s false. - then normally, we still manage to deny only that [Mark is homosexual]”. Therefore, the derogatory content of slurs persists in negations like [2] and, moreover, it cannot be blocked through denial.

[1a] Antony: “Mark is a faggot”.
[1b] John: “That’s not true”.

According to Camp, we are still faced with the projective behavior: denial would target the descriptive component (the fact that Mark is a homosexual), without modifying the derogatory content. Again, it could seem that the projection of the derogatory content can be better accounted for in a non-truth-conditional account. Let us see if this is the case. This conversational phenomenon can be divided in two parts: (1) the negation of the evaluative content; (2) the negation of the descriptive content. It is straightforward that:

i. if John interprets Antony’s utterance literally, and
ii. if John’s communicative intention is to oppose to the evaluative derogatory content,

as a competent speaker, he will realize that both the formulas ‘That’s not true’ and ‘That’s false’ are not apt. That is, John will not opt for denial. So, let us give a linguistic explanation for this fact. Check the table:

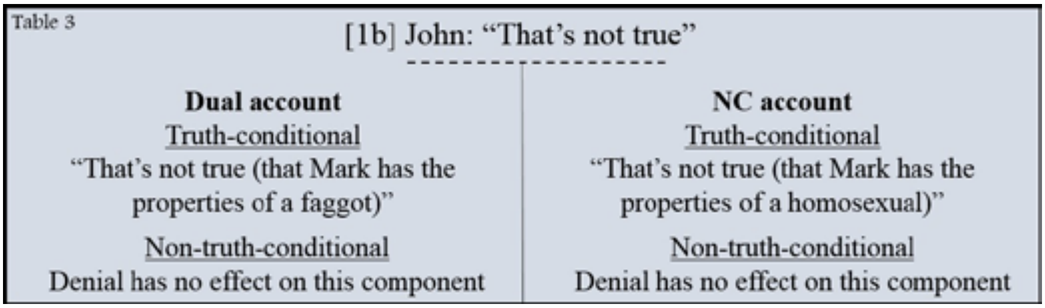

Table 3

[1b] John: "That's not true"

Dual account	NC account
Truth-conditional "That's not true (that Mark has the properties of a faggot)"	Truth-conditional "That's not true (that Mark has the properties of a homosexual)"
Non-truth-conditional Denial has no effect on this component	Non-truth-conditional Denial has no effect on this component

From a NC perspective, the linguistic issue that John must face is clear: the derogatory content has "pragmatic" and/or "not-at-issue" nature and cannot be blocked by a denial. From a dual perspective, the issue is far less clear, but we can agree that the result is the same. Indeed, it is reasonable that a non-bigot speaker who wants to oppose the derogation conveyed by [1a] focuses on the non-truth-conditional content. Because, in some sense, the truth-conditional content appears to be nothing more than a concept-token ("*Mark* is despicable because of *his* homosexuality") of the concept-type represented by the non-truth-conditional content ("*Homosexuals* are despicable because of *their* homosexuality"). The attempt to preserve Mark from a negative evaluation of his homosexuality would be irrational if not combined with a refusal of the general negative evaluation of homosexuality.

Thus far, it seems that both the accounts we have considered predict that answering to Antony's derogatory utterance through denial would be inappropriate. In this sense, the non-deniability of the evaluative component does not support the non-truth-conditional theories *over* the dual account, because the projection of a non-deniable content is also predicted by the dual account.

That said, in the quotation opening this section, we read that with a denial, opposed to an assertion containing a slur, we "manage to deny" the membership of the subject in the set denoted by the NC. The use of the expression 'manage' presupposes an attempt to do something. This sounds odd because speakers ordinarily do not *try,* they know well how to use a denial. In this sense, if we want to understand what Camp meant, we must probably look at the information about the target available to the speaker.

Consider table 4. If John believes that Mark is homosexual, the interpretation of a dialogue like [1a-b] will proceed as suggested above. No attempts: John will not opt for the denial. The point seems to be that speakers assume that 'faggot' refers to homosexuals.

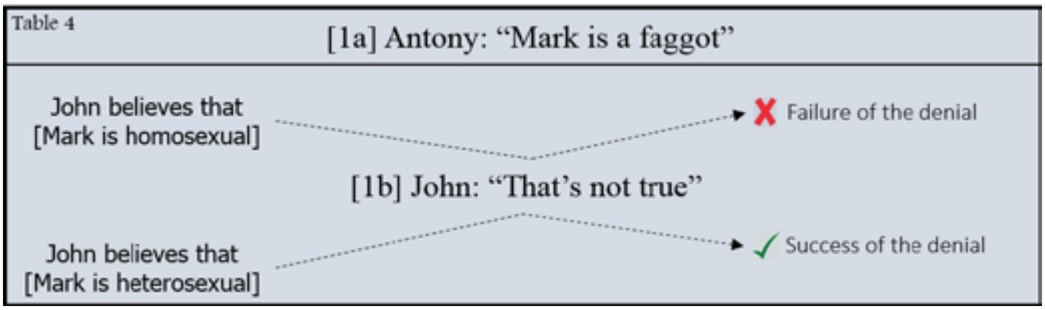
Table 4
[1a] Antony: "Mark is a faggot"
John believes that [Mark is homosexual]
✗ Failure of the denial
[1b] John: "That's not true"
John believes that [Mark is heterosexual]
✓ Success of the denial

Let us now imagine that John believes that Mark is not homosexual. Something changes, indeed in this latter case:

i. the use of the denial is allowed, and so ordinarily,
ii. even non-bigot speakers like John tend to object to a certain piece of content (e.g. homosexuality) intentionally allowing the categorical derogation to remain standing.

Here is what Camp (2013) probably means when she writes that, through denial, we "manage to deny" just the membership in the set denoted by the NC. She refers to contexts in which the speaker using the denial (John) entertains a specific belief ("the target is not a member of the set denoted by the NC").
In any case, here is also the puzzle for the dual account supporters: if the truth-conditional content of 'faggot' corresponds to *more than* homosexuality (evaluative + descriptive content), why can we use a denial to negate *just* the homosexuality (only descriptive content) of the target? The possible answers depend basically on what we are licensed to infer from *deniability*. As an example, one may wonder if deniability is a good diagnostics for truth-conditional relevance.

2.3. Deniability and At-issueness

According to Potts, deniability is useful to distinguish "entailments" from "nonentailments": "The meanings divide into two subclasses, entailments ('commitments') and nonentailments; the main factor in the split is the notion of deniability. [...] Nonentailments are deniable [...]. In contrast, entailments are not deniable" (Potts, 2005, p. 27). However, as Potts (2012) clarifies, truth-conditional contents and entailments are not the same. And even if Camp (2013) chooses not to address the question of the "theoretical status" of the evaluative content, the puzzle that she pointed out seems to depend on that status. Indeed, she quotes McCready (2010), according to whom, in general: "In ordinary denial, the truth of any at-issue part of a sentence can be called into question" (p. 7). And, in particular: "the negative part of the meaning of Kraut, and, by extension, pejoratives in general is CIE content [Expressive Conventional Implicature], and not part of the at-issue meaning" (p. 10). So, according to McCready (2010), as well as to Potts (2005, 2007a, 2007b), what the previous tables show is that the evaluative content of slurs is not at-issue. It seems that the "at-issue content" corresponds to the *most important* content conveyed via an utterance. "'At-issue entailment' sets up a useful contrast with CIs, which are secondary entailments that cooperative speakers rarely use to express controversial propositions or carry the main themes of a discourse" (Potts, 2005, p. 4). Consider [3]:

> [3] I spent part of every summer until I was ten with my grandmother, who lived in a working-class suburb of Boston.

With the purpose of clarifying what he means with "at-issue content", Potts (2005) highlights that, in [3], there are two assertions, but one of them plays "a secondary role relative to the information conveyed by the main clause" (p. 28). Therefore, first of all, it seems that "at-issueness" concerns the relation between propositions and the "main theme of a discourse". Indeed, Simons *et al.* (2010, p. 323) give the following definition of the notion:

a) A proposition *p* is at-issue iff the speaker intends to address the Question Under Discussion (QUD) via ?*p* (thus, the question *whether p*).
b) An intention to address the QUD via ?*p* is felicitous only if:
   i. ?*p* is relevant to the QUD, and
   ii. the speaker can reasonably expect the addressee to recognize this intention.

Now, if at-issueness is a matter of primacy concerning a (potentially implicit) QUD, it seems that co-text and context are fundamental to establish if a content is at-issue. Accordingly,

> identifying whether a particular proposition is at-issue according to the definition [...] requires judgments on whether one question is relevant to another [...] and judgments on whether a speaker can reasonably expect the addressee to recognize a particular intention (Tonhauser, 2012, p. 241).

Not only at-issueness is a notion that has to do with the QUD in a conversation, but also with "competent, cooperative addressees" able to "identify" (*ibidem*) the main themes of a discourse.
Concerning the previous example [1a], one may wonder if it is possible to *identify* the at-issue content, given that we do not know anything about the context. Consider:

A. Antony could be asserting just that Mark is a homosexual (from a homophobic perspective). Than the evaluative content would be secondary. Otherwise,
B. Antony could be asserting that Mark is despicable because of being homosexual. In this case, the evaluative content would be primary and thus, one can suppose, at-issue.

Here is an exemplification of the case B:

> [4a] John: "Why are you so hostile to Mark?"
> [4b] Antony: "Mark is a faggot".

Given the explicit QUD in [4a], it seems clear that the evaluative content of [4b] is at-issue.
A first conclusion is the following: if at-issueness does not strictly concern the lexical content of slurs (as we saw, the at-issue content depends on a sort of conversational salience), one may wonder if this "theoretical status" is really useful for the debate. In addition, as a collateral outcome, the same skepticism may affect the deniability test. Consider, as an example, that even the supporters of presuppositional accounts use the argument of deniability against truth-conditional accounts. Yet, according to Potts (2005), even presuppositions (and conversational implicatures) are deniable.
So let us analyze the deniability of [4b]:

> [4b] Antony: "Mark is a faggot".
> [4c] John: "That's not true".

Probably, the most natural interpretation of [4c] is the one according to which the argument of the denial is the proposition 'I am so hostile to Mark because Mark is a faggot'. In other terms, the interpretation according to which John is discussing the reason of the hostility. However, we are not interested in this case. Indeed, for this interpretation, whether 'faggot' and 'homosexual' are coreferential has no relevance.
Then we proceed to analyze [4a-c] as we did for [1a-b]. See table 5.

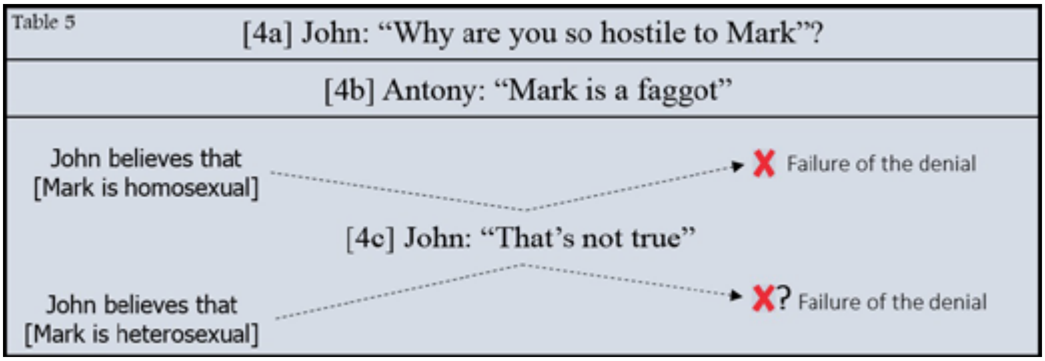

Table 5

[4a] John: "Why are you so hostile to Mark"?

[4b] Antony: "Mark is a faggot"

John believes that [Mark is homosexual]
✗ Failure of the denial
[4c] John: "That's not true"
John believes that [Mark is heterosexual]
✗? Failure of the denial

If John, believing that Mark is homosexual, wants to block the derogatory content expressed by Antony, denial ([4c]) works as usual. Thus, it continues to sound inappropriate (while being a potential failure). Imagine that John believes that Mark is *not* homosexual. Our previous analysis predicts that John will answer through denial in order to target the proposition 'Mark is a homosexual'. Yet, something is wrong. Although it is clear that John *is allowed* to use [4c] in that way, the prediction is somewhat surprising because (1) the denial should target the at-issue content and (2) the at-issue content of [4b] is something like 'Mark is despicable because of being homosexual'. The QUD in [4a-c] is not the sexual orientation *per se*, but rather the evaluation of the sexual orientation. However, the alternative is to accept different analysis of the same proposition (compare table 5 to table 4). In sum:

i. the contribution 'faggot' offers to the at-issue content of an utterance is context-dependent;[4]
ii. the evaluative component can be at-issue;
iii. even if this content is at-issue, denial will sound inappropriate.

Then, in conclusion, it seems that according to speakers the evaluative content of slurs is not *a matter of truth*. However, it is controversial that the notion of at-issueness can help us to better understand the relation between slurs' lexical content and the truth of the utterances in which they are embedded.

In any case, in the following section, we will use the apparatus presented so far (theories, projection and denial) in order to investigate a real case of non-standard use of slurs (e.g. 'faggot' used to derogate a heterosexual man).

3. Non-standard Uses

Croom (2015) argues against the co-referentiality thesis. If we accept his argument, we should actually acknowledge that the supposed NCs of the slurs are not NCs. Briefly, the idea is that since slurs and NCs are ordinarily used to denote different sets of entities they thereby differ in their meaning. 'Faggot', as an example, would refer to individuals in the world on the basis of a set of properties that *can or cannot* contain the property of "being homosexual". In this framework, the undeniable relevant role played by the property of being homosexual in the standard use of 'faggot' is explained by the notion of "conceptual

[4] An anonymous reviewer expressed her/his concern about this conclusion. In particular, s/he suggested that the at-issue content of a word cannot be context-dependent because it should depend on its lexical meaning. I agree. It is precisely that kind of concern that makes me worry about how a "question under discussion" conceived as "the primary goal" and/or "the immediate topic" of a conversation (Roberts, 1996) can be useful for investigating the lexical content codified by an isolated word.

anchor": "which may be understood as the most relevantly salient (rather than necessary) default descriptor that helps communicative agents ground the apt application of S(lur) towards its prototypical (rather than essentially categorical) targets" (Croom, 2015, p. 35). According to Croom, we need such a theory essentially to account for those uses that Jeshion (2013a) calls *nonliteral*:

> Let us distinguish these basic uses from two broader uses that I will dub "nonliteral." One involves applications to only those members of the group referenced by the slur that are stereotype-conforming. The other involves applications to those perceived to be exhibiting properties in the stereotype of the slur's neutral counterpart, yet who are not members of the group (p. 324).

In particular, Croom presents four sources of supporting evidence showing that slurs and NCs are in fact not co-referential and then, one can suppose, showing that so-called non-literal uses do not constitute an independent category:

i. the discussion provided by Szekely (2008); here the author reports that the slur 'faggot' was in fact used to apply to some but not all male homosexuals;
ii. the discussion provided by MacDonald (1999); here the author discusses how slurs were used in his linguistic community and reports that 'nigger' was in fact used to apply to some but not all African Americans;
iii. the discussion provided by Troyani (2013); here the author reports that the slur 'guido' was in fact used to apply to some but not all Italian-Americans;
iv. the discussion provided by the comedian Chris Rock in his famous routine completely based on the conceptual clash "Niggas vs. Black People".

For sake of brevity, I shall not go into details; however, Croom's discussion raises the following issue, to which we shall now turn: if a speaker makes a non-standard use of a slur, will we find differences in the ways in which the addresses use denials? Let us consider a recent episode. "I wanted to poke fun at Mancini for the fact that he enters the field as for a wedding party. I meant '*fighetto*', not that thing about sex!".[5] The quote is from the newsweekly *Chi*: in the article, Maurizio Sarri, current coach of *Società Sportiva Calcio Napoli*, tries to explain why he addressed Roberto Mancini, former coach of Football Club Internazionale Milano, with '*frocio*' ('faggot'). '*Fighetto*' is a derogatory Italian term used to mean something between 'snooty' and 'posh'.[6] Consider the following exchange:

> [5a] Maurizio: "Roberto is a faggot".
> [5b] Sinisa: "That's not true".

In the previous paragraphs, we assumed that to exhibit the typical property of the NC (being homosexual, being Jew, being Italian, etc.) is the critical factor for being target of the corresponding slur ('faggot', 'kike', 'wop', etc.). Nevertheless, what Maurizio Sarri is claiming is that the meaning of 'faggot' in [5a] is something like "snooty person who enters the field as for a wedding party". Let us see how the two theories account for this use.

[5] My translation; for the original quote, see: http://bit.ly/1TdHlyC.

[6] Sometimes these attributes come together with something like "effeminate". This connection is a good candidate to explain this non-standard use of 'faggot'.

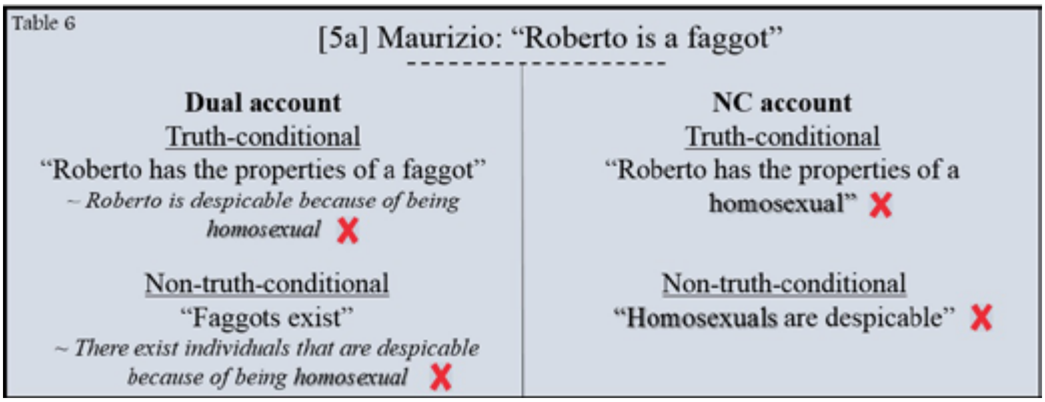

Table 6

[5a] Maurizio: "Roberto is a faggot"

Dual account	NC account
Truth-conditional "Roberto has the properties of a faggot" ~ *Roberto is despicable because of being homosexual* ✗	Truth-conditional "Roberto has the properties of a homosexual" ✗
Non-truth-conditional "Faggots exist" ~ *There exist individuals that are despicable because of being homosexual* ✗	Non-truth-conditional "Homosexuals are despicable" ✗

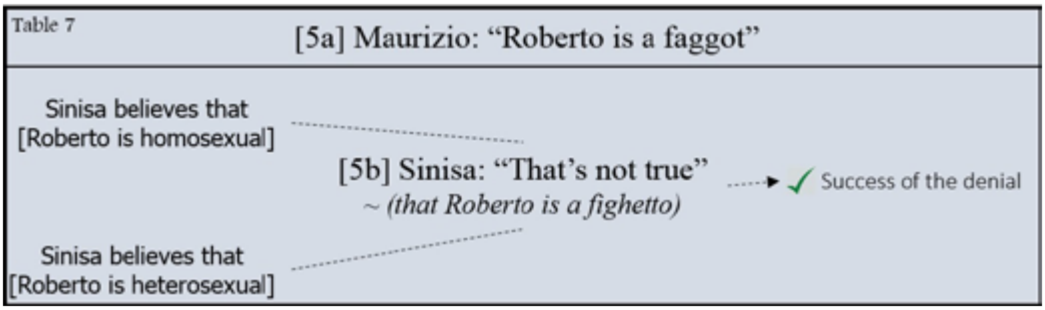

Table 7

[5a] Maurizio: "Roberto is a faggot"

Sinisa believes that [Roberto is homosexual]

Sinisa believes that [Roberto is heterosexual]

[5b] Sinisa: "That's not true"
~ *(that Roberto is a fighetto)*

✓ Success of the denial

In table 7, we focus on the relevance of the beliefs that Sinisa may entertain about Roberto's homosexuality. Given that we are assuming that Sinisa interprets Maurizio's utterance in the way Maurizio suggested a posteriori, that is, as the attribution to Roberto of the property (or set of properties) of *being a snooty person who enters the field as for a wedding party*, one could conclude that those beliefs do not influence the potential use of denial.
Now, let us focus on the relevance of the beliefs that Sinisa may entertain about the *actual* predication.

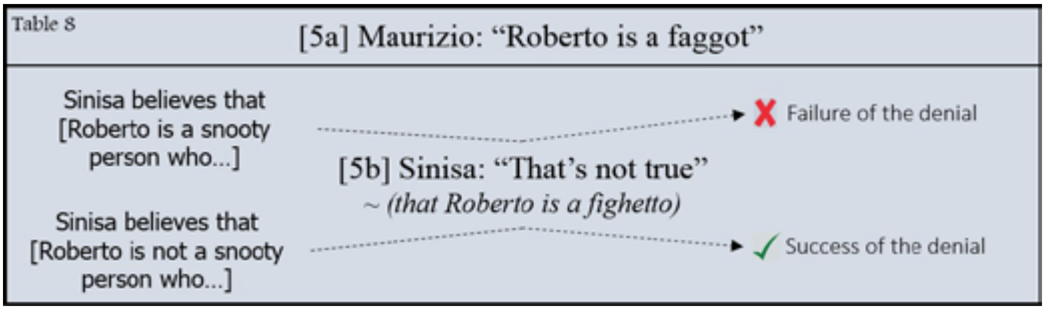

Table 8

[5a] Maurizio: "Roberto is a faggot"

Sinisa believes that [Roberto is a snooty person who...] → ✗ Failure of the denial

[5b] Sinisa: "That's not true"
~ *(that Roberto is a fighetto)*

Sinisa believes that [Roberto is not a snooty person who...] → ✓ Success of the denial

Here, the point is that Sinisa may deny the assertion by Maurizio either because he believes that Maurizio is saying something false or because, whatever he thinks about Roberto, he wants to protect his friend from the derogatory content conveyed by Maurizio. Table 8 sums up the two cases and we see that in [5a-b], exactly like in [1a-b] (see table 4), if the speaker:

i. is interested in generically denying the evaluative content, with no interest in denying the descriptive content, he will not use the denial formula (failure);
ii. wants to negate the descriptive content, he may successfully use the denial (success).

At this point, it will be clear that both the theories we discussed run into some problems. Briefly, according to what Maurizio claimed, sexual orientation would not be at issue in [5a] and if that was correct, it would follow that:

i. the fact that the speaker (Sinisa) believes or not that the target (Roberto) exhibits the typical properties of the NC is irrelevant for the effectiveness of the denial (table 7). Indeed, in this case it is improper to say that Sinisa *manages to deny* only that Roberto is homosexual;
ii. in general, both the NC account and the dual account appear unsatisfactory (table 6).

Let us now subsume this explanatory weakness under a more general semantic normativity problem. Following Marconi (1997), we know that Sarri, like everyone else, should

> accept (and is regarded as socially obliged to accept) the consequences of his assertions taken in the sense in which semantically authoritative speakers take them, independently of whether such a sense coincides with the sense he intended them to have (p. 129).

Therefore, given our theories, we expect that the projected content smoothly enter into the common ground. The audience of Maurizio, even if confronted with the utterance at a different time, should tend to react to that content (derogatory towards homosexual people). But what happened in this case? The interpretation of Sarri's words divided public opinion:

i. on the one hand, some people interpreted Sarri's utterance according to the predictions of the accounts we have considered, thus they attributed to him the commitment to a homophobic content like 'Homosexuals are despicable';[7]
ii. on the other hand, some people interpreted Sarri's utterance as lacking any reference to homosexuality.[8]

Note that, although the victim prompted the standard interpretation ("He used racist words [...] shouting, saying *frocio*, finocchio; [...] in England, if anyone used those words, he would be banished from any kind of field"[9]) even the institution in charge has decided for the second non-standard interpretation: "The decision by the Sport Judge formally clarifies the absence of any racist or homophobic connotation in coach Sarri's words".[10] A significant part of the linguistic community recognized that Sarri's intended meaning reflects an existing use, and it is likely that Sarri used the slur, "not just with the *intention* of using it as everybody else in the community does but under the *assumption* that [he was] using it as everybody else does" (Marconi, 1997, p. 216). For these reasons, non-standard uses of 'faggot' cannot be conceived as occurrences of a *private*

[7] "Sarri fell back into it. After two years, again with homophobic insults". My translation; for the original, see: http://bit.ly/1QSADqP.

[8] "Stop hypocrisy: Sarri is not a homophobic and Mancini is not gay." My translation; for the original, see: http://bit.ly/2mFcwrw.

[9] Mancini during the after match press conference. My translation; for the original, see: http://bit.ly/1SeR1qY.

[10] My translation; for the original, see: http://bit.ly/1UFRinw.

language (Wittgenstein, 1953). On the other hand, Croom (2015) suggested that it is strongly controversial to consider these uses "figurative": Croom proposes to account for non-standard uses of slurs by adopting a family resemblance conception of category membership (Rosch & Mervis, 1975; Wittgenstein, 1953). Accordingly, in any use of 'faggot', HOMOSEXUALITY would work as a "conceptual anchor". However, Croom's proposal may not be conclusive:

i. it does not take into account Sarri's claim – supported by part of the linguistic community and by the judge – according to which in his use of '*frocio*', sexuality was *not* salient, rejecting the accusations of homophobia;
ii. opponents can say that, if HOMOSEXUALITY is the conceptual anchor "that helps communicative agents ground the apt application of ['faggot'] towards its prototypical targets" (Croom, 2015, p. 35), it remains somehow salient even in non-standard uses.

The supposed optionality of the "conceptual anchor" cannot account for the fact that there are at least two distinct linguistic sub-communities with different opinions about the application conditions of the term.

> The communicative intentions behind that which we communicate must be specific, determinate, and definite if they are to be calculable; [...] If an individual speaker comes to associate with the use of a particular expression a set of conditions dissimilar to that which another competent speaker associates with it, there is no problem if both are constant in their distinct associations; this is what we normally attribute to idiolect or ambiguity (Lepore, 2015, p. 7).

The reported case is just a clear example of a linguistic fact concerning 'faggot': there exists a non-standard association between the word and a set of conditions of application no less clear, constant and public than the standard one. We can account for this fact in two different ways:

i. if we assume the classical theory of concepts and then we assume that belonging to the target class is a necessary condition for the application of the slur 'faggot', then we should recognize that Sarri's use of 'faggot' features another word, different from the slur 'faggot' in that it has another meaning. The term 'faggot' would be semantically ambiguous under this reading;
ii. otherwise, if we assume the family resemblance theory, according to which none of the semantic traits associated to 'faggot' is necessary for its application, we should recognize that Sarri's use of the slur 'faggot' was not really different from any standard use of the slur.

4. Conclusion

In the literature (in particular, see McCready, 2010 and Camp, 2013), it has been said that the denial ('It's not true that P' / 'P is false') is a linguistic formula capable to highlight the projection of the slurs' derogatory content. In this article, we have investigated the interaction between an utterance containing a slur and a denial.

In the first part of the article (§2), we briefly presented our apparatus. We introduced the two main approaches to analyze slurs meaning (dual account *vs.* NC account), showing how both theories account for the projection in negations (table 2) and for the inappropriateness of answering back to the derogatory content of slurs through denial (table 3). However, as Camp (2013) suggested, the fact that we can use denial to deny the membership of the target in the set denoted by the NC constitutes a puzzle for truth-conditional approaches. In this regard, if some scholars proposed to explain this phenomenon by way of the notion of at-issueness (Potts, 2005; McCready, 2010), we think that it is important to consider that that notion apparently has to do with the "topics of discussion" (Roberts, 1996) rather than with the lexical content of isolated words. In particular, we would tend to assume that, differently from at-issue contents (see Simons *et al.*, 2010; Tonhauser, 2012), the lexical content of a term should not be something that (a) speakers *identify in context* through (b) a *cooperative behavior*. We also tried to show that the at-issue content of (even) an atomic proposition containing a slur is context-dependent.

Leaving aside this line of speculation, in the second part of the article (§3), we used the apparatus to analyze non-standard uses of slurs (e.g. 'faggot' used to derogate an heterosexual man). Considering the public opinion reactions to a recent incident (see the article: "Inter's Roberto Mancini: Napoli manager Maurizio Sarri called me a faggot"[11]), we argued that both the NC account and the dual account would predict that Sarri cannot possibly be right in claiming that his use of 'faggot' was not homophobic, because of the crucial role attributed to the typical trait of the supposed NC (in that case, homosexuality).

In this sense, non-standard uses seem to pose an unsolved (but *politically* relevant) semantic normativity problem. On the one hand, there is a linguistic sub-community, according to which each occurrence of the slur would be derogatory towards homosexuals. On the other hand, there is a linguistic sub-community (consisting *at least* of Sarri and the judge but possibly many more speakers), according to which *'frocio'* could be used with no reference to homosexuality. So, when observing the existence of these conflicting readings, one may wonder whether the judge made the right decision concerning the Sarri-Mancini incident.

We suggested two different ways to deal with the problem. If one assumes that being gay is a necessary trait for the application of 'faggot' in standard uses and that Sarri's use reflects an existing use, then one should recognize that the Italian derogatory term '*frocio*' is semantically ambiguous. On the other hand, if we assume that none of the semantic traits associated to the term is necessary for its standard application, we should recognize that when a slur like 'faggot' (as well as other slurs) is applied to individuals who do *not* belong to the target class the speaker still runs the risk of being derogatory to the target class itself. In any case, the judge's decision is controversial because it is very doubtful that a use of 'faggot' which do not refer to homosexuality really exists. So, if we choose the first option (classical theory of concepts), the meaning-related source of the conflict between the two communities needs further explanation.[12] If we choose the second option (family resemblance theory), the anti-homophobic prohibition on the use of 'faggot' seems to rest on the idea according to which all the possible uses of the slur are strongly "related" (Wittgenstein, 1953).

REFERENCES

Anderson, L., & Lepore, E. (2013). What did you call me? Slurs as prohibited words. *Analytic Philosophy*, 54(3), 350-363.

Camp, E. (2013). Slurring Perspectives. *Analytic Philosophy*, 54(3), 330-349.

Cepollaro, B. (2015). In Defense of a Presuppositional Account of Slurs. *Language Sciences*, 52, 36-45.

Croom, A. (2011). Slurs. *Language Sciences*, 33, 343-358.

Croom, A. (2014). The Semantics of Slurs: a Refutation of Pure Expressivism. *Language Sciences*, 41, 227-242.

Croom, A. (2015). The Semantics of Slurs: a Refutation of Coreferentialism. *Ampersand*, 2, 30-38.

[11] *The Guardian* online: http://bit.ly/20dbz59.

[12] The hypothesis of an "edict of prohibition" (Anderson & Lepore, 2013) seems appropriate for this approach.

Frege, G. (1892). On sense and reference. In P. Geach & M. Black (Eds.). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 56-78.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hedger, J. (2012). The Semantics of Racial Slurs: Using Kaplan's Framework to Provide a Theory of the Meaning of Derogatory Epithets. *Linguistic and Philosophical Investigations*, 11, 74-84.

Hom, C. (2008). The Semantics of Racial Epithets. *Journal of Philosophy*, 105, 416-440.

Hom, C., & May R. (2013). Moral and Semantic Innocence. *Analytic Philosophy*, 54(3), 293-313.

Jeshion, R. (2013a). Slurs and Stereotypes. *Analytic Philosophy*, 54(3), 314-329.

Jeshion, R. (2013b). Expressivism and the Offensiveness of Slurs. *Philosophical Perspectives*, 27, 231-259.

Kaplan, D. (1999). The Meaning of Ouch and Oops: Explorations in the Theory of Meaning as Use, manuscript.

Lepore, E. (2015). On the Perspective-Taking and Open-Endedness of Slurring, manuscript.

Macià, J. (2002). Presuposición y significado expressivo. *Theoria: Revista de Teoria, Historia y Fundamentos de la Ciencia*, 3(45), 499-513.

Marconi, D. (1997), *Lexical Competence*. Cambridge, MA: MIT Press.

Mac Donald, M.P. (1999). *All souls: A Family story from southie*. New York: Ballentine.

McCready, E. (2010). Varieties of Conventional Implicature. *Semantics & Pragmatics*, 3, 1-57.

Panzeri, F., & Carrus, S. (2016). Slurs and Negation. *Phenomenology and Mind*, 11, 170-180.

Potts, C. (2007a). The Expressive Dimension. *Theoretical Linguistics*, 33, 165-197.

Potts, C. (2007b). The Centrality of Expressive Indices. *Theoretical Linguistics*, 33(2), 255-268.

Potts, C. (2015). Presupposition and Implicature, in S. Lappin & C. Fox (Eds). *The Handbook of Contemporary Semantic Theory* (2nd ed.). West Sussex, UK: Blackwell Publishing.

Putnam, H. (1970). Is semantic possible?. In H. Putnam (2003), *Mind, Language and Reality. Philosophical Papers*, 2. Cambridge: Cambridge University Press.

Putnam, H. (1975). The Meaning of 'meaning'. In H. Putnam (2003), *Mind, Language and Reality. Philosophical Papers*, 2. Cambridge: Cambridge University Press.

Roberts, C. (1996). Information Structure: Towards an Integrated Formal Theory of Pragmatics. In J.H. Yoon & A. Kathol (Eds.). *Papers in Semantics*, 49. Columbus, OH: Ohio State University Press, 91-136.

Rosch, E., & Mervis, C. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7, 573-605.

Schlenker, P. (2007). Expressive Presuppositions. *Theoretical Linguistics*, 33, 237-245.

Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2010). What projects and why. *Semantics and Linguistic Theory*, 21, 309-327.

Szekely, L. (2008). Offensive Words (track 1). In *Chewed Up*, Los Angeles: Image Entertainment.

Tonhauser, J. (2012). Diagnosing (not-)at-issue content. *Proceedings of Semantics of Under-represented Languages of the Americas (SULA)*, 6, 239-254.

Troyani, S. (2013). Guido culture: The destabilization of Italian-American identity on Jersey Shore. *California Italian Studies*, 4(2). http://escholarship.org/uc/item/1m95s09q.

Williamson, T. (2009). Reference, Inference and the Semantics of Pejoratives. In J. Almog & P. Leonardi (Eds.). *The philosophy of David Kaplan*. Oxford: Oxford University Press, 137-158.

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

ANDRÉS SORIA RUIZ
ENS-PSL, Paris - Institut Jean Nicod
andressoriaruiz@gmail.com

THOMASON (UN)CONDITIONALS

abstract

Thomason conditionals are sentences of the form if p, ~Kp. Given plausible assumptions, these sentences cause trouble for epistemic theories of indicative conditionals. Our aim is to show that Thomason examples are not indicative conditionals, but alternative unconditionals, in the sense put forward by Rawlins (2013). This hypothesis solves the difficulty and explains certain features that set Thomason examples apart from run-of-the-mill indicative conditionals.

keywords

indicative conditionals, unconditionals, epistemic modality, semantics

Phenomenology and Mind, n. 12 - 2017, pp. 98-108
DOI: 10.13128/Phe_Mi-21109
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Two weeks into your new office job and things are starting to look a bit eerie. You are happy with the job, but there is something unsettling about your coworkers: they are extraordinarily reserved. You greet them as you arrive, bid them good evening as you leave, yet nothing but a nod comes out of them. You walk to the coffee machine and no one raises their gaze from their cubicles. On the few occasions that you have had lunch with them, the conversation was brief and noncommittal. In these circumstances, you may be justified in thinking:

(1) If my coworkers hate me, I have absolutely no idea.

The present paper is about the right semantics for sentences like (1). Van Fraassen (1980, p. 503) attributes examples like these to Richmond Thomason, so – following Bennett (2003) – we will dub sentences like (1) *Thomason conditionals* or *examples*. In particular, (1) is very similar to an example by Stalnaker (1984, p. 105). Constructions like these cause trouble for broadly epistemic theories of indicative conditionals. These theories maintain, roughly, that the role of antecedents is to temporarily update a knowledge state with the information that the antecedent is true, and then check whether the consequent holds with respect to the updated knowledge state. But the consequent denies knowledge of the antecedent, so it cannot hold true with respect to a knowledge state updated with the antecedent.

What I propose is to treat Thomason examples not as *bona fide* indicatives, but rather as *alternative unconditionals* (Rawlins, 2013). These are sentences whose syntax is superficially very similar to that of an indicative conditional, but where the consequent holds *unconditionally*, that is, regardless of whether the antecedent is true or false.

The paper is structured as follows: in section 1, I present the simplest version of the epistemic theory of indicative conditionals – namely Ramsey's test – and show how Thomason conditionals cause trouble for it. I then present the problem for more contemporary versions of the theory, representing states of information *via* epistemic modal bases. Section

[1] I would like to thank Chris Barker, Kyle Blumberg, David Bordonaba, Paul Egré, Diego Feinmann, Víctor Fernández, María José Frápolli, Daniel Galdeano, Javier González de Prado, Dan Hoek, Benjamin Icard, Natalia Karczewska, Jose de Luna, Mora Maldonado, Eduardo Pérez, Manuel de Pinedo, Mirco Sambrotta, Isidora Stojanovic, José Ramón Torices, Neftalí Villanueva and audiences at the 8th SEFA Congress in Oviedo and NYU Linguistics for their help and comments. Support for this work comes from the Spanish Fulbright Commission, Obra Social La Caixa and grant numbers ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL. Thanks to the editors of this issue and two anonymous reviewers.

2 considers three relatively obvious ways of avoiding the problem posed by Thomason conditionals, but finds them lacking. In section 3, the hypothesis that Thomason conditionals are unconditionals is put forward. This hypothesis receives support from the uncommon behavior of Thomason examples under paraphrase with *only if* and contraposition. Truth-conditions for claims like (1) are provided. Section 4 concludes.

1. The Ramsey Test and Thomason Conditionals

Epistemic theories of indicative conditionals go back at least to Ramsey, who held the following view: "If two people are arguing 'If *p*, will *q*?' and are both in doubt as to *p*, they are adding *p* hypothetically to their stock of knowledge and arguing on that basis about *q*..." (Ramsey 1931/2001, p. 247, n. 1). Ramsey's test, as this proposal has come to be known, appears to be applicable to many indicative conditionals, insofar as these sentences convey their utterer's ignorance with respect to the claim in their antecedent. Ignorance about the antecedent is most often taken to be a presupposition triggered by the presence of indicative morphology in both antecedent and consequent, in contrast to the subjunctive morphology of counterfactuals (see Stalnaker, 1975 for an early statement of this view). It is easy to see why Thomason conditionals make trouble for this view. These constructions are of the form *if p, ~Kp*, and they are in the indicative mood, so the Ramsey test should be applicable to them. But herein lies the difficulty: if the role of the antecedent is to update our stock of knowledge, then the update guarantees *Kp*, which is what the consequent denies (see Chalmers and Hájek, 2007 for a clear and brief state of the problem). Note too that the trouble caused by Thomason conditionals appears only when the conditionals are about participants in the conversation. Compare (1) and (2) with (3):

(2) If your coworkers hate you, you have absolutely no idea.
(3) If Alice's coworkers hate her, she has absolutely no idea.

In (2) we have the same problem as in (1): once we (temporarily) add the information that your coworkers hate you to *our* knowledge, it follows that you know it. By contrast, the trouble disappears in (3), assuming that Alice is not taking part in our conversation: adding the information that her coworkers hate her to our knowledge will certainly not guarantee that she is in the loop.

The trouble remains when we turn to more contemporary versions of the epistemic theory (among many others, Gillies, 2010; Kratzer, 1986; Stalnaker, 1975, 2014). Broadly speaking, these theories represent indicative conditionals as first intersecting a contextually determined epistemic modal base with the proposition in the antecedent and then evaluating its consequent with respect to the modal base thus updated. Let us define epistemic modal bases as follows (I take this and the *(Definedness)* condition below from Gillies, 2010):

(*Epistemic modal base*): given a context c and a world w, C is a modal base (for c,w) only if $C^{c,w} = \{w' : w'$ is compatible with the c-relevant information at $w\}$

Truth-conditions for indicative conditionals are given thus:

(*Indicatives*): $[[\text{if } p, q]]^{c,w} = 1$ iff $C^{c,w} \cap [[p]]^{c} \subseteq [[q]]^{c}$

To see how Thomason conditionals make trouble for this view, we need to make three assumptions: first, assume that an indicative conditional is defined at a context c only if its antecedent is an open possibility relative to a modal base C (for c,w):

(*Definedness*): $[[\text{if } p, q]]^{c,w}$ is defined only if p is compatible with $C^{c,w}$

Next, assume a simple, text-book semantics for *knows* such as the one to be found in Heim & Kratzer (1998, p. 306). Where $K^{x,w}$ is the set of worlds that are epistemically accessible for knower *x* at *w*,

(*Know*): $[[\text{know}]]^{c,w} = \lambda p \lambda x.\forall w' \in K^{x,w} : [[p]]^{c,w'} = 1$.

Finally, assume that the modal base $C^{c,w}$ is a superset of any interlocutor *m* at *c*'s set $K^{m,w}$ of epistemic alternatives. In other words, any participant in a conversation knows at least as much as what is known with respect to $C^{c,w}$:

($K \subseteq C$): for any interlocutor *m* in *c,w*, $K^{m,w} \subseteq C^{c,w}$.

With these assumption in place, consider a felicitous utterance of (1) at context *c* and world *w*: by (*Definedness*), we will restrict $C^{c,w}$ with the information in the antecedent *that my coworkers hate me*; then, by our second and third assumption that information either restricts the speaker's knowledge $K^{speaker,w}$ or is already true across it. Either way, the proposition in the consequent, *that I have absolutely no idea*, is false, since $K^{speaker,w}$ has been updated with the information in the antecedent. So, if the epistemic theory is right, sentences like (1) should be invariably false. However, they *can* be true, and therefore the epistemic theory is in trouble.

2. Three Obvious Escapes

Before we move to our preferred solution to this problem, let us revise three relatively obvious escape routes. The first is to treat the antecedents of Thomason conditionals as satisfying a presupposition in their consequent. The second is to drop the requirement that indicatives quantify over *epistemic* possibilities. The third solution is to allow the epistemic operator in the consequent of a Thomason conditional (or any indicative conditional, for that matter) to access worlds outside the updated modal base.

Dynamic semantics treatments of conditionals and presupposition (stemming from Heim, 1983/2002; see Schlenker, 2011 for an overview) suggest a way of interpreting sentences like (1) that would explain at least some of these cases away: Thomason conditionals may be sentences whose antecedents satisfy a presupposition of their consequents. Take a sentence like (1), whose consequent is a negative knowledge claim, *~Kp*. *Kp* presupposes *p*, and since presuppositions project under negation, *~Kp* also presupposes *p*. The idea is that in a sentence of the form *if p, ~Kp*, the antecedent is there just to satisfy the presupposition of the consequent clause *~Kp*, the bare utterance of which would be infelicitous otherwise.

This solution is suggestive, and it would place Thomason conditionals along constructions such as '*If Jane used to smoke, she has stopped*'. However, Thomason conditionals are also of the form:

(4) If my coworkers hate me, it's not obvious that they do.
(5) If my coworkers hate me, I can't be sure about it.
(6) If my coworkers hate me, I can't tell.

Sentences like (1) and the different sentences (4)-(6) display the same structure (that is, the structure of a conditional with indicative morphology, with a certain clause in the antecedent and an epistemic verb embedding that same clause in the consequent) and they seem to have roughly the same meaning – they all express essentially *ignorance* about the coworkers' feelings towards the speaker. But the different consequent clauses in (4)-(6) do not presuppose the truth of the clause embedded under the attitude verbs in (4)-(6). Thus, the proposed explanation could not be extended to Thomason conditionals like (4)-(6). Insofar as one aims at offering a general account of these constructions, this one will not do.

The second solution consists on giving up the distinction between indicative and subjective conditionals in terms of the idea that indicatives quantify over *epistemic* possibilities. This solution may seem *prima facie* attractive, insofar as the root of the problem caused by Thomason conditionals appears to lie in the assumption that their antecedent intersects a knowledge state. If we drop this assumption, the problem vanishes: the relevant updates need no more carry over to our knowledge. This can be seen by noting that Thomason *subjective* conditionals are not problematic:

(7) If my coworkers hated me, I would have absolutely no idea.

Since it is not assumed that the antecedent of a subjunctive conditional restricts anyone's knowledge, it does not follow from (7) that anyone *knows* that my coworkers hate me. Furthermore, this seems to be the intuitively right way of interpreting a sentence like (1): entertaining the possibility that my coworkers hate me is not thereby entertaining the additional possibility that I am aware of it, since different things follow from each possibility: if they hate me, then life will probably go on as it has for the past two weeks; but if I *learn* that they hate me, I may quit my job. Nonetheless, if we hold with the epistemic theory that such a (momentary) update is an update on our knowledge, then whatever follows from adding *that we possess* that information should follow already from simply *adding* the information to our knowledge. We may therefore give up the idea that the antecedent of an indicative restricts a knowledge state.[2]

However, dropping the characterization of indicatives as operating on epistemic possibilities has a price, since this theory has both theoretical appeal and philosophical pedigree: among other virtues, it straightforwardly accounts for the epistemic *feel* of indicatives, that is, the idea that indicatives are concerned with how things might turn out to be (rather than with how things would or might *have* been). The epistemic theory also captures the observation that speakers possessing different pieces of evidence may be justified in uttering indicative conditionals with contradictory consequents – Gibbard's (1981) Sly Pete and Bennett's (2003) Top Gate example are two well-known cases. Finally, in dynamic semantics, this theory of the meaning of indicatives provides a good explanation of the projection behavior of presuppositions embedded in conditionals (Schlenker, 2011). In sum, abandoning a semantics for indicatives in terms of operations over epistemic possibilities seems too rash a move to make in light of Thomason conditionals. We aim to show that it is also unnecessary.

The third way of avoiding the problem caused by Thomason conditionals is to note that there could be accessible non-antecedent worlds even after the relevant modal base has been updated with the antecedent. This way, we would allow the ignorance claim in the consequent to come out true even when the antecedent is true throughout the modal base.

This is the most natural way out. We presented the problem as originating, in part, in our assumption ($K \subseteq C$) that any update on $C^{c,w}$ ought to carry over to $K^{m,w}$ (for any interlocutor *m* in *c,w*). But in fact, the ignorance claim embedded in the consequent of (1) is not to be evaluated at the context of utterance *w*, but at each world in the updated modal base:

(*Truth-conditions for* (1)): $[[\text{if } p, \sim Kp]]^{c,w} = 1 \text{ iff } C^{c,w} \cap [[p]]^{c} \subseteq [[\sim Kp]]^{c}$

[2] Not everyone who works with this kind of theory of indicatives makes this assumption. Yalcin (2007), for example, is careful to state his observations about epistemic contradictions embedded under *if* and *suppose* without assuming that those operators range over epistemic possibilities. He just takes them to operate on an information parameter in the circumstances of evaluation.

So even though an update on $C^{c,w}$ would effect the corresponding update on $K^{m,w}$, an update on $C^{c,w}$ need not effect a similar update on $K^{m,w'}$, where *w’* is any world in the updated modal base. Where *p* is the proposition *that my coworkers hate me*, what the previous truth-conditions state is that the intersection of the modal base $C^{c,w}$ and *p* is a subset of the set of worlds at which I am ignorant of *p*. For it to be true that each world *w’* in the update is a world at which I do not know that they hate me, there ought to be at least one non-hate world in $K^{m,w'}$. If nothing prevents us from accessing such non-hate worlds, then we are not in trouble. The reply to the purported counterexample would be straightforward: *no*, epistemic theories do not predict sentences like (1) to have false consequents; the illusion that they do so is due to a confusion caused by evaluating the knowledge claim in the consequent at the world of utterance, instead of evaluating it at each world in the updated modal base.

This is a fine solution as far as it goes. But something *does* prevent us from accessing non-hate worlds from the updated modal base. This third solution is in tension with a principle that Gillies claims characterizes modal bases generally, namely that they are *well-behaved*:

(*Well-behavedness*): For any world *w* and context *c*, a modal base $C^{c,w}$ is well-behaved iff
a. $w \in C^{c,w}$
b. if $w' \in C^{c,w}$, then $C^{c,w} \subseteq C^{c,w'}$

The first condition is simply that modal bases are *reflexive*; the second is that they are *euclidean*. Taken together, these conditions entail that if a modal base is well-behaved, then it is also *closed*:

(*Closedness*): for any $w' \in C^{c,w}$, $C^{c,w} = C^{c,w'}$.

This just means that no world inside a modal base $C^{c,w}$ opens (or closes) more possibilities than are open at *w*.[3]

For Thomason conditionals, *(Well-behavedness)* forecloses the possibility that, after we update the modal base with the antecedent, non-antecedent worlds are still accessible: trivially, $C^{c,w} \cap [[p]]^c \subseteq [[p]]^c$. Now, take any *w’* in $C^{c,w} \cap [[p]]^c$. For *~Kp* to be true at *w’* for an interlocutor *m*, $K^{m,w'}$ has to be compatible with *~p*; that is, $K^{m,w'} \not\subseteq [[p]]^c$. By our second assumption ($K \subseteq C$), ignorance gets transmitted upwards to the modal base at *w’*, so that $C^{c,w'} \not\subseteq [[p]]^c$. But if $C^{c,w} \cap [[p]]^c$ entails *p* whereas $C^{c,w'}$ does not, then $C^{c,w} \cap [[p]]^c \neq C^{c,w'}$. In other words, $C^{c,w} \cap [[p]]^c$ is not *closed*.

In sum: pointing out that the embedded knowledge claim in a Thomason conditional is to be evaluated at worlds in the updated modal base – from which non-antecedent worlds might be accessible – is of little help, since the well-behavedness of modal bases impedes that non-antecedent worlds are accessible from the updated modal base.

In reply to this, perhaps we could drop (*Well-behavedness*), or at least the euclidean condition – reflexivity is clearly out of question. But the euclidean condition is intuitive too, as it forbids that modal bases at worlds within a modal base *C* be more informative than *C* itself. The intuitive idea is that to consider it epistemically possible that one possesses more information than one actually possesses *just is* to possess more information. But that prevents us from accessing the non-antecedent possibilities that would make the consequent of a Thomason conditional true.

[3] *Proof* (again, taken literally – except for change of notation – from Gillies, 2010, p. 6): at any context *c* and world *w*, “suppose $w' \in C^{c,w}$. Consider any $w'' \in C^{c,w'}$. Since *C* is euclidean and $w' \in C^{c,w}$, $C^{c,w} \subseteq C^{c,w'}$. Since *C* is reflexive, $w \in C^{c,w}$ and thus $w \in C^{c,w'}$. Appeal to euclideanness again: since $w'' \in C^{c,w'}$, $C^{c,w'} \subseteq C^{c,w''}$; but $w \in C^{c,w'}$ and so $w \in C^{c,w''}$. And once more: since $w \in C^{c,w''}$, $C^{c,w''} \subseteq C^{c,w}$. And now reflexiveness: $w'' \in C^{c,w''}$ and so $w'' \in C^{c,w}$. (The inclusion in the other direction just is euclideanness.)”.

So once we take on board (*Closedness*), epistemic theories of indicatives are still in trouble in the face of Thomason conditionals. Here however, the plot takes an interesting turn. By incorporating (*Closedness*), we can show that Thomason conditionals entail that the relevant modal base is ignorant with respect to their antecedents. Truth-conditions for indicatives were given in terms of restricted epistemic necessity. We can represent this with an epistemic necessity operator $\Box_C$ (a box operator appropriately restricted by the relevant modal base *C*) and material implication ($\rightarrow$). That is, [[if *p*, *q*]] is equivalent to $\Box_C(p \rightarrow q)$. (*Closedness*) can be recast as the principle that $\Box_C p \rightarrow \Box_C \Box_C p$ (assuming reflexivity, which we are doing all along, this addition makes this system S4). Now take a Thomason example, whose structure is *if p*, *~Kp*. By our third assumption ($K \subseteq C$), *if p*, *~Kp* entails *if p*, $\sim\Box_C p$. By the epistemic theory as we have just stated it, [[if *p*, $\sim\Box_C p$]] is equivalent to $\Box_C(p \rightarrow \sim\Box_C p)$, which is equivalent to $\Box_C p \rightarrow \Box_C \sim\Box_C p$. Now suppose for *reductio* that $\Box_C p$. By $\Box_C p \rightarrow \Box_C \sim\Box_C p$ and *modus ponens*, $\Box_C \sim\Box_C p$; and by (*Closedness*) and *modus ponens*, $\Box_C \Box_C p$. By reflexivity from $\Box_C \sim\Box_C p$, we get $\sim\Box_C p$; and by reflexivity from $\Box_C \Box_C p$, we get $\Box_C p$ (i.e. if *p* is necessary at a world *i*, then *p* is true in all worlds accessible from *i*. By reflexivity, *i* is accessible from *i*, so *p* is true at *i*). But this is a contradiction, so by *reductio* we obtain $\sim\Box_C p$. In other words, Thomason conditionals entail that the relevant modal base is ignorant with respect to the antecedent.

In a one-person context (where the relevant modal base is just the speaker's knowledge) (1) entails that I do not know that my coworkers hate me. This is puzzling, since utterances of conditionals that entail their consequents are usually infelicitous (obvious examples are conditionals with a tautology in their consequent: *if I get home early, either I will eat or I will not*). Thomason conditionals do not fit this bill – their consequents are not tautologous and they do not sound infelicitous.

In sum, we have made a surprising set of observations: we have considered three ways of escaping the problem caused by Thomason conditionals, and all of them are unsatisfying. The first one was to treat the antecedents of Thomason conditionals as satisfying a presupposition in their consequent. But we saw that this solution could not be extended to Thomason conditionals whose consequents lack such a presupposition. The second solution was to give up the connection between indicatives and epistemic possibilities. But given the theoretical virtues of that view, that seemed a high price for a small payoff. The final escape route was to allow modal bases indexed to worlds in the intersection of the initial modal base with the antecedent to access non-antecedent worlds. This, however, clashed with the principle that modal bases are *closed*. Furthermore, by taking (*Closedness*) on board it was shown that Thomason conditionals entail something very much like the ignorance claim in their consequents, which is unexpected. I want to argue that these observations make sense if we cease treating constructions like (1) as conditionals, and we start treating them as *un*-conditionals.

3. Thomason Conditionals are Unconditionals

We have observed that, assuming (*Closedness*) and ($K \subseteq C$), Thomason conditionals entail their consequents, which is not a normal thing for an indicative conditional to do. But who said that Thomason conditionals are *normal*? For one, sentences like (1) do not seem to welcome the kind of paraphrase that indicative conditionals are normally subjected to. Consider paraphrasing (1) with *only if*:

(8) My coworkers hate me only if I have absolutely no idea.

This paraphrase sounds very different from the original. Here's an attempt at saying why: conditionals express some sort of connection, causal, probabilistic or other, between their antecedent and consequent. That is what the attempted paraphrase (8) seems to convey: that there is some sort of dependency between my coworkers' hate and my ignorance. But on its most natural interpretations, (1) conveys no such dependency. Rather, an utterance of (1) is appropriate in just the kind of context described at the start, where the speaker just cannot read her coworkers. But no relevant connection between her ignorance and the truth of that possibility is conveyed by uttering (1).

Things get worse when we try paraphrase by contraposition:

(9) If I have any idea (i.e. *if I know*) that my coworkers hate me, then my coworkers do not hate me.

The antecedent in (9) entails that my coworkers hate me, but then the consequent denies it. This is a no-go.

The strange behavior of Thomason examples with respect to *only if* paraphrase and contraposition suggests that these constructions may not be indicatives at all. This possibility is also supported by the observation that the particle *if* can be interrogative instead of conditional. In such cases, it is interchangeable for *whether*:

(10) I am not sure if they fought.
(11) I am not sure whether they fought.

Note that one can place these *if/whether* clauses at the start of the sentence, thereby constructing what looks very much like a Thomason conditional:

(12) Whether they fought, I am not sure.
(13) If they fought, I am not sure.

We submit that this is what Thomason examples are: constructions whose structure resembles that of an indicative conditional, but whose antecedent clause is an interrogative clause. This means that those *if*-clauses have the same denotation as questions (Heim, 2000; Karttunen, 1977/2002), more specifically *yes-or-no* or *alternative* questions. Thus, a prominent option opens up, namely that constructions like (1) are *alternative unconditionals*, in the sense of Rawlins (2013).

Unconditionals are constructions like the following:

(14) Whoever bakes the cake, it will be delicious.
(15) Whatever he said, you should not feel bad.
(16) Whether or not Silvio comes, it will be fun.

Informally, unconditionals differ from indicative conditionals in the following sense: whereas indicative conditionals restrict the circumstances in which we evaluate the consequent, unconditionals force us to consider all the range of alternatives that the antecedent presents us with. That is, (14) can be paraphrased by saying that, if Philip bakes the cake, it will be delicious; if Yining bakes it, it will be delicious; if Paloma bakes it, it will be delicious... and so on for all the salient alternatives. It is thus natural to give a standard interrogative semantics (in terms of sets of possible answers) to unconditional adjuncts, and that is just what Rawlins defends. In the case of alternative unconditionals like (16), just like in the case of whether-questions, the alternatives are a proposition and its negation.

An interrogative clause effects a partition on the domain of possible worlds *W*, rearranging it in cells whose worlds agree on each answer to the question. For instance, the denotation of ‘Who framed Roger Rabbit’ is a partition of the domain of propositions, where the worlds in each cell of the partition agree on each possible answer to that question: in cell #1, every *w* is such that the proposition that Valiant framed him is true in *w*; in cell #2, every *w’* is such that the proposition that Judge Doom framed him is true in *w’*, etc. Alternative interrogative clauses, on the other hand, partition *W* in cells according to whether worlds provide a negative or positive answer to a yes-or-no question: so the denotation of ‘whether or not my coworkers hate me’ partitions *W* in two cells, one wherein every *w* is such that the proposition that my coworkers hate me is true in *w*, and another cell in which every *w’* is such that the proposition that my coworkers hate me is false in *w’*.[4]

Hence, the proposal is that the antecedent clause of a Thomason conditional has the semantic value of an alternative question, that is, a set of propositions whose members are a proposition and its negation. In a somewhat simplified manner (see Rawlins, 2013 for more details, especially for much compositional detail over which I’m skipping), truth-conditions for alternative unconditionals may be given as follows:

(*Alternative unconditionals*): $[[\text{if/whether or not } p, q]]^{c,w} = 1$ iff $C^{c,w} \cap [[\text{whether or not } p]]^{c} \subseteq [[q]]^{c}$

All we’re doing is swapping the denotation of the antecedent clause in the truth-conditions for indicative conditionals given at the outset for the denotation of an alternative question, whose content is the antecedent clause and its negation. Admittedly, this requires some formal violence, since a set of propositions is not the right type of object to be intersected with a set of worlds. However, recall that denotations of alternative questions are partitions of logical space. It is thus suggestive to interpret that what ought to be intersected with the modal base is just the set of worlds on which the partition is performed, that is, *W* (*minus* worlds that fail to satisfy any presuppositions that the alternative question may have, see n. 4).

The antecedent clause of an unconditional like (1) is no longer a proposition, but the set {my coworkers hate me, my coworkers don’t hate me}. Substituting this denotation type for the antecedent, the truth conditions for (1) look like this:

$[[(1)]]^{c,w} = 1$ iff $C^{c,w} \cap$ {my coworkers hate me, my coworkers don’t hate me}$\subseteq$ [[I have absolutely no idea]]^{c}.

The alternative question {my coworkers hate me, my coworkers don’t hate me} denotes a partition of *W*, and since $C^{c,w}$ is a subset of *W* (assuming that $C^{c,w}$ also satisfies the presuppositions of the antecedent clause), the intersection with respect to which we must evaluate the consequent is just $C^{c,w}$ itself. (1) comes out true just in case the modal base is a subset of or equal to worlds in which I’ve no idea of whether my coworkers hate me or not. Given our third assumption ($K \subseteq C$), every world in which I’ve no idea of whether my coworkers hate me is a world whose modal base is also ignorant as to whether my coworkers hate me.[5] Thus, what the truth conditions for (1) demand is that $C^{c,w}$ is a subset of or equal to

[4] A further question is whether such partition is exhaustive, that is, whether every $w \in W$ is in one or other cell of the partition. Intuitively, the answer will be negative if the interrogative clause carries a presupposition. In our case, for example, worlds in *W* where I am unemployed should not be in either cell of the partition corresponding to *whether my coworkers hate me*.

[5] *Proof*: at a context *c*, take any *w* such that I don’t know whether my coworkers hate me at <*c*,*w*>. Given our semantics for ‘knows’, this is true just in case some world *w’* in my knowledge state $K^{c,w}$ is such that my coworkers don’t hate me in *w’*. But by ($K \subseteq C$), if $w' \in K^{c,w}$ then $w' \in C^{c,w}$. Thus, $C^{c,w}$ is ignorant as to whether my coworkers hate me.

the set of w' such that $C^{c,w'}$ is ignorant as to whether my coworkers hate me. In other words, that for every $w' \in C^{c,w}$, some $w'' \in C^{c,w'}$ is such that my coworkers don't hate me in w''. Given (*Closedness*), this will be the case if and only if some $w''' \in C^{c,w}$ is such that my coworkers don't hate me in w''', as one would expect if Thomason examples express ignorance about their antecedent.

Assuming that Thomason examples are unconditionals explains the awkwardness of the paraphrases with *only if* and contraposition: as we mentioned, such paraphrases appeared to bring out a dependency between the antecedent and the consequent that is characteristic of conditionals. But such connection seemed to be absent from sentences like (1) on their most natural interpretation. If sentences like (1) are unconditionals however, that makes sense, since unconditionals are characterized precisely by the lack of such a connection: given that the consequent is true under any of the set of circumstances considered in the antecedent, an unconditional expresses the independence of the consequent with respect to the antecedent. Rawlins (2013, p. 112) calls this feature of unconditionals *relational indifference*. If Thomason examples are really unconditionals, then given relational indifference, it is to be expected that those paraphrases are odd.

Finally, unconditionals entail their consequents, so it is no wonder that Thomason examples do so as well. In this view, what a sentence like (1) turns out to express can be more fully paraphrased by saying that *if my coworkers hate me, I have absolutely no idea; if they do not, I have absolutely no idea either*; alternatively, that *whether or not my coworkers hate me, I have absolutely no idea.*

4. Conclusion

In the last section of this paper, we have defended the view that Thomason examples are not after all indicative conditionals, but *alternative unconditionals*. This view respects all the assumptions that we have made and requires no substantial changes in the epistemic theory of indicatives. It also accounts for a number of observations about these sentences, namely, the strangeness of *only if* paraphrase and contraposition, as well as the fact that, given our assumptions, Thomason examples entail their consequents. Finally, this view also respects the intuition that Thomason unconditionals are appropriate as a way of conveying ignorance about their antecedents.

REFERENCES

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Chalmers, D.J. & Hájek, A. (2007). Ramsey + Moore = God. *Analysis*. 67(2), 170-172.

Gibbard, A. (1981). Two Recent Theories of Conditionals. In W.L. Harper, R. Stalnaker, & G. Pearce (Eds.), *IFS: Conditionals, Belief, Decision, Chance and Time*. Dordrecht: Springer Netherlands, 211-247.

Gillies, A.S. (2010). Iffiness. *Semantics and Pragmatics* 3, 1-4.

Heim, I. (2000). Notes on Interrogative Semantics. [Lecture Notes from MIT]. http://www.sfs.uni-tuebingen.de/~astechow/Lehre/Wien/WienSS06/Heim/interrogatives.pdf.

Heim, I. (1983/2008). On the Projection Problem for Presuppositions. In P. Portner & B. Partee (Eds.), *Formal Semantics: The Essential Readings*. Oxford: Blackwell, 249-260.

Heim, I. & Kratzer, A. (1998). *Semantics in Generative Grammar*. Malden, MA: Blackwell.

Karttunen, L. (1977/2008). Syntax and Semantics of Questions. In P. Portner & B. Partee (Eds.), *Formal Semantics: The Essential Readings*. Oxford: Blackwell, 382-420.

Kratzer, A. (1986). Conditionals. *Chicago Linguistics Society*, 22(2), 1-15.

Ramsey, F.P. (1929/2013). *Foundations of Mathematics and Other Logical Essays*. London: Routledge.

Rawlins, K. (2013). (Un)conditionals. *Natural Language Semantics*, 21(2), 111-178.

Schlenker, P. (2011). Presupposition Projection: Two Theories of Local Contexts. Part I. *Language and Linguistics Compass*, 5(12), 848-857.

Stalnaker, R. (2014). *Context.* Oxford: Oxford University Press.
Stalnaker, R. (1975). Indicative conditionals. *Philosophia*, 5(3), 269-286.
Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.
Van Fraassen, B.C. (1980). Review of Brian Ellis, Rational Belief Systems. *Canadian Journal of Philosophy*, 10(3), 497-511.
Yalcin, S. (2007). Epistemic modals. *Mind*, 116(464), 983-1026.

PAOLO LABINAZ
University of Trieste
plabinaz@units.it

ASSERTION AND THE VARIETIES OF NORMS[1]

abstract

This paper challenges Cappelen's claim that the speech act category of assertion is to be discarded since there is no principled way to distinguish between utterances that are assertions and those that are not. Using an Austin-inspired framework, I will argue that, in opposition to his claim, there are some norms that can be seen to apply to assertion in a more intimate way than others, and these norms can be shown to be constitutive of it, since it is by means of them that we can account for specific defects pertaining to the making of an assertion, which the reliance on contextually variable norms (such as the conversational maxims of Grice to which he refers) does not seem able to do.

keywords

assertion, norms of assertion, conversational maxims, speech act theory, J.L. Austin

Phenomenology and Mind, n. 12 - 2017, pp. 110-120
DOI: 10.13128/Phe_Mi-21110
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

This paper challenges Herman Cappelen's claim that the speech act category of assertion, as characterized in the contemporary philosophical debate, is an empty category with no explanatory power, due to the fact that any attempt to identify a principled way to distinguish between those utterances that are assertions and those that are not is bound to fail (Cappelen, 2011). In particular, I intend to focus specifically on two of the arguments he provides against the most prominent theories of assertion, namely normative ones.[2] These arguments do indeed raise an interesting question about the ways in which normative theories deal with norms that are supposed to govern assertion essentially and how they relate to those governing assertion in a more extrinsic way, as conversational maxims do. My aim will be to show that, insofar as we want to defend the speech act category of assertion against arguments such as those proposed by Cappelen, we have to clarify the roles that different kinds of norms play in the making of assertions, and I shall argue that this can be done within an Austin-inspired framework for speech act analysis (see Sbisà, 2017).

2. The Debate Over the Nature of Assertion

In the last twenty years or so, there has been considerable debate over the nature of assertion among those who work in the philosophy of language and other related fields of research, such as epistemology.[3] At the core of this debate is Timothy Williamson's claim that the members of the speech act category of assertion are those speech acts whose performance is governed by the knowledge norm (henceforth KNA), which he articulates in the following way:[4]

> (KNA) One must: assert *p* only if one knows *p* (Williamson, 2000, p. 243).

[1] I would like to thank the editors of this special issue of *Phenomenology and Mind* for their support and patience, and also an anonymous reviewer for helpful comments and suggestions.

[2] Cappelen (2011, pp. 30-37) provides four arguments against normative theories that take assertion to be governed by a constitutive rule. For a detailed, critical examination of these arguments, see Montgomery (2014).

[3] As pointed out by a referee, the term 'assertion' has often been used ambiguously in this debate to refer either to all those utterances of indicative sentences involving the speaker's commitment to the truth of the proposition expressed, which Mitchell Green (2013) has grouped together under the label "the assertive family", or just to assertion proper. Here, when speaking of the nature of assertion, I am referring to what it is that qualifies a speech act to be an assertion proper. Later on, I will widen the discussion to other members of the assertive family, arguing that Cappelen's view cannot account for the differences among these types of act.

[4] (KNA) has been defended, though on different grounds, by Keith DeRose (2002), John Hawthorne (2004), and John Turri (2011), among others.

According to Williamson (2000), (KNA) gives the condition "on which a speaker has the *authority* to make an assertion. Thus asserting *p* without knowing *p* is doing something without having the authority to do it, like giving someone a command without having the authority to do so" (p. 257). On this ground, (KNA) is conceived as constitutive of the speech act of assertion, that is, it is essential to this act, governing necessarily every instance of it. According to Williamson, indeed, (KNA) specifies a necessary condition for properly asserting. It is worth noting that, although (KNA) governs every instance of the speech act of assertion, when one violates it, one is still making an assertion, even if this assertion is improper and so may be subject to criticism from its audience. Moreover, there are some cases in which, for example, due to the urgency of the situation, one can legitimately assert that *p*, even if one knows that one does not know *p*. In these cases, Williamson (2000) holds, (KNA) "can be overridden by other norms not specific to assertion" (p. 256), making one's assertion that *p* permissible.

Among those who support the constitutivity claim, but do not think that knowledge is the norm of assertion, there is wide disagreement about how the norm of assertion is to be characterized. Some hold that truth is the norm of assertion (Weiner, 2005), while others claim that it is governed by a norm of belief (Bach, 2008), others again conceive the norm of assertion as involving epistemic requirements other than knowledge, such as justified or rational belief (see, respectively, Kvanvig, 2009; and Douven, 2006), or that it is credible or reasonable for the speaker to believe the asserted content (Lackey, 2007), and the like. There are also those, such as Goldberg (2015), who argue that the norm of assertion is context-sensitive.[5]

3. Cappelen's No-Assertion View

Herman Cappelen (2011) has recently entered the debate over the nature of assertion, arguing in favor of what he calls "the No-assertion view", according to which the speech act of assertion can be discarded since it can be shown to be a theoretically useless category, at least as characterized by the most prominent theories of assertion, particularly the normative ones. According to him, all that we need to account for those utterances that philosophers traditionally place under this category is the notion of "saying", which refers here to the act of expressing a (complete) proposition by means of uttering a declarative sentence, together with some contextually variable norms (Cappelen, 2011, pp. 22-26). According to this view, "declarative" sayings are governed by norms that, since they vary across contexts, cannot be constitutive of the act of saying. Cappelen takes Grice's conversational maxims as a paradigmatic case of these norms because they "are norms that guide behavior, not norms that are essential to (or constitutive of) the behavior they guide" (Cappelen, 2011, p. 24). More specifically, since norms like these guide speech behavior, rather than being constitutive of it or of the conversational exchanges in which it is involved, they can never be stably associated with a certain context. Accordingly, given the contextually salient features of the situation, sayings may be evaluated by appealing to various norms, such as moral norms, norms of politeness, norms involved in communicative cooperation and so on.

While both Pro-assertion views (like those presented in the previous section) and Cappelen's No-assertion view concur that in uttering a declarative sentence one expresses a (complete) proposition, whatever it may be, what is at issue, in Cappelen's opinion, is the essentialist claim supported by normative theories of assertion, according to which an utterance counts as an assertion in virtue of its being governed by some norm. Accordingly, Cappelen holds that, insofar as the essentialist claim is shown to be false (regardless of its content), the No-assertion

[5] For an overview of the main theories involved in the debate over the nature of assertion, see Pagin (2007/2014).

view should be preferred to its main opponents. To this end, he gives four arguments, of which the first two support norm variability, while the other two regard our ordinary ways of speaking of unwarranted or insincere assertions and of reporting one's assertion, respectively. As mentioned in Section 1, I shall focus here on the first two arguments, since they purport to show that normative theories cannot account for the data supporting norm variability. Here I shall simply present the two arguments, while in Section 4.2 I shall proceed to discuss them in the light of the Austin-inspired framework.

The first argument is concerned with modal judgments about norm variability (Cappelen, 2011, pp. 30-32). Cappelen points out that, although Williamson (together with many other proponents of the norm of assertion) argues in favor of an essentialist claim about the norm governing the speech act of assertion, he provides no modal argument in its support, but merely aims to show that when making assertions, people actually conform to (KNA). Is Williamson justified in claiming that (KNA) *necessarily* governs every instance of assertion? If so, it would be impossible to conceive of assertion "as governed by a variety of norms" (Cappelen, 2011, p. 31). But according to Cappelen, this is easily conceivable. Think of Mia performing, by saying that *p*, a certain act E, corresponding to what the proponents of normative theories would take to be a paradigmatic case of assertion, and then ask yourself:

> Could Mia have done that, i.e. performed E, if the default assumption was that she only assert that *p* if she believes that *p*?
> Could Mia have done that, i.e. performed E, if the default assumption was that she assert *p* only if she is committed to defending *p* in response to objections?
> Could she have done that, i.e. performed E, if the default assumption was that she only assert *p* if *p* is true? (Cappelen, 2011, p. 31).

According to Cappelen, while proponents of normative theories, such as Williamson, would say that only in one case has Mia performed E, that is, in the situation in which the default assumption corresponds to their favored norm, in his view we can feel justified in holding that she has performed the act E in every one of these cases.[6] Consider now a case in which a game, that is, a paradigmatic case of a rule-governed activity, is involved:

> Could [Mia] have played that game, i.e. tennis, if serves were thrown by hand, without a racket, and no ball could be hit by a player unless she had a foot on one of the lines .... (fill out for completeness)? (Cappelen, 2011, p. 31).

According to Cappelen, it would be hard to hold that in this case she would be playing the same game. In his view, a comparison between the two cases makes it clear that "assertion is not like a game, and that it's not governed by rules" (Cappelen, 2011, pp. 31-32). If so, (KNA), or any other norm of assertion, cannot be conceived as constitutive of it, as there is no rule *necessarily* governing it. Accordingly, the proponents of assertion's norm have no principled way to distinguish between those sayings that they suppose to be assertions and those that are not.

The second argument focuses on the actual variability of the norms involved in the making of an assertion (Cappelen, 2011, pp. 32-35). In Cappelen's view, if it can be shown that such

[6] One reviewer has pointed out that here Cappelen is misrepresenting the point of view of normative theorists. Be that as it may, my focus will not be on Cappelen's reconstruction of their point of view, but on his claim (to be discussed below in Section 4.2) that regardless of what norm is in force, one can recognize that by saying *p* Mia has performed an act E corresponding to what normative theorists would take to be an assertion.

a variability actually exists, then those "declarative" sayings that are taken to be assertions cannot be said to be governed necessarily by some norm. Here, Cappelen is referring to the discussions among the proponents of the norm of assertion. As in any philosophical debate, when someone argues for a certain norm of assertion, someone else provides one or more cases that are presented as going against the view that such a norm governs every instance of that speech act and is thereby constitutive of it. In response to these cases, the proponent of the norm will present some defensive maneuvers to account for them, holding, for example, that "the case in question wasn't really an assertion, or it wasn't performed in a default context, or it has some kind of second order justification that blinds us to the first order violation" (Cappelen, 2011, p. 33). What matters here is that the wide variety of counterexamples presented in the debate over the content of norm's assertion provide overwhelming evidence that the same instance of assertion can be legitimately accounted for by appealing to one norm rather than another. According to Cappelen, this overwhelming evidence suggests that it is impossible to identify one single norm governing all cases of assertion.[7]

4. How Do Norms Relate to Assertion?

In this section, I will argue that, in opposition to Cappelen's No-Assertion view, there is no need to discard the speech act category of assertion and replace it with a "minimalist" account centered on the notion of saying. On the one hand, Cappelen's proposal can be shown to be unsatisfactory. Indeed, he gives too much weight to norms such as Grice's conversational maxims, holding that we can explain what people do in conversation and how they do that by appealing solely to those norms, but this appears to be debatable (Section 4.1). On the other hand, if we want to defend the speech act category of assertion against arguments such as those presented above, one promising route to follow could be that of referring to the Austin-inspired framework recently developed by Marina Sbisà (2017) to account for "the different roles [speech act norms] play in the dynamics of illocution" (p. 1).[8] On the basis of this framework, we can shed light on the relationship between norms which need to be observed in the making of an assertion and norms governing assertion in a more extrinsic way, as conversational maxims do (Section 4.2).

4.1. Sayings and Grice's Conversational Maxims

As seen above, according to Cappelen's No-assertion view, in uttering a declarative sentence one performs a saying, expressing a certain proposition. But, and this is fundamental to his "minimalist" account, there is no rule governing a saying: for this reason, sayings can be assessed in a variety of ways, depending on the contextually salient features of the situation. However, in accounting for controversial cases in which putative assertions are involved, the norms referred to by Cappelen are almost always Grice's conversational maxims, regarded by him as the paradigmatic case of norms which, while operating on sayings, are not constitutive of the acts being performed. Consider the case of the conjunction of the form "*p*, but I don't know that *p*", whose apparent wrongness is usually provided as evidence in support of (KNA) (see, e.g., Williamson, 2000, pp. 253-254). Cappelen (2011, pp. 37-40) points out that, insofar as its wrongness is assumed to support a constitutive rule of that kind, the same could be said of many other similar sentences. Consider the following two examples taken from a list of sentences proposed by him:

[7] Even if it could be argued that Cappelen's conclusion is a non-sequitur (as a reviewer has rightly observed), I will challenge his argument from another point of view (see Section 4.2).

[8] The page numbers of the quotations refer to the preprint version of Sbisà's paper available on her ResearchGate page (see *References*). The paper is in press and forthcoming in a theme issue of the *Poznan Studies in the Philosophy of the Sciences and the Humanities*.

(1) *p*, but I don't want you to believe that *p*
(2) *p*, but *p* is irrelevant to what we are talking about.

If we follow Williamson's line of argument, says Cappelen, each one of these sentences would provide evidence for a different constitutive norm of assertion, but this would be in stark contrast with the essentialist claim supported by normative theories of assertion. Moreover, Cappelen (2011) maintains that "the feelings of badness triggered by these conjunctions are easily explained by standard Gricean considerations" (p. 38). According to him, if a speaker uttered (2), she would directly violate the maxim of relevance, since, while presenting herself as conforming to it in the first conjunct ("*p*"), she would then be denying it in the second one ("but *p* is irrelevant to what we are talking about"). The same can be said of "*p*, but I don't know that *p*", except for the fact that here the speaker would be violating the maxim of quality (see Cappelen, 2011, p. 39). However, it seems to me that reliance on Grice's conversational maxims is not enough to account for the defects of speech behavior corresponding to the utterance of sentences such as those presented above. Note that already Austin (1975) pointed out that "in order to explain what can go wrong with statements we cannot just concentrate on the proposition involved" (p. 52), since what is required is a reference to the total situation in which they are made: however, he was referring not only to aspects external to the performed act, as Cappelen does by appealing to Grice's conversational maxims, but also to those aspects that make it possible. If both these kinds of aspects are taken into account, (1) and (2) can be shown to be radically different from "*p*, but I don't know that *p*" because, while in the latter *p* turns out to be infelicitous, in the former *p* merely appears to be put forward in an awkward or unusual way. Indeed, we can easily conceive of conversational patterns involving (1) and (2), but that would be very difficult in the case of "*p*, but I don't know that *p*". If during a conversation a speaker utters the sentence "The red pen is on the table", and then adds "...but what I have just said is irrelevant to what we are talking about", she appears only to be making explicit that what she has said is not relevant to the purposes of the conversation, but in so doing she does not seem to be questioning any aspect that is characteristic of the practice of asserting. Although more complex, something similar can be said as regards (1). There can be many reasons why a speaker, after making the assertion that *p*, might point out that she does not want the hearer to believe that *p*, but that does not mean that the utterance of *p* would result in an infelicitous assertion; it may instead count as an assertion followed by a warning ("I warn you that I don't want you to believe that *p*"). Note, *inter alia*, that uttering (1), as well as (2), can stimulate the hearer's interest in *p*: there would be nothing wrong if she were to challenge the speaker by asking "How do you know that *p*?", even if *p* is not relevant to the discussion, or the speaker does not wish the hearer to believe that *p*, respectively. However, if we focus on the second conjunct of both (1) and (2), a standard reaction to their being uttered may be a request for clarification, due to the fact that, as said above, their conjunction with *p* makes the utterance of the latter seem unusual or awkward. Indeed, the request for clarification will be not concerned with the act of asserting itself, but with the apparently inappropriate contribution the speaker has given to the conversation by uttering those sentences. For example, when a sentence such as (2) is uttered, a hearer could ask the speaker "So, why did you say that?": note that this is not a challenge to her asserting that *p*, but rather a request to explain why she has given this "unusual" conversational contribution. If, instead, a speaker asserts only that *p*, and *p* is irrelevant to the goal of conversation, she may be liable to criticism for violating the maxim of relevance ("Hey, what you have just said is not relevant to what we are talking about!"), and the criticism can be then followed, if necessary, by a request for clarification ("So, what do you mean by saying that *p*?"). Very different is the case of "*p*, but I don't know that *p*". When a sentence such as this is uttered, it looks as though something has gone wrong with what the first conjunct is designed to accomplish, which, insofar as this is an act of assertion, goes beyond the plain expression of a proposition. In particular, it is not only conversationally inappropriate to utter the two conjuncts together, but it looks as if uttering the second conjunct in some way questions what the speaker has done in uttering the first one. Indeed, if the speaker maintains that she does not know that *p*, how could she seriously assert that *p*? She could have made many other moves: for example, she could have made an assertive speech act not involving any claim to knowledge, such as a conjecture or a guess or the utterance of a hedged sentence, e.g. "I think that *p*" or "It may be that *p*". Regardless of the precise content of the norm of assertion, the second conjunct appears to question something that is at the very core of the practice of assertion. It seems to me, then, that contrary to Cappelen's view, making reference to Grice's maxims is not enough to account for this case: while uttering a sentence such as "*p*, but I don't know that *p*" involves a specific defect that in some way questions the act the first conjunct is designed to accomplish, that is, an assertion that *p*, (1) and (2) exhibit generic defects that have no impact on the act being performed by the utterance of the first conjunct. In a conversation, any linguistic move gives rise to certain expectations between the interlocutors, which are unrelated to the common goal of that conversation and so cannot be reduced to expectations of cooperativeness. If I choose to make an assertion, rather than a guess, and thereby utter a declarative sentence, my audience will form some expectations, which differ in some way from those raised by uttering "I guess that…", regardless of the goal(s) of the conversation in which I am involved. But the differences among these types of act do not seem to be accountable within Cappelen's framework because according to it, there cannot be rules that are constitutive of one or another assertive speech act. Moreover, the evaluation of speech behavior as appropriate or inappropriate to the goal of the conversation, as polite or impolite etc. also depends (in part) on it being recognized as a certain kind of illocutionary act. In Cappelen's framework, there is no need for the hearer to recognize the utterance of a declarative sentence as being an assertion: it is enough to recover the proposition expressed and evaluate this saying against the norms one assumes to be the most relevant, given the contextually salient features of the situation. But as I have tried to show, whether a declarative sentence is used to do something is dependent on more than just what the speaker says by means of uttering it, and this "more" (whatever it may be) cannot be explained merely by appealing to some contextually variable norms that are external to the act being performed.

4.2. Assertion and the Dynamics of Illocution

In order to defend the speech act category of assertion, it is not enough simply to question Cappelen's "minimalist" account. Two more issues still need to be addressed. On the one hand, we have to clarify the different functions served by those norms which are (or may be) involved in the making of an assertion. On the other hand, insofar as we want to show that there is a principled way to distinguish between those utterances that are assertions and those that are not, and that this can be characterized in normative terms, we also have to show why Cappelen's arguments should be rejected. To tackle these issues, I will rely on an Austin-inspired framework for speech act analysis, focusing on what Marina Sbisà calls "the dynamics of illocution", by which she refers to "the interactional mechanisms that make it possible for the utterance of one interlocutor to bring about an illocutionary effect, recognized by the other interlocutor" (Sbisà, 2017, p. 1).

Let us first consider Cappelen's argument for the actual variability of norms. In support of his claim, Cappelen offers a lengthy quote from an article by Janet Levin (2008), in which she argues that there is no single norm of assertion, but many norms that "are always pragmatically determined" (Levin, 2008, p. 371). Indeed, in her view, there are so many reasons why we can make assertions which, while not conforming to (KNA), can be legitimately regarded as normatively appropriate, that (KNA) itself could be conceived as one among the many norms locally governing assertion. Levin (2008) points out that an assertion can be made, among other things, “to establish authority, to demonstrate conviction, to encourage others to believe, to commit themselves publicly to a belief about which they themselves had been (perhaps irrationally) wavering” (p. 367). This leads her to conclude that when one makes an assertion, e.g., to demonstrate conviction, one’s audience understands that the assertion is being made for that reason, regardless of whether (KNA) has been respected or not, and that the same can be said for many other cases. Williamson might observe here that Levin’s examples correspond to non-default uses of declarative sentences. However, it is far from clear how to distinguish between default and non-default uses. A more viable reply can be found in the light of the Austin-inspired framework mentioned above. Indeed, the dynamics of illocution is centered on the bringing about of illocutionary effects, while Levin’s examples include consequential effects pertaining to the perlocutionary dimension of speech acts. Here I am relying on the received distinction between illocutions and perlocutions: illocutions are the kinds of acts done in saying something according to a procedure designed to have characteristic effects (Austin, 1975, p. 14), while perlocutions are the kinds of acts done by saying something amounting to the production of consequential effects (Austin, 1975, p. 101). For example, by means of asserting that *p*, I can convince the audience to believe that *p* or that I am very competent with regard to the subject matter of *p*. But the assertion being made in no way depends on the reasons for making it. The fact that an assertion has been made, as well as the conditions upon which its performance is dependent, differ in important ways from the reasons why one has made that assertion. There is thus a sharp distinction between the illocutionary and perlocutionary dimension of a speech act. Note that, in challenging Cappelen’s argument about the actual variability of norms, I have also delimited the field of illocutionary act dynamics, since strictly speaking, the reasons why one makes an assertion cannot be conceived as part of that dynamics.

Let us now turn, firstly, to the functions served by the norms that are (or may be) involved in the making of an assertion, and secondly, once this point is clarified, to Cappelen’s argument concerning modal judgments about norm variability. With reference to the dynamics of illocution, Sbisà (2017) distinguishes between three different kinds of norms: constitutive rules, maxims, and objective requirements. Here I will focus exclusively on the first two kinds of speech act norms, since they are more relevant to the issues raised by Cappelen. As to constitutive rules, Sbisà (2017) takes them to be those norms that “when complied with, enable us to perform the acts they define” (p. 1). In other words, success in the performance of an illocutionary act is dependent upon compliance with those norms. Contrary to appearances, Sbisà’s constitutive rules cannot easily be equated with those proposed by the supporters of mainstream normative theories. Indeed, while Williamson (2000, p. 239) holds that, insofar as (KNA) necessarily governs any instance of assertion, it cannot be conventional because it is neither contingent nor replaceable, according to Sbisà (2017), there are good reasons to regard constitutive rules as socially accepted norms, fixing “procedures or routines that are repeatable and recognizable from one occasion to another and whose function (the production of illocutionary effects) is only exercised against a background of intersubjective agreement” (p. 1; see also Witek, 2015). Accordingly, within this framework, any illocutionary act, be it an assertion or not, can be described as the execution of a procedure fixed by socially accepted norms which is designed to produce a conventional effect; its achievement depends on the interlocutors’ intersubjective agreement, be this explicit or tacit (Sbisà, 2007; 2009). The reference to tacit agreement is due to the fact that a hearer does not usually check whether the speaker’s utterance complies with constitutive rules; instead, compliance is presumed by default, unless there are specific reasons for doubting it. For example, a speaker who utters a simple declarative sentence, unless special conditions obtain, will be expected to be in a position to successfully make an assertion by means of it,[9] since she has used a linguistic form that makes the act she purports to perform recognizable. Within this framework, then, constitutive rules specify the initial conditions (as well as the appropriate steps) required to execute the procedure successfully: failure to comply with them leads "to failure in performing the act and therefore in bringing about its illocutionary effect" (Sbisà, 2017, p. 15). In the case of assertion, beyond requiring the utterance of a declarative sentence, constitutive rules fix the epistemic position that is required of a speaker for asserting successfully (i.e., the speaker's epistemic authority with respect to the content she purports to assert) and the circumstances in which the procedure for asserting can be invoked (see Labinaz and Sbisà, 2014).[10] As said above, if special conditions do not obtain, compliance with constitutive rules is presumed by default: the speaker and her audience are entitled to form certain expectations of each another connected with the correct execution of the procedure for asserting, which differ, for example, from the expectations raised by the correct execution of the procedure for performing other kinds of assertive speech acts (e.g., a guess or a conjecture). When an assertion is performed successfully, the bringing about of its illocutionary effect, which can be characterized in terms of commitments and entitlements, is also expected to be obtained. Now consider once more "*p*, but I don't know that *p*". In the light of our Austin-inspired framework, we can say that, insofar as it states that a constitutive rule of the act of asserting is violated, the utterance of the second conjunct impedes the generation of the commitments and entitlements that a speaker and her audience would expect of one another when a (successful) assertion is made. What happens then is that it is not clear what the speaker has done in uttering that sentence. On the one hand, one could suppose that the speaker has intentionally made an unsuccessful assertion, pointing out her lack of relevant knowledge for some particular reason (as observed in the previous section, it is very difficult to conceive of a conversational pattern involving the utterance of "*p*, but I don't know that *p*"). The reason why she has made this unsuccessful assertion might be explained in terms of intended perlocutionary effects. On the other hand, in uttering that sentence, a speaker could be regarded (in a more charitable way) as executing the procedure for performing another kind of assertive speech act, such as a conjecture, whose successful performance does not require the speaker being in the same epistemic position as appropriate for assertion. Something different happens when a violation of a conversational maxim, even a non-repairable one, is discovered. In these cases, there is a penalty for the speaker, which consists of criticism, damage to reputation, or loss of reliability, but which in no way questions the achievement of the conventional effect characteristic of assertion. As noted by Sbisà (2017, p. 4), "[t]his does not make sense with conversations, since there is no single conventional effect to be associated with conversation as such".[11] Speaking of conversation brings us to the second kind of norms applying to speech acts, maxims. Maxims can be seen as half-way between constitutive rules (in the sense described above) and Grice's conversational maxims: despite being in some sense associated with the procedure for performing the speech act, they are not conventional, but are based on rational motivations, just like Grice's conversational maxims.

[9] By 'successfully' I mean here that there is no misfire (in the Austinian sense), but there may still be an abuse of the procedure of asserting.

[10] Since it is not a goal of this paper to determine the precise content of these constitutive rules, I will remain neutral on how their content should be articulated.

[11] As pointed out by a reviewer, even if it is true that there is no single conventional effect to be associated with conversations as such, we can still conceive of conversations as sequences of speech acts having some aim or other (such as to answer a question or settle on a plan) (see, in particular, Green, 2017).

Indeed, according to Sbisà (2017, p. 4), both maxims applying to speech acts and conversational maxims can be characterized as advice that must be followed in the attempt to optimize one's communicative behavior. In particular, compliance with these maxims should lead to the performance of "a speech act optimally satisfactory in the perspective of the participants" (Sbisà, 2017, p. 15). In the case of such an "optimal" assertion, maxims require that the speaker should (at least) believe that what she has asserted is true, and that her subsequent behavior (be this verbal or non-verbal) is consistent with it. As said above, although these maxims are associated with the procedure for performing the speech act, and thus we normally expect from a speaker making an assertion that she will be sincere and consistent in her subsequent behavior, their violation does not lead to questioning the assertion's successfulness, at least insofar as its constitutive rules are respected. As in the case of the violation of a conversational maxim, here too, if the speaker turns out not to be complying with one of the maxims, she will be blamed for this (with a resulting loss of credibility or reliability), but these consequences have no impact on the act being performed. For example, when someone lies by asserting that *p*, the act being performed is still recognized as an assertion: what may happen is that, if she is recognized as not believing what she has asserted to be true, her assertion is taken to be insincere and therefore subject to criticism. We can say that the speaker has made a successful assertion, which is however not an optimal one; that is to say, it does not amount to an optimal speech act performance from the perspective of the participants in the conversation (which would require full cooperativity).

We can now return to Cappelen and to his argument concerning modal judgments about norm variability. Remember that this argument aims to show that, contrary to Williamson's claim, according to which assertion does not exist without being governed by (KNA), because that rule necessarily governs any instance of this speech act, we can conceive of cases in which, regardless of which of the norms mentioned in Section 3 above is in force, one can recognize that by saying *p* a speaker has performed an act corresponding to what normative theorists would take to be an assertion. But the three norms to which Cappelen refers highlight aspects that are actually involved in the making of an assertion, i.e., when taking an utterance to be an assertion, we normally expect that the asserter believes what she asserts, that she is committed to defending the asserted content if appropriately challenged, and that the asserted content is true. That is why at first sight one might be convinced by Cappelen's argument that each of the norms considered can be regarded as an equally plausible candidate for the norm of assertion. But this conclusion does not hold. As our Austin-inspired perspective makes clear, mere compliance with any of those rules, taken individually, does not suffice to outline the complete procedure of asserting, to which they all in fact contribute. Moreover, even compliance with them all fails to put the speaker in a position to successfully make an assertion, since none of them makes any reference to the speaker's epistemic authority with respect to the content she purports to assert.

5. Concluding Remarks

In this paper, I have argued that, contrary to Cappelen's No-assertion view, some norms can be seen as applying to assertion in a more intimate way than others, because of the role they play in what Sbisà has called "the dynamics of illocution". Among these norms, while constitutive rules are those the observance of which makes assertion possible, the compliance with maxims makes the performed assertion optimally satisfactory in the perspective of the participants in the conversation. As I have tried to show, distinguishing between these two kinds of norms can help us to account for specific defects pertaining to the making of an assertion, which, contrary to what is claimed by Cappelen, reliance on Grice's conversational maxims alone does not seem able to do. Clearly, I have only laid the basis here for a deeper analysis of the role that different kinds of norms may play in the making of assertions. Many issues still need to be tackled, such as the relationship between the maxims involved in the dynamics of illocution and Grice's conversational ones, particularly with reference to the maxim of quality, which appears to have a special relationship with the making of assertions (see, e.g., Goldberg, 2015). For the moment, however, I hope to have shown that, thanks to the Austin-inspired framework presented above, we can still conceive of assertion as a category to be accounted for in speech act-theoretical terms.

REFERENCES

Austin, J. L. (1975). *How to Do Things with Words* (2nd ed.). Oxford: Oxford University Press.
Bach, K. (2008). Applying pragmatics to epistemology. *Philosophical Issues*, 18, 68-88.
Cappelen, H. (2011). Against Assertion. In J. Brown & H. Cappelen (Eds.), *Assertion: New Philosophical Essays*. Oxford: Oxford University Press, 21-48.
DeRose, K. (2002). Assertion, Knowledge, and Context. *Philosophical Review*, 111, 167-203.
Douven, I. (2006). Assertion, Knowledge, and Rational Credibility. *Philosophical Review*, 115, 449-485.
Goldberg, S. C. (2015). *Assertion: The Philosophical Significance of a Speech Act*. Oxford: Oxford University Press.
Green, M.S. (2013). Assertions. In M. Sbisà & K. Turner (Eds.), *Handbook of Pragmatics. Vol. II: Pragmatics of Speech Actions*. Berlin: de Gruyter, 387-410.
Green, M.S. (2017). Conversation and common ground. *Philosophical Studies*, 174, 1587-1604.
Hawthorne, J. (2004). *Knowledge and Lotteries*. Oxford: Oxford University Press.
Kvanvig, J. L. (2009). Assertions, Knowledge, and Lotteries. In P. Greenough & D. Pritchard (Eds.), *Williamson on Knowledge*. Oxford: Oxford University Press, 140-160.
Labinaz, P., & Sbisà, M. (2014). Certainty and Uncertainty in Assertive Speech Acts. In A. Zuczkowski, R. Bongelli, I. Riccioni & C. Canestrari (Eds.), *Communicating Certainty and Uncertainty in Medical, Supportive and Scientific Contexts*. Amsterdam: Benjamins, 31-58.
Lackey, J. (2007). Norms of Assertion. *Noûs*, 41, 594-626.
Levin, J. (2008). Assertion, Practical Reason, and Pragmatic Theories of Knowledge. *Philosophy and Phenomenological Research*, 76, 359-384.
Montgomery, B. (2014). In Defense of Assertion. *Philosophical Studies*, 171, 313-326.
Pagin, P. (2007/2014). Assertion. In E. N. Zalta (Eds.), *The Stanford Encyclopedia of Philosophy* (last version: Spring 2014 Edition). http://plato.stanford.edu/archives/spr2014/entries/assertion/.
Sbisà, M. (2007). How to Read Austin. *Pragmatics*, 17, 461-473.
Sbisà, M. (2009). Uptake and Conventionality in Illocution. *Lodz Papers in Pragmatics*, 5, 33-52.
Sbisà, M. (2017). Varieties of Speech Act Norms. Forthcoming in *Poznan Studies in the Philosophy of the Sciences and the Humanities*. www.researchgate.net/profile/Marina_Sbisa/publications.
Turri, J. (2011). The Express Knowledge Account of Assertion. *Australasian Journal of Philosophy*, 89, 37-45.
Weiner, M. (2005). Must We Know What We Say?. *Philosophical Review*, 114, 227-251.
Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
Witek, M. (2015). An Interactional Account of Illocutionary Practice. *Language Sciences*, 47, 43-55.

ENRICO CIPRIANI
University of Turin
enrico.cipriani@edu.unito.it,
ciprianienricoec@gmail.com

CHOMSKY ON ANALYTIC AND NECESSARY PROPOSITIONS[1]

abstract

My aim is to critically discuss Chomsky's position concerning the analytic-synthetic distinction and necessary propositions. To do so, I present Chomsky's objection to Quine's criticism of the analytic-synthetic distinction, and I point out that Chomsky's defense of such a distinction can be efficacious only under the assumption of conceptual innateness. I then focus on Chomsky's analysis of necessary propositions. In particular, I present Chomsky's objection to Kripke's essentialism, and Chomsky's hypothesis that the distinction between necessary and contingent truths is determined by the structure of the conceptual system and its relations with other systems of common-sense understanding. I highlight that this hypothesis is not compatible with Chomsky's own objection to Kripke.

keywords

Analytic-Synthetic distinction, "humbler" notion of analyticity, Innateness Hypothesis, essential properties, conceivability and possibility, categorization

Phenomenology and Mind, n. 12 - 2017, pp. 122-131
DOI: 10.13128/Phe_Mi-21111
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

CHOMSKY ON ANALYTIC AND NECESSARY PROPOSITIONS

1. Chomsky (2000, ch. 3) presents a criticism of Quine's (1951) thesis against the analytic-synthetic distinction. According to Chomsky, we have strong evidence that some analytic sentences hold in natural language, and it is plausible to assume that such sentences are a by-product of the structure of the innate conceptual system. Chomsky (2000, p. 47) mentions the semantic connection between 'to kill' and 'to die', entailing "a qualitative distinction – determined by the language itself – between the sentences 'John killed Bill, so Bill is dead' and 'John killed Bill, so John is dead'", and the relation of referential dependence adopted in Binding Theory (Chomsky, 2007, 2008; Hinzen, 2016): "it would be difficult to find a study of referential dependence in natural language that does not conclude that the language itself determines that" the relation of referential dependence holds between 'Mary' and 'herself' in "Mary expects to feed herself", and that does not hold between the same constituents in "I wonder who Mary expects to feed herself" (Chomsky, 2000, p. 47). Chomsky uses the poverty of stimulus argument to argue that the a priori knowledge ("cognizing") of the syntactic-semantic relations mentioned above is due to innate conceptual connections:

> Acquisition of lexical items poses what is sometimes called "Plato's problem" in a very sharp form. [...] At peak periods of language acquisition, children are acquiring ("learning") many words a day, perhaps a dozen or more, meaning that they are acquiring words on very few exposures, even just one. This could appear to indicate that the concepts are already available, with much or all of their intricacy and structure predetermined, and that the child's task is to assign labels to concepts, as might be done with limited evidence given sufficiently rich innate structure. And these conceptual structures appear to yield semantic connections of a kind that will, in particular, induce an analytic-synthetic distinction, as a matter of empirical fact.
>
> To the extent that anything is understood about lexical items and their nature, it seems that they are based on conceptual structures of a specific and closely integrated type. It has been argued plausibly that concepts of a locational nature – including goal and source of action, object moved, etc. – enter widely into lexical structure, often in quite abstract ways. In addition, notions like actor, recipient of action, instrument, event,

[1] I am grateful to Prof. Diego Marconi and Prof. Giorgio Graffi for their suggestions.

> intention, causation and others are pervasive elements of lexical structure, with their specific properties and interrelations (Chomsky, 2000, pp. 61-62).

These concepts (goal, etc.), that Jiulius Moravcsik has interpreted as "generative factors" (see Moravcsik, 1975, 1990), have a crucial role in the formal analysis of many syntactic-semantic phenomena, like causative (see Dowty, 1979; Levin & Malka, 1995; Parsons, 1990; Ramchand, 2008; Schäfer, 2009) and locative alternation (see Baker, 1997; Grimshaw, 1990; Jackendoff, 1990; Levin, 1993; Talmy, 2000), meaning transfer, and, more generally, in the generative analysis of lexical meaning (see Larson and Segal, 1995; Pustejovsky, 1995). Pinker (1989, 1994) has advocated from an empirical point of view that such factors are innate elements of the mind, having important roles in language acquisition and language processing.
According to Chomsky, the interrelations of such factors are responsible for the distinction between truths of meaning and truths of fact. Although he does not explain exactly what he means by these notions, Chomsky mentions some examples of truths of meaning. If I state that John has been killed, then I know a priori that John is dead: this inference is a truth of meaning, that depends uniquely on the meaning of words. The same is true when we consider that 'John persuaded Bill to go to the university' *necessarily* entails that Bill went *sua sponte* to the university, and was not obliged. These are just some of the many examples that can highlight a strict network of semantic connections. Since the child is not explicitly instructed about such connections, but acquires them during language acquisition, Chomsky (2000) suggests that the child

> approaches language with an intuitive understanding of concepts involving intending, causation, goal of action, event, and so on; furthermore, it must be that the child places the words that are heard in a nexus that is permitted by the principles of universal grammar, which provide the framework for thought and language, and are common to human languages as systems that enter into various aspects of human life. These elements also appear to enter into an integrated "conceptual scheme" a component of the initial state of the language faculty that is fleshed out in specific ways, with predetermined scope and limits, in the course of language growth, one aspect of cognitive development (p. 62).

The innateness interpretation of the semantic connections just mentioned induces Chomsky to argue that

> one of the central conclusions of modern philosophy is rather dubious: namely, the contention – often held to have been established by work of Quine and others – that one can make no principled distinction between questions of fact and questions of meaning, that it is a matter of more or less deeply held belief. This conclusion has been supported by reflection on an artificially narrow class of examples; among them concepts that have little or no relational structure. In the case of such sentences as 'cats are animals' for example, it is not easy to find evidence to decide whether the sentence is true as a matter of meaning or fact, or whether there is an answer to the question in this case, and there has been much inconclusive controversy about the matter. When we turn to concepts with an inherent relational structure such as *persuade* or *chase*, or to more complex syntactic constructions such as those exhibiting referential dependence or causative and relative constructions, then it seems that semantic connections are readily discerned (p. 63).

The innateness perspective is *necessary* to enable Chomsky to defend the analytic-synthetic distinction from Quine's criticism. Indeed, the fact that we are able to identify some truths of meaning in natural language *does not constitute* a sufficient reason to reject Quine's criticism: although his purpose was to show "that there was no truths that are both immune from revision based on experience and directly accessible to the linguistically informed mind" (Marconi, 1997, p. 33), Quine himself admitted that some sentences are trivial cases of analyticity, as Putnam (1975a) had already noticed. Thus, Quine (1986, p. 208) argued that it was possible to maintain a "humbler" notion of analyticity, restricted to the domain of "empirical semantics". He observed that there are sentences "that we learn to recognize as true in the very process of learning one or another of the component words": among them, we find such sentences as 'Bachelors are unmarried', and logical words ('if', 'or', etc.). Quine concludes that "some truths are learned by learning words", and this is the "worthwhile insight" that makes up the only content of the notion of analyticity as "an intelligible and reasonable" notion (pp. 94-95) (see Marconi, 1997, pp. 33-34 for discussion). Thus, to point out, as Chomsky does, that we have intuitive knowledge of some truths and semantic relations is not a sufficient reason to prefer the innateness hypothesis to other explanations; an equally plausible hypothesis is that our intuitive knowledge is determined by the fact that such truths and connections "are learned by learning words" (perhaps, this is not true for referential dependence relations, that, as such, belong however to the domain of syntax). Therefore the innateness hypothesis is not the only possible (the best) explanation of why some analytic sentences (expressing truths of meaning) hold in natural language; it is instead a necessary assumption to justify the existence of "*not* humbler" analytic sentences. Arguing that analytic truths (truths of meaning) are determined by innate conceptual schemas allows saving the notion of analytic truth as it was interpreted by logical positivists. Indeed, logical empiricists identified analytic sentences "with (a) necessary truths, (b) a priori truths, (c) unrevisable pieces of knowledge, and (d) sentences constitutive of (lexical) semantic competence" (Marconi, 1997, p. 30). It is clear that these conditions are satisfied by those sentences that are taken to reflect innate conceptual relations.

The notion of analyticity advocated by logical positivists has been criticized not only by Quine. Kripke (1980) has shown that (a) and (b) do not extensionally coincide, and Putnam (1975a) has pointed out that the same is true in the case of (c) and (d): for example, 'Cats are animals' is part of lexical semantic competence, but is revisable, while '37 is the thirteenth prime number' is unrevisable but is not part of the lexical semantic competence of a common speaker (see Putnam's linguistic labor division theory: Putnam, 1975b). Although it is not relevant here, Chomsky's judgment of Putnam's theory can be summarized as follows: roughly, according to Chomsky, Putnam's conclusions are beyond linguistic theory, namely they do not belong to the naturalistic investigation of language. Chomsky supports this criticism pointing out that the notions Putnam uses (such as that of linguistic community, of rule, etc.) are common sense notions and they are not scientific nor technical. Let's now consider Chomsky's criticism to Kripke.

2. In *Naming and Necessity* Kripke shows that a priori truths and necessary truths not always coincide. Thus, he distinguishes the epistemic level (a priori *vs.* a posteriori) from the metaphysical level (necessary *vs.* contingent): in this perspective, we can distinguish between *a priori necessary truths* (i.e., 'Hesperus is Hesperus') and *a posteriori necessary truths* (i.e., 'Hesperus is Phosphorus', under the theory of rigid-designation for proper names). Among the latter, we can include those truths which depend on the essential properties of the objects, namely on the properties that objects necessarily possess in virtue of their metaphysical essence (see Casalegno, 1997, pp. 237-238). Kripke provides some examples of essential properties: if an individual is a human being in the actual world, then she is a human being in all possible worlds. Thus, if we consider Richard Nixon, we can assume that there are possible worlds where he did not win the 1969 elections, but there is no possible world where he is not a human being; furthermore, since being a human being necessarily entails being an animate object, then we cannot imagine a world where Nixon is a human being but is not an animate object:

> If we can't imagine a possible world in which Nixon doesn't have a certain property, then it's a necessary condition of someone being Nixon. Or a necessary property of Nixon that he [has] that property. For example, supposing Nixon is in fact a human being, it would seem that we cannot think of a possible counterfactual situation in which he was, say, an inanimate object; perhaps it is not even possible for him not to have been a human being. Then it will be a necessary fact about Nixon that in all possible worlds where he exists at all, he is human or anyway he is not an inanimate object (Kripke, 1980, p. 45).

Kripke's hypothesis is based on the notion of conceivability. According to Kripke, conceivability corresponds to logical possibility; thus, if I cannot imagine a world where Nixon, *this* very person, is not a human being, then it is logically impossible for Nixon not to be a human being, namely he is necessarily a human being. If this is so, then 'Nixon is a human being' is a necessary a posteriori truth.

In *Reflections on Language* (but see also Chomsky, 2000, pp. 60-61), Chomsky argues that the notion of essential properties can be grounded in the connections holding within the common-sense understanding and the language system:

> The necessity of this statement ['the person Nixon is an animate object'] follows without any attribution of necessary properties to individuals apart from their designation. [...] The necessary truth of [this statement] is a consequence of the necessary truth of the statement that people are animate objects. This necessary truth may be grounded in a necessary connection between categories of common-sense understanding, or an analytic connection between the linguistic term "person" and "animate". Under any of these assumptions, we need not suppose that an essential property is assigned to an individual, Nixon, apart from the way he is named or the category of common-sense understanding to which he is assigned (Chomsky, 1975, p. 47).

This kind of explanation is used by Chomsky to discuss another example of essential property that Kripke advocates. In *Naming and Necessity*, Kripke argues that an essential property of objects is their material origin: "if a material objects has its origin from a certain hunk of matter, it could not have its origin in any other matter" (p. 114, fn. 56). A consequence of this stipulation, Kripke says, is that it is not possible to imagine a world where an individual has parents other than her actual ones. In other words, an individual would not be the same person if she had different parents. For example, it is not possible to imagine a world where Queen Elizabeth II, *that* person, is not the daughter of George VI and Elizabeth Bowes-Lyon. Chomsky criticizes Kripke's argument in two ways. Firstly, he highlights that such intuitions can differ from speaker to speaker:

> My own intuition differs about [this] example. Thus, it does not seem to me that a logical problem would arise if Queen Elizabeth II were to write a fictionalized autobiography in which she, this very person, had different parents; we might, I think,

> take this as a description of the history of this person in a different "possible world" a description of a possible state of this world with the very objects that appear in it (Chomsky, 1975, p. 48).

Secondly, he argues that *even if we assumed* that having specific parents is an essential properties of the Queen, then anyway this stipulation would derive from our categorization and designation of the entity in the systems of common-sense understanding:

> But suppose that it is not so, and that having a particular origin is an "essential property". Is it, then, an essential property of the thing itself, apart from its designation or categorization in common-sense understanding? I think not. We name the entity with a personal name, "Queen Elizabeth II". We assign it to a category of common-sense understanding, Person. It might be [...] that an object taken to be a person could not be the same person if it had parents other than its own. If so, this is a property of the conceptual system of common-sense understanding, and perhaps also of the related system of language; it is a property of the concept Person. Given the cognitive structure, we can distinguish necessary and contingent properties. We can distinguish between what might have been true of an object categorized and designated within these systems, and what could not have been otherwise. The intuitive force of the argument for essential properties seems to derive from the system of language in which the name is placed and the system of common-sense understanding, with its structure, in which the object is located (Chomsky, 1975, p. 49).

The notion of common-sense understanding was used by Chomsky in the 1970s and 1980s to denote the cognitive systems interacting with the mental organ of faculty of language (which is responsible for language acquisition). This is the definition of common-sense understanding provided in *Rules and Representations*:

> One basic element of what is loosely called "knowledge of language" [...] is knowledge of grammar, now analyzed in terms of a certain structure and rules, principles and representations in the mind. This grammar generates paired representations of sound and meaning, along with much else. It thus provides at least partial knowledge of sound-meaning relations. A fuller account of knowledge of language will consider the interactions of grammar and other cognitive systems, specially, the systems of conceptual structures and pragmatic competence, and perhaps others, for example, systems of knowledge and belief that enter into what we might call "common sense understanding" (Chomsky, 1980, p. 92).

It is reasonable to interpret common sense understanding as including the innate conceptual systems plus other systems that process extralinguistic knowledge, such as beliefs, assumptions about the world, intentions, and others. We can track the same idea in later works, in the claim that the generative procedure of the faculty of language generates structural descriptions that are mapped to the articulatory-perceptual systems and to the conceptual-*intentional* systems, where these latter systems include not only the conceptual system, but also other cognitive systems, whatever may be, which process extralinguistic factors such as beliefs, etc. (see Chomsky, 2000, 2007, 2008). Thus, the conceptual-intentional systems are taken to be responsible both for the semantic interpretation and for the pragmatic one.

Now, Chomsky is perhaps right in saying that no logical problem arises when we imagine the counterfactual situation where the Queen, this very person, has different parents: assuming that conceivability corresponds to logical possibility, if we can imagine that (conceive the situation where) the Queen has different parents, then it is *a fortiori* logically possible that she has different parents (thus, no logical problem arises, as Chomsky states). This objection is relevant for Kripke's theory, and deals with the problematic relation between conceivability and possibility, a well-known topic in philosophy (see Gendler & Hawthorne, 2002). Chomsky's objection is useful to highlight a problematic assumption of Kripke's theory, namely that we know a priori that a truth concerning an essential property is a necessary truth (if P concerns essential properties and P is true, then *a priori* P → Nec P). Consider the way in which Marconi (2010, pp. 143-144) describes Kripke's theory:

> For [Kripke], there are necessary features of the world that do not depend on our description of the world, but on the nature of things. As such, they need to be discovered by investigating nature [...]. They are, in this sense, *a posteriori*. More precisely, that the world has such features – for example, that salt is NaCI – can only be known *a posteriori*, if at all; whereas we know *a priori*, thanks to philosophical argument and, perhaps, intuition, that such features are necessary features (that salt is necessarily NaCI if it is NaCI at all).
>
> [...] for Kripke some necessary facts cannot be known *a priori* for it cannot be known *a priori* that there are such facts. Once it is established that they are facts, it is indeed *a priori* that they are necessary: once we know the world has such and such features, it follows *a priori* that they are necessary features ('p → Nec p' is *a priori*).

That said, a possible objection to Kripke (which is in a sense presupposed by Chomsky's objection) is that it is not sure that "philosophical argument(s)" and "intuition(s)" are sufficient to conclude that a truth concerning essential properties is a necessary truth. On the other hand, however, if we assume, as Chomsky seems to do, that the conceptual system is innate and, as such, common to all human beings (constitutional uniformity of individuals: see Graffi, 2001, p. 46), and if we assume moreover that our intuitions about essential properties are determined by the structure of the conceptual system, then we should reasonably conclude that our intuitions should be uniform across individuals and it should be impossible to have divergent judgments about essential properties and necessary a posteriori truths. This consequence *is not* taken into account, however, by Chomsky, who bases his criticism of Kripke on the fact that his intuition differs from Kripke's. This is my main objection to Chomsky's explanation of essential properties. Let's consider it.

Chomsky's words can be interpreted as stating that our intuitions about essential properties and necessary a posteriori truths are determined by the conceptual scheme and by other cognitive systems that are involved in our interpretation of language and world. Through this explanation, Chomsky seems to adhere to a "long philosophical tradition in modern philosophy, stemming from Kant" according to which "necessary features of the world are a by-product of our conceptual scheme (or schemes)" since "we regard as necessary what we could not experience, or conceive, or linguistically describe as being otherwise. Necessity does not inhere in the things themselves: it is projected onto them by us" (Marconi, 2010, p. 141). Kant (1787, p. 239) argued that the "categories of modality [...] only express the relation of the concept to the faculty of knowledge": according to Kant, "to say that something is possible, or actual, or necessary amounts to saying that its concept bears a certain relation to the faculty of knowledge" (Marconi, 2010, p. 141). In particular, with respect to modality Kant said that "that which in its connection with the actual is determined in accordance with universal conditions of experience, is (that is, exists as) necessary" (Kant, 1787, p. 239).

This perspective was adopted also by the later Wittgenstein, who argued that necessity is based on our "form of representation". Philosophers endorsing this perspective did not deny that "there are necessary features of the world" (such as that necessarily every thing is identical to itself); they insisted on the fact that "the necessity of such features comes from how we experience, or linguistically describe, or conceptualize the world, not from the world itself" (Marconi, 2010, p. 141). Kripke stood against this tradition: according to him, necessary features are not determined by the structure of the conceptual scheme, but by the nature of things. Thus, some necessary truths, such as mathematical or analytic truths, are known a priori, while others are known a posteriori.

If we consider how Chomsky (2000, pp. 41-42) discusses Kripke's intuition that "Nixon would be the same person if he had not been elected President of the USA in 1968, while he would not be the same person if he were not a person at all", it is clear that Chomsky follows in a sense the Kantian tradition: "doubtless there is an intuitive difference" between these intuitions, but "that follows from the fact that Nixon is a personal name, offering a way of referring to Nixon as a person; it has no metaphysical significance". Indeed, "if we abstract from the perspective provided by natural language, which appears to have no pure names in the logician's sense", then "intuitions collapse".

The problem in Chomsky's explanation can be understood by considering how Chomsky explains why we are prepared to say that Nixon would not be the same person if he was not a human being, while he would be if he had not won the 1969 elections. Chomsky writes that the "necessary truth" of this statement "is a consequence of the necessary truth of the statement that people are animate objects. This necessary truth may be grounded in a necessary connection between categories of common-sense understanding, or an analytic connection between the linguistic term 'person' and 'animate'" (Chomsky, 1975, p. 47).

Now, if our judgments about necessary truths are determined by *necessary* connections inside the categories of common-sense understanding (reflected in connections inside the language system), then our judgments should be a priori, and, as such, uniformly shared by all speakers: if speakers had divergent judgments about an allegedly necessary truth, this would entail that such truth is in fact not necessary. Thus, when Chomsky notices how his own intuition about the property of having certain parents differs from Kripke's, he should conclude that this property cannot be an essential property; *instead*, he argues that if we assume, contrarily to his own intuition ("but suppose that it is not so"), that having specific parents is an essential property, then this intuition should be explained by focusing on the structure of "the conceptual system of common-sense understanding".

Perhaps, one way to respond to this criticism is to argue that the innateness character of the conceptual system does not entail a priori knowledge of necessary features: the necessary features we associate to an entity depend on *how we decide* to conceptualize that entity (assuming that many conceptual systems are available). For example, the fact that we decide to categorize the Queen in the conceptual category Person entails that we attribute to the Queen specific essential properties (such as that of being a human being, an animate object, or of having specific parents, etc.). This reply is in line with Chomsky's idea that our conceptualization of the world depends on a wide range of factors, such as intentions, beliefs, assumptions, and others. Chomsky has often advocated this perspective, especially in his discussions concerning the notions of nameable object and of reference.

> There are complex conditions – poorly understood, though illustrative examples are not hard to find – that an entity must satisfy to qualify as a "naturally nameable" thing: these conditions include spatiotemporal contiguity, *Gestalt* qualities, functions within the space of human actions (Chomsky, 1975, p. 43).

> The object in front of me is not essentially a desk or a table; that very object could be any number of different things, as interests, functions, intentions of the inventors, etc. vary (Chomsky, 2000, p. 42; see also p. 127).

At the same time, however, this reply does not really solve the problem I have put forward in this paper: if the fact that we attribute to the Queen certain essential properties depends on the fact that we categorize the entity Queen as falling under the concept of Person, and if the concept Person, being part of an innate conceptual structure, is necessarily related to other concepts (being a human being, etc.), then our intuitions should *not* differ. But, as Chomsky reminds us as he discusses Kripke's example, they do. The only way to justify (from Chomsky's perspective) Chomsky and Kripke's divergent intuitions about the property of having specific parents would be to assume that one categorizes the Queen with the concept Person, while the other does not; but it is plausible to assume that they both recognize that the Queen, although a Queen, is a person. In order to resist such a conclusion (namely, that Chomsky and Kripke categorize differently the Queen), one should assume that the conceptual system is only *one of the many factors*, not *the only factor*, involved in categorization and the divergence in speakers' intuitions would then be attributed to differences in intentions, beliefs, etc.

REFERENCES

Baker, M.C. (1995). Thematic Roles and Syntactic Structures. In L. Haegeman (Ed.), *Elements of Grammar: Handbook of Generative Syntax*. Dordrecht: Kluwer, 73-173.

Casalegno, P. (1997). *Filosofia del linguaggio*. Florence: La Nuova Italia Scientifica.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheom Book.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge, MA: MIT Press.

Chomsky, N. (2007). Approaching UG from Below. In U. Sauerland, H.-M. Gärtner (Eds.), *Interfaces + Recursion = Language?*. Berlin: Mouton de Gruyter, 1-29.

Chomsky, N. (2008). On Phases. In R. Freidin, C.P. Otero & M.L. Zubizarreta (Eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Cambridge, MA: MIT Press, 133-166.

Dowty, D. (1979). *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel.

Gendler, T.S., & Hawthorne, J. (2002). *Conceivability and Possibility*. Oxford: Clarendon Press.

Graffi, G. (2001). *200 Years of Syntax. A Critical Survey*. Benjamins: Amsterdam and Philadelphia.

Grimshaw, J. (1990). *Argument Structure*. Cambridge, MA: MIT Press.

Hinzen, W. (2016). On the Grammar of Referential Dependence. *Studies in Logic, Grammar, and Rhetoric*, 46(59), 11-33.

Jackendoff, R. S. (1990). *Semantic Structures*. Cambridge, MA: MIT Press.

Kant, I. (1787). *Kritik der reinen Vernunft*. Eng. Transl. by N. Kemp Smith. London: Palgrave Macmillan, 2003.

Kripke, S.A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Larson, R., Segal, G. (1995). *Knowledge of Meaning*. Cambridge, MA: MIT Press.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Levin, B., & Malka, R.H. (1995). *Unaccusativity: At the Syntax-Semantics Interface*. Cambridge, MA: MIT Press.

Marconi, D. (1997). *Lexical Competence*. Cambridge, MA: MIT Press.

Marconi, D. (2010). Wittgenstein and Necessary Facts. In P. Frascolla, D. Marconi, & A. Voltolini (Eds.), *Wittgenstein: Mind, Meaning and Metaphilosophy*. London: Palgrave Macmillan, 140-166.

Moravcsik, J. (1975). Aitia as Generative Factors in Aristotle's Philosophy. *Dialogue*, 14, 622-636.
Moravcsik, J. (1990). *Thought and Language*. London: Routledge.
Parsons, T. (1990). *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge, MA: MIT Press.
Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.
Pinker, S. (1994). *The Language Instinct*. New York: Harper Perennial Modern Classics.
Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press;
Putnam, H. (1975a). The Analytic and the Synthetic. In H. Putnam (Ed.), *Philosophical Papers, 2, Mind, Language, and Reality*. Cambridge: Cambridge University Press, 33-69.
Putnam, H. (1975b). The Meaning of "Meaning". In K. Gunderson (Ed.), *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science*, 7. Minneapolis: University of Minnesota Press, 131-193.
Quine, W.V.O. (1951). Two Dogmas of Empiricism. *The Philosophical Review*, 60(1), 20-43.
Quine, W.V.O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
Quine, W.V.O. (1986). Replies. In L.E. Hahn & P.A. Schilpp (Eds.), *The Philosophy of W.V.O. Quine*. La Salle, Ill.: Open Court.
Ramchand, G. (2008.) *Verb meaning and the lexicon: a first-phase syntax*. Cambridge: Cambridge University Press.
Schäfer, F. (2009). The Causative Alternation. *Language and Linguistic Compass*, 3(2), 641-681.
Talmy, L. (2000). *Towards a Cognitive Semantics. Concepts Structuring Systems*. Cambridge, MA: MIT Press.

HASHEM RAMADAN
Boğaziçi University
hashem.ramadan@boun.edu.tr

THE TWO-WAY RELATIONSHIP BETWEEN LANGUAGE ACQUISITION AND SIMULATION THEORY[1]

abstract

In this paper, I try to draw a two-way connection between simulation theory and language acquisition. I argue that an individual with better simulation capabilities is at an advantage when it comes to foreign language acquisition, but this also works in the opposite direction in that exposure to many languages leads to better simulation capacities and more empathy. A number of studies relating to the subject of language and simulation will be presented in this paper. An evolutionary explanation and an analysis of the case of children with autism will also be presented to argue in favor of simulation theory over theory theory.

keywords

simulation theory, language acquisition, empathy

Phenomenology and Mind, n. 12 - 2017, pp. 132-140
DOI: 10.13128/Phe_Mi-21112
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

LANGUAGE ACQUISITION AND SIMULATION THEORY

1. Introduction

Interaction between humans is often theorized to be the result of a cognitive capacity that allows one to understand or predict the mental states of others. Such a capacity is called the *Theory of Mind*. Through attributing mental states to others, one is able to form predictions of the person's thoughts or feelings and act accordingly (Goldman, 2012). There are a few explanations as to how one comes to have a theory, one of them being *Simulation Theory*. Simulation theory basically claims that in order for humans to understand others and anticipate how they might think or act, they employ a kind of simulation where they put themselves in people's shoes. In other words, how they themselves would act or feel in certain situations would be the main criterion in determining how the person they are trying to understand would act. They would more or less project their own emotions onto others to determine how they feel or think. Goldman (2006) divided simulation capabilities within humans into two: high-level and low-level. In short, low-level simulation is the automatic response one does like mirroring someone's facial expressions without rationalizing it, while high-level simulation involves a bit of background information along with imagining of a scenario. In this paper, I will assess the question as to whether this form of cognitive activity is directly related to language learning and acquisition. I will tentatively describe the connection for both low-level and high-level simulation to language learning and acquisition. The first issue I will discuss is the possibility for people with stronger simulation capacities and those who display a higher degree of empathy towards others to have better language acquisition skills. Conversely, it is possible that exposure to different linguistic environments may increase one's ability to simulate other minds. I will argue that both ideas could be phenomena working simultaneously in that a child raised in a multilingual setting develops better communication skills and a more empathic character which in turn gives her/him an advantage to more accurately simulate other individual's intentions. This, consequently, makes her/him more skilled at foreign language acquisition. I support the hypothesis through examples of studies showing the connection between communication skills and empathy. An alternative explanation to these phenomena will be presented relying on theory-theory. Theory-theory is a contender to simulation theory. I will introduce it briefly, but argue against it by demonstrating that it falls short of accounting for the role empathy

[1] I would like to thank Professor Lucas Thorpe whose guidance throughout his social cognition seminar has helped me immensely and where a precursor to this article was submitted as a term paper.

plays in human's lives. This idea will be supplemented through analyzing empathy from an evolutionary perspective. Finally, I will mention some studies conducted with children with autism spectrum disorder (ASD) pertaining to their learning skills. The results of these studies will show that a simulation theory of mind could account for empathy's effect on language acquisition and multilingualism's effect on empathy and simulation more so than theory-theory.

2. High-Level *vs.* Low-Level Simulation

The distinction between high-level and low-level simulation needs to be fully understood first. According to simulation theory, and specifically for high-level simulation, when one wants to guess what another person would do in a certain situation, the question they would ask themselves would normally be something like "what would I do in her/his place?" (Goldman, 2012). After this initial thought, acting upon that question would entail taking some background knowledge about both the situation and the person being simulated into consideration. What this means is that a fair amount of imagination would be required in order to get to a satisfying conclusion of what the simulated person might do. The person engaging in the simulation would have to pretend mental states: s/he would assume what beliefs or desires s/he would have if s/he were the simulated person. This of course has a big margin of error due to one's own desires and way of thinking coming into play, the biggest contributor for this error being the use of one's own possible reaction to a situation as the model for simulating the opposing person's reaction. To avoid this, some inhibition needs to be employed for the simulator to get to as accurate a simulation as possible (Goldman & Jordan, 2013, p. 452). As for low-level simulation, it is more about unintentional or automatic responses to stimuli from other people. These stimuli would mainly be reactions to pain or disgust, or maybe situations where something seemingly painful or disgusting in the eyes of the simulator has happened to the person whose actions are being simulated. It is more of a primitive type of simulation where the main actors are mirror neurons. These are essentially neurons that fire after a certain action from the individual as well as after observing that action by another individual (Goldman, 2006, p. 134). The simulating side is often unaware that a process of simulation is occurring, i.e. it is possible that there would be no knowledge of the matching of mental states between the simulator and the simulated. This could entail that mirroring is not proof of effective mind reading. It might, however, be the case that it is a basis for it (Goldman, 2006, p. 134). Both high and low-level simulation seem to have correlations with how subjects are effective in communicating with one another, making them worth studying in relation to language as both a factor in their shaping and a possible result in terms of how well it is utilized.

3. Foreign Language Exposure and High-Level Simulation Skills

Starting with high-level simulation, research was done by Fan *et al.* (2015) to study the impact of exposure to multilingual environments on children's abilities to interpret a speaker's intentions. The procedure consisted in a setup where a set of objects are placed on shelves in front of the child with the instructor on the opposite side behind the shelves. The child is able to see all the objects, but some objects are hidden from the instructor by a barrier. There would be an object which the instructor would tell the child to pick and a distracting object similar to the intended one but hidden from the instructor's view. The instruction would be something along the lines of "I see such and such, can you pick it?" all while the instructor is wearing black matte sunglasses to prevent her/his gaze from affecting the child's judgement. Tests were conducted to measure the children's verbal ability, executive function, visual-spatial intelligence, and perspective taking, and it was concluded that all groups within this research had comparable levels of language understanding and could follow the instructor in the absence of a distractor. The researchers observed that children who had been exposed to a multilingual environment were able to understand which object the instructor was pointing to more often than children who were brought up in a monolingual one. It was a 50% chance for the monolingual children to get the right answer, but those who were either bilingual or had been merely exposed to more than one language had 77% and 76% success rates respectively. They concluded that either exposure to a diverse linguistic environment at any time during one's life may help in the development of more effective communicative skills, or that a child needs to be exposed at a certain stage in their development for this to take effect (Fan *et al.*, 2015). But in both cases, I believe this reveals superior higher-level simulation capacities among bilingual children or children exposed to a multilingual environment. The correct interpretation of the instructor's intentions points to this since from a simulation theory perspective, the child would need to imagine her/himself in the instructor's shoes and interpret what they would be thinking if put in that situation given certain background knowledge. In this case, the placement of the objects and being able to see or not see them acts as background knowledge. Gordon corroborates this when he describes how in a close knit community, people do not require much imagination to predict what others within their community mean when they act or speak since there is a shared set of values and norms. On the other hand, someone placed in a foreign setting would need to do a lot of pretending in order to predict other's behaviors and understand their intentions (Gordon, 1986). We tend to experience this when we move abroad by trying to adjust to our new community's norms. Our behavior changes when we move and try to adjust to our new environment. Perhaps if we go to our home country for a visit we would shift to our old behavior to accommodate that temporary change. This ease in shifting between two methods of communication, similar to shifting between two languages, could mean that the individual has developed an efficient way of understanding others' intentions and predicting their behavior. And since infants start developing the ability to interpret others' intentions and actions through repeated exposure to certain types of actions from an early age (Goldman & Jordan*,* 2013, p. 448), I believe it is reasonable to hold that children brought up in a monolingual environment would have more difficulties in discerning others' intentions than those brought up in multilingual ones. Multilingual settings provide for a much wider variety of linguistic stimuli for the infant to be exposed to and interpret behaviors through trial and error, making her/him quicker at adapting to new experiences and better able to faithfully predict or imagine what people's actions mean.

4. Empathy as a Precursor to Understanding Others

Regarding low-level simulation, I think there is a good chance that it is related to empathy in the sense that empathy requires response to minute emotional cues (Guiora *et al.,* 1968). The reactions to such cues are often quick. They don't require imagination from the receiver and come about with probably less need for background information to back them up. Perhaps those who are more empathic than others tend to react to more situations and to a more diverse range of stimuli from people of different cultures who would behave differently. It would be interesting if this has effects on acquiring foreign languages. One definition of empathy is

> a process of comprehending in which a temporary fusion of self-object boundaries, as in the earliest pattern of object relations, permits an immediate emotional apprehension of the affective experience of another, this sensing being used by the cognitive functions to gain understanding of the other (Guiora, 1965, p. 782).

On the assumption that this definition provides a correct picture of empathy, individuals who more accurately understand others' emotions and care to fruitfully interact with those they deal with the most – albeit without necessarily giving it much thought – might have better motivation to communicate with the others more deeply and seriously than most people do. Having a pronunciation that is closer to that of native speakers facilitates communication within a certain linguistic community. So being more empathic would be the driving force for striving to sound more native-like. In other words, empathy pushes individuals to want to understand others, which in turn could drive their actions to achieve this understanding. In this particular case, having native-like pronunciation is a tool to push communication forward, and consequently to ensure mutual intelligibility. This road from empathy to desiring mutual intelligibility could also have a step in the middle where another action is required in order to be able to produce the correct pronunciation. Since empathy here is formed through a low-level mechanism, mirror neurons may be at play. Consequently, "mirroring" or imitating others would be an expected outcome of people who show high levels of empathy. As a result of imitating people's accents, the imitator would be able to understand those imitated more so than individuals who do not usually imitate a foreign accent while speaking (Adank *et al.*, 2010). Furthermore, I think the repeated imitation may also be an opportunity for training oneself in speaking the language which could give the imitator an advantage over non-imitators.

5. An Alternative Account Through Theory Theory

There could be, however, an alternative to simulation theory that would explain the apparent correlation. According to theory theory, a child predicts how someone is thinking by positing theories in her/his mind about others' emotions and putting those theories to the test. This requires much more background information than simulation. The information gets built up in the child's mind as a result of trial and error, i.e. the child tests out her/his theory in real life by observing if her/his predictions about the person being theorized about hold (Goldman & Jordan, 2013, p. 450). In the case of multilingual or "exposed-to-multiple-languages" children, it could be that what gives them an advantage are the years of dealing with multiple instances where there was a chance to test out many theories, rather than simulation itself. Differences in language sometimes entail differences in thinking about certain issues due to possible intricacies or grammatical features that are present in one language but not in another. One's culture, in addition, often influences how s/he thinks about many issues: the child, upon being exposed to more than one language, will be exposed to more than one way of thinking or viewing the world. This will give her/him the opportunity to test out the same theory on both cultures. If, for example, the theory held for one and not for the other, the child will be able to understand the differences between people's methods of analyzing and thinking about the world more so than a child brought up in a monolingual society. As mentioned earlier, a closely knit and homogeneous society would require less predictions since various situations are thought about the same way by many people. This is a way to account for high-level simulation in relation to knowledge of many languages. As for theory theory, rather than putting her/himself in the person's shoes, the child tests out theories known to be true in relation to a specific group of people and not necessarily to all groups.

6. Where Theory Theory Falls Short

But since theory theory requires one to have developed a way of thinking about others after much trial and error within her/his environment, it does not explain why there could be a difference in attitudes and beliefs between people with the same background. What could account for this is a phenomenon explaining individuals' behaviors that are not the result of conscious considerations and previous knowledge. Since empathy is more often than not acted upon automatically without much regard to background information (Sonnby-Borgström, 2002), it seems to fit the description of such a phenomenon, and according to a simulation theory model, this could be explained by appealing to low-level simulation. This means that empathy would not fit with theory theory since being empathic would not always require conscious reasoning. For example, consider two individuals who have lived in the same place and encountered the same people their entire lives, yet one of them is prejudiced against a target group *G* while the other is not. The decisive factor may be that the prejudiced individual is less empathic than the other (Gutsell & Inzlicht, 2010). Without the need for a theory about other minds, and utilizing low-level simulation, one could be empathic of others and in turn be able to predict their genuine emotions and actions to a degree close enough to reality, and this is what most likely would be present in the non-prejudiced person. For this reason, I think that simulation theory provides a better account of human interaction and understanding of one another than theory theory.

7. The Evolutionary Benefit of Empathy as a Facilitator of Effective Communication

An evolutionary analysis of a mother's interaction with her child[2] may also be in favor of adopting simulation theory over theory theory. Rather than a theory theory based interaction, empathy, being involuntary and not needing much prediction and effort, would facilitate keeping the mother alert when caring for her child. We see this in human mothers where the slightest potential harm to their children automatically triggers a response to attend to them and try to shield them from danger. Perhaps mirror neurons in mothers are very similar to those of their children to facilitate this. The mother having to breastfeed would put the male in a position to have to be the main hunter in the family which would give the mother more time with her child than the father. The change from hunter-gatherer societies to agriculture and a more sedentary lifestyle wasn't made so long ago on the evolutionary calendar (Whyte, 1977), so much of these traits should normally be seen in humans today, and we would expect females to generally show higher levels of empathy than males. This is indeed what we see according to Olivares-Cuhat in a study conducted to "investigate possible relationships between empathy and foreign language learning performance on the one hand, and between emotional empathy and academic achievement, on the other hand" (Olivares-Cuhat, 2012, p. 62). Although the study was mainly to figure out if differing levels of empathy had anything to do with differing levels of achievement in education in general and foreign language learning in specific, they found that females were significantly more empathic than males (Olivares-Cuhat, 2012, p. 67). Females generally score better academically as well (Conger & Long, 2010, p. 184; Whalen *et al.*, 2003) and have higher foreign language performance (Gu, 2002, p. 35). This quote from that study summarizes what I was saying earlier:

> "a desire, willingness, or affective ability to adopt features of another cultural community and make them part of one's own behavioral repertoire ... can serve as an important influence on the individual's motivation to learn a second language" (Gardner, 2010, p. 114). From this, it follows that students endowed with positive empathic characteristics could be more able to recognize and identify with cultural differences that would, at first, promote their interest and motivation to learn a target language and, subsequently, help them become better language learners (Olivares-Cuhat, 2012, p. 69).

The method to test the participants' empathy was that developed by Caruso and Mayer (1998) in which items pertaining to certain empathy related concepts are rated by the participants.

[2] I am discussing the mother-child interaction rather than the parents-child one for two reasons: (i) the research I discuss regrettably focused on mother-child interaction only (Sullivan, 2011), and (ii) female subjects were found to be more empathic than male subjects in one study (Olivares-Cuhat, 2012). However, it is difficult to assess, from these studies alone, whether having empathic characteristics is evolutionarily convenient only for mothers or for parents in general.

For example, the participant would rate something like: "Suffering: I get very upset when I see a young child being treated meanly". The participant's score would then be calculated by taking the mean of her/his ratings. The study showcases the results as showing a correlation between academic scores and the level of empathy (Olivares-Cuhat, 2012, p. 69). This discussion of the evolutionary basis of empathy is, in my opinion, adequate to make sense of why we have developed empathy as a mechanism to understand and deal with others, and I believe that it presents a stronger case than theory theory in explaining human-human interaction and mind reading.

8. Autism Spectrum Disorder and Deficiencies in Simulation and Foreign Language Learning

The last point I would like to discuss is whether we could learn something from the case of children with ASD in regards to the topic of language and its relation to empathy. Children with ASD typically don't show a response to their caregivers' faces different than that to the faces of strangers (Powell, 2004, p. 1055). Reasoning from an evolutionary point of view, it is in the infant's interest to have a unique reaction to her/his caregivers' faces. The survival advantage of this may be that the child would be able to inform her/his caregivers when s/he is hungry (Sullivan *et al.*, 2011).[3] Perhaps the lack of a unique child-to-caregiver reaction in children with ASD is an indication of an indifference rather than an inability to tell faces apart. In addition to this, children with ASD mainly keep to themselves, and don't engage much in verbal communication, and it has been observed that foreign language learning is more difficult for children with ASD than it is for non-ASD children (Wire, 2005). This is merely speculation, but it is possible that a correlation may be found between not being socially connected to people on a scale found in non-ASD children and exhibiting weaknesses in learning a language. If, for children with ASD, the deficit to recognize minute cues and signals from others, including from the child's caregiver, leads to a lack of overall empathy on the long run, then according to what I have argued in this paper, these children should have a significantly harder time in learning a foreign language than their non-ASD peers. The longstanding notion about autism – which is one of the autism spectrum disorders – is that those who suffer from it have a lack of empathy. There has been some research, however, suggesting that the opposite may be true in that people on the autism disorder spectrum may in fact have heightened sensitivity, making it difficult to phase out the unimportant cues and process what is necessary in order to understand others (Favre *et al.*, 2015). I believe that in either case, there would likely be a deficiency in simulation capabilities. In the first case, assuming that ASD would result in a lack of empathic feelings, the person suffering from ASD could have a deficiency in her/his mirror neuron activity if we were to analyze this from a low-level simulation perspective. Such a deficiency has been observed experimentally (Dapretto *et al.*, 2005). For high-level simulation, being unable to put oneself in someone else's position would result in such indifference. On the assumption that people with ASD have heightened concentration and can pick up things non-ASD people would normally ignore, this could also lead to overwhelming the receiver with stimuli to the point that s/he would not be able to adequately analyze people around her/him both on the low and high-level. For the former, her/his mirror neurons would react to unnecessary stimuli in certain situations, and as for the latter, not understanding what's important in imagining someone's situation might lead to wrong predictions. Language is known to be generally deficient in people with ASD, which supports the hypothesis that having more accurate capacities of simulation allows

[3] Regrettably, this study did not investigate the parents-child attachment and bonding, only mother-child. There was only a brief mention of the possibility of the father providing the necessary bonding relationship and sensory stimulation for normal development, as opposed to mothers.

for better language learning. Also, a study showed that bilingual children suffering from ASD "were more likely to vocalize and utilize gestures" (Valicenti-McDermott *et al.*, 2012, p. 945). And although this study didn't show significant advantages in expressive language for bilingual children over their monolingual peers, there was more pointing and pretend-play activity from bilinguals (Valicenti-McDermott *et al.*, 2012). I think this shows that there is a high chance that there might be a connection going both ways in that individuals with better than average simulation capacities are generally capable of grasping a foreign language more easily than others, and those who have had previous experience or exposure to a foreign language are able to understand people more than those who have not.

9. Conclusion

I have attempted to show in this paper that there could be a connection between both low-level and high-level simulation on the one hand and language learning and acquisition on the other hand. On the low-level side, being able to mirror others' emotions could give an advantage to the person mirroring them in that s/he will be able to learn a foreign language and reproduce its pronunciation in a more native-like fashion. Moreover, being more empathic could make her/him more open to accept different norms and cultures which in turn provides some motivation for mutual understanding. Being more empathic encourages subjects to look for what facilitates understanding: a native-like pronunciation on the one hand, and a near native language proficiency on the other. For high-level simulation, I have shown that exposure to more than one language can provide the individual with more opportunities and experiences to refine her/his notions and ideas of others. This in turn may be the reason for better simulation capabilities since the simulator, through her/his many different experiences, will be more likely to put her/himself in others' shoes and try to understand them from different perspectives. Through reasoning and analyzing how humans may have most likely evolved, I believe that simulation theory provides a better overall picture of human social cognition than theory theory. The data concerning children with ASD support the hypothesis that having communication problems could lead to a harder time in learning languages, and that being bilingual leads to a higher possibility of communication albeit not necessarily verbal. This seems to be a good account of why simulation and language learning and acquisition may be related.

REFERENCES

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation Improves Language Comprehension. *Psychological Science*, 21(12), 1903-1909.

Caruso, D.R. & Mayer, J.D. (1998). A Measure of Emotional Empathy for Adolescents and Adults. Manuscript. https://mypages.unh.edu/sites/default/files/jdmayer/files/empathy_article_2000.pdf.

Conger, D. & Long, M.C. (2010). Why Are Men Falling Behind? Gender Gaps in College Performance and Persistence. *Annals of the American Academy of Political and Social Science*, 627(1), 184-214.

Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., & Iacoboni, M. (2005). Understanding emotions in others: minor neuron dysfunction in children with autism spectrum disorder. *Nature Neuroscience*, 9(1), 28-30.

Fan, S.P., Liberman, Z., Keysar, B., & Kinkler, K.D. (2015). The Exposure Advantage: Early Exposure to a Multilingual Environment Promotes Effective Communication. *Psychological Science*, 26(7), 1090-1097.

Favre, M.R., La Mondola, D., Meystre, J., Christodoulou, D., Cochrane, M.J., Markram, H., & Markram, K. (2015). Predictable enriched environment prevents development of hyper-emotionality in the VPA rat model of autism. *Frontiers in Neuroscience*, 9 (127).

Gardner, R. C. (2010). *Motivation and Second Language Acquisition: The Socio-Educational Model.* New York: Peter Lang.

Goldman, A.I. & Jordan, L.C. (2013). Mindreading by simulation: The roles of imagination and mirroring. In S. Baron-Cohen, H. Tager-Flusberg, & M.V. Lombardo (Eds.), *Understanding Other Minds: Perspectives from Developmental Social Neuroscience* (3rd ed). New York: Oxford University Press, 448-466.

Goldman, A.I. (2012). Theory of Mind. In E. Margolis, R. Samuels, & S.P. Stich (Eds.), *The Oxford Handbook of Philosophy of Cognitive Science.* New York: Oxford University Press, 402-424.

Goldman, A.I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mind Reading*. New York: Oxford University Press.

Gu, Y. (2002). Gender, Academic Major, and Vocabulary Learning Strategies of Chinese EFL Learners. *RELC Journal*, 33(1), 35-54.

Gordon, R.M. (1986). Folk Psychology as Simulation. *Mind and Language*, 1(2), 158-171.

Guiora, A.Z., Taylor, L.L., & Brandwin, M.A. (1968). The Role of Empathy in Second Language Behavior. *Center for Research on Language and Language Behavior*. http://files.eric.ed.gov/fulltext/ED024952.pdf.

Guiora, A.Z. (1965). On clinical diagnosis and prediction. *Psychological Reports*, 17, 779-784.

Gutsell, J.N., & Inzlicht, M. (2010). Empathy constrained: Prejudice predicts reduced mental simulation of actions during observation of outgroups. *Journal of Experimental Social Psychology*, 46(5), 841-845.

Olivares-Cuhat, G. (2012). Does Empathy make a difference in the Foreign Language Classroom?. *Alicanto*, 5, 62-72.

Powell, K. (2004). Opening a Window to the Autistic Brain. *PLoS Biology* 2(8): e267, 1054-1058.

Sonnby-Borgström, M. (2002). Automatic Mimicry Reactions as Related to Differences in Emotional Empathy. *Scandinavian Journal of Psychology,* 43(5), 433-443.

Sullivan, R., Perry, B.S., Sloan, A., Kleinhaus, K., & Burtchen, N. (2011). Infant Bonding and Attachment to the Caregiver: Insights from Basic and Clinical Science. *Clinics in Perinatology* 38(4), 645-655.

Valicenti-McDermott, M., Tarshis, N., Schouls, M., Galdston, M., Hottinger, K., Seijo, R., Shulman, L., & Shinnar, S. (2012). Language Differences Between Monolingual English and Bilingual English-Spanish Young Children With Autism Spectrum Disorders. *Journal of Child Neurology* 28(7), 945-948.

Whyte, R.O. (1977). The Botanical Neolithic Revolution. *Human Ecology* 5(3), 209-222.

Wire, V. (2005). Autistic Spectrum Disorders and learning foreign languages. *Support for Learning*, 20(3), 123-128.

MARCO FENICI
University of Florence
mfenici@hotmail.com

REBUILDING THE LANDSCAPE OF PSYCHOLOGICAL UNDERSTANDING AFTER THE MINDREADING WAR[1]

abstract

'Mindreading war' refers here to the intricate net of connected debates both in the philosophy and the cognitive sciences concerning the onset, the development, and the nature of the cognitive mechanisms underlying mindreading – i.e., the alleged ability to attribute mental states to predict and explain others' behavior. The mindreading war has lasted for almost forty years by now with apparently no winners or losers. This article argues that the present stalemate results from the lack of initial theoretical discussion about foundational issues that led to the conflict. Recovering the dialogue between psychologists and philosophers is necessary if we are to start rebuilding the landscape of psychological understanding once this long war is over.

keywords

folk psychology, mindreading, social cognition, theory of mind

Phenomenology and Mind, n. 12 - 2017, pp. 142-150
DOI: 10.13128/Phe_Mi-21113
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Mindreading, or Theory of Mind (ToM), is usually defined as the ability to attribute mental states (e.g., beliefs, desires, and intentions) to others to predict and explain their behavior.

By 'mindreading war', I refer here to the intricate network of connected debates both in the philosophy and the cognitive sciences concerning the onset and the development of this capacity in infancy and early childhood as well as the description of the cognitive mechanisms underlying it.

The mindreading war has lasted for almost forty years by now with apparently no winners or losers. In psychology, the confrontation between the proponents of nativist and constructivist accounts persists to the present day, while dual-system theorists of social cognition have also joined the discussion more recently. In philosophy, the debate has dissolved in a plethora of proposals addressing a range of different issues with unclear consequences on the empirical debate (see Apperly, 2008, 2009; Stich & Nichols, 1997).

This article surveys the outcomes of the conflict but first clarifies its onset and development. Accordingly, it describes the start of the war in section 2, the empirical debate in psychology in section 3, and the first and second theoretical confrontation in philosophy in sections 4 and 5, respectively. Finally, section 6 tries to clarify the relevance of each of these phases to the wider context of the mindreading war. A clear understanding of the foundational issues that led to the mindreading war is necessary if we are to start rebuilding the landscape of psychological understanding once this long war is over.

2. The *Casus Belli* and the Start of the Mindreading War

In 1978, two comparative psychologists devised a procedure to assess whether a subject would understand the intention behind an observed action (Premack & Woodruff, 1978). They found that a chimpanzee was extremely reliable in passing the task, and concluded that chimpanzees have a *Theory of Mind* (ToM) – i.e., a theoretical capacity to infer the mental states of other agents, and to exploit these attributions to predict others' actions.

Philosophers, who at that time were debating about the meaning and the scientific relevance of mental state terms, acknowledged the validity of the procedure to assess a subject's understanding of mental states but argued that the proposed experiment did not test the possession of the concept most distinctive of human psychological understanding, that is, belief (Bennett, 1978; Dennett, 1978; Harman, 1978).

[1] I wish to thank Jay Garfield for useful comments on a previous version of this article.

Developmental psychologists rose to the challenge, and responded a few years later by devising a new procedure, the false belief test (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983), which requires subjects to predict an actor's behavior that crucially depends on her possession of a false belief. Strikingly, they found that children below age four could not pass the task, which suggested to them that younger children lacked a proper concept of belief as well as full-fledged ToM (Wellman, Cross, & Watson, 2001).

3. The First Phase of the Conflict: The Theory-of-Mind Debate in Psychology

The earliest results from the false belief test started a debate among developmental psychologists, which has developed exponentially. On the one hand, advocates of modularism interpreted the results as demonstrating that children are endowed with a ToM module (ToMM) – i.e., a cognitive mechanism specific for social cognition – that has been shaped through natural selection, and is underpinned by dedicated neural processes. Some of them initially proposed that ToMM reaches full maturation right when children pass the false belief test (Baron-Cohen, 1995). Over time, however, modularists have converged on the claim that ToMM matures in early infancy (Baillargeon, Scott, & He, 2010; Leslie, 2005), and that children younger than four struggle with the false belief test because of limited computational resources (Bloom & German, 2000).

On the other hand, advocates of constructivism claimed that children progressively construct a ToM through a process of hypothesis testing and revision that is akin to the construction of a scientific theory (Gopnik & Meltzoff, 1996; Gopnik & Wellman, 1994; Perner, 1991; Wellman, 1990; see also Perner, Huemer, & Leahy, 2015 for a very recent proposal). Accordingly, they argued that younger children fail the false belief test because they lack a full-fledged concept of belief, which is acquired only after age four.

Such a vibrant confrontation has become even more intense after the findings that even infants manifest a sensitivity to others' (false) beliefs when assessed by measuring different behavioral indices (e.g., anticipatory and preferential gazing, fixation time) rather than by direct questioning (see Baillargeon, Scott, & Bian, 2016 for a review). According to modularists, these results from "spontaneous-response" – as opposed to traditional "elicited-response" – false belief tasks demonstrate the very early onset of the ToM mechanism while constructivists have replied by explaining the same data in the terms of basic capacities to form an expectation about observed actions requiring no attribution of mental states (Perner & Ruffman, 2005; Rakoczy, 2012; Ruffman, 2014; see also Banovsky, 2016; Fenici, 2014; Fenici & Zawidzki, 2016).

The debate is still far from being settled. It has been given new impetus by the appearance of dual-system theorists on the battle field (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). In agreement with modularists, they contend that infants are endowed with socio-cognitive abilities that are specific to the processing of others' mental states – although they initially track what another agent has seen – i.e., her perceptual states – rather than what she believes. In agreement with constructivists, however, they claim that these basic abilities are significantly limited, and that the mature possession of ToM depends on the later emergence of distinct forms of sophisticated mindreading after age four.

4. The Second Phase of the Conflict: The Mindreading Debate in Philosophy

While psychologists argued about the onset and the development of ToM through the ontogeny, some philosophers noticed that both modularists and constructivists had each assumed that the capacity to attribute mental states relies on either tacit (modularism) or explicit (constructivism) knowledge of the principles of folk psychology, and thereby had endorsed a theoretical account about ToM – what they called the *Theory Theory* (TT) of the mind. In contrast, these partisans of the *Simulation Theory* (ST) argued that the attribution of mental states – or *mindreading*, as they re-labelled it – has a practical nature, and depends on the capacity to simulate another agent by putting oneself "in her shoes".

Some of these scholars focused on the personal-level features of the simulatory mechanism (Heal, 1986), others on the subpersonal mechanism itself (Gallese & Goldman, 1998; Goldman, 2006; Gordon, 1986). Simulation theorists also disagreed about whether simulation recruits the simulator's introspection capacities. In general, however, they all agreed that attributing mental states to an agent to predict her behavior proceeds in three steps: (i) the simulator projects herself in the place of the agent, and simulates the mental states she would have if she were there; (ii) the simulator practically decides what she would do if she had the mental states she pretends to have; (iii) the simulator interrupts the simulation, and projects on the agent the intention to act according to what she planned to do during the process.

A novel conflict then arose (see Carruthers & Smith, 1996; Davies & Stone, 1995 for a review), with some philosophers taking up the arms against their colleagues, and in defence of some of the traditional "theoretical" solutions to the ToM debate in psychology (Botterill, 1996). Progressively, the two sides acknowledged that there are compelling cases favoring each of the contenders, and that people may attribute mental states through both theorizing and simulation depending on the context – although each argued that their proposed strategy is the default (e.g., Carruthers, 1996; Goldman, 2006). In this context, proponents of "hybrid" accounts also introduced combined cognitive architectures implementing mindreading through a mix of theorizing and simulation (Currie & Ravenscroft, 2004; Meltzoff, 2002, 2005; Mitchell, Currie, & Ziegler, 2009; Nichols & Stich, 2003). The contenders seemed to having agreed to a truce but the spark for a novel conflict was smouldering under the ashes, and peace was far from definitive.

5. The Third Phase of the Conflict: The Debate About Social Cognition in Philosophy

The memory of the clash between the supporters of TT and ST was still alive when a novel movement of opposition rose among philosophers. Several dissenters rebelled against the traditional parties arguing that both TT and ST misconstructed the role of mindreading in social interaction. These *Interaction Theorists* (ITs), as we may call them, did not regroup around a central claim but all gave a prominent role to social interaction (possibly with no attribution of mental states) in the explanation of traditional cases of mindreading.

Some supporters of IT criticized the idea – shared by TT and ST – that mental states can be only inferred indirectly from the observation of behavior (Gallagher, 2008, 2011; Krueger & Overgaard, 2012; Overgaard, 2015; Zahavi, 2008, 2011). Inspired by phenomenology (Scheler, 1954, p. 260; Husserl, 1982, p. 51), or by Wittgenstein (see Overgaard & Zahavi, 2009 for an analysis), they claimed instead that mental states can be the object of direct perception so that mindreading requires neither theorizing nor simulation.

Other ITs argued that action understanding often depends on the practical knowledge of the social context in which it occurs. Accordingly, many cases of everyday social interaction do not really require mindreading (Gallagher, 2001, 2012; Gallagher & Hutto, 2008; Hutto, 2017). Some also endorsed the complementary narrative practice hypothesis (NPH), according to which genuine mindreading (whenever it occurs) involves a narrative capacity to evaluate and discuss one's reasons for acting but significantly does not require either simulation or theorizing (Hutto, 2004, 2008).

Finally, some ITs affiliated to enactivism argued that social cognition – as every case of cognitive activity – is essentially constituted by the dynamic processes of interaction between the body and the environment (both physical and social) (Fuchs & De Jaegher, 2009; De Jaegher, 2009; McGann & De Jaegher, 2009). Consequently, they criticized both TT and ST for defining mindreading as the spectatorial capacity of an observer that looks at other agents from a detached ("third-person") perspective. In contrast, they claimed, our engagement with others is mostly embedded in a social ("second-person") interactional context.

6. Rebuilding the Landscape of Psychological Understanding

After decades of trench warfare with increasingly sophisticated theoretical weaponry deployed, it seems that no winners or losers have emerged. In psychology, modularists, constructivists, and dual-system theorists still battle to gain supremacy in the field. Their debate is far from being settled, however, with each party bringing new evidence to bear favoring one view or another. Importantly, little attention is given to how much the dispute is actually empirical and how much it depends on the theoretical framework that psychologists implicitly assume to account for human development (Overton, 2015).

In philosophy, the confrontation between TT and ST have ended without apparent resolution while IT have also occupied an area in the debate. These factions seem to have divided the field in different zones of influence, each patrolling her own area without challenging their rivals. Importantly, little of the philosophical discussion between TT, ST, and IT seems to influence the empirical debate in psychology.

I argue that the present stalemate results from the lack of initial theoretical discussion (both in psychology and philosophy) about foundational issues at the origin of the debate. The earliest studies employing the false belief test paradigm implicitly assumed that there must be a ToM competence that is possible to test. By acknowledging the validity of this procedures in assessing one's understanding of belief, and moving to discuss its most proper description, most of philosophers in the initial TT-ST debate corroborated the idea that we mostly make sense of others' actions by tacitly attributing mental states. Accordingly, ToM or mindreading quickly turned from an operational construct related to an experimental procedure into a reified cognitive capacity responsible for most of our social interaction. As such, it was then immediately available as an object for (diverging) psychological and philosophical inquiries.

The more recent discussion in philosophy has however shown that there are important reasons to question the initial setting of the ToM debate. If, as ITs have argued, many cases of everyday social interaction may not require at all the attribution of mental states, we should reconsider the *explanandum* – not only the *explanans* – of (either spontaneous- or elicited-response) false belief tests. Because mindreading may be not so pervasive in everyday social interaction, developmental psychologists may have got it wrong in assuming that infants and children must attribute mental states in order to succeed in these tasks (see, e.g., Andrews, 2012, pp. 22-34; de Bruin & Newen, 2012, p. 254).

This consideration also suggests that we reconsider the plausibility of mindreading (or ToM) as a unified cognitive capacity. If, as ITs have also argued, action prediction depends on a variety of cognitive mechanisms, mindreading (or ToM) may not exist as a unified cognitive capacity (Fenici, 2012, 2017; Garfield, Peterson, & Perry, 2001; Hutto, 2008). In contrast, we may attribute mental states to others more as a way to explain and rationalize their (past) actions rather than to predict their (future) ones (McGeer, 2007).

If this analysis is at least partially correct, psychologists and philosophers would greatly benefit from an armistice conference – a collective reassessment of the foundational issues that led to the mindreading war. I conjecture that recovering the dialogue between the two disciplines will help restructuring the present landscape of psychological understanding, and enter a new era of peaceful successful collaboration between them.

I conclude offering reflection about four different issues that, I believe, require additional theoretical clarification, and should attract the new generations of philosophers for the more general benefit of the scientific community.

1) *The role of philosophical theories of content within the empirical debate*. As I noted in section 3, psychologists are still debating whether the infant's selective response to an agent's (false) beliefs in spontaneous-response false belief tasks really demonstrates that, in their second year, infants understand the concept of belief – with proponents of modularist accounts of ToM claiming that it does, and advocates of constructivist accounts claiming that it does not (see Fenici, 2015 for discussion). Reaching two opposite conclusions, Buckner (2014) and Hutto (2017) have argued that the issue is partially decided by what theory of mental content one assumes. Philosophers should make explicit how some differences among the opposed accounts in the ToM debate among psychologists might actually depend on their different philosophical assumptions about representational concepts.
2) *The traditional debate between TT vs. ST*. It is surprising that neither TT nor ST have taken advantage of the vast amount of new evidence brought forward by psychologists to decide the empirical ToM debate. To some, this might indicate that, despite its appearance, the TT–ST debate is underdetermined by the experimental data because it aims to provide a high-level, functional description of mindreading capacities that can be accommodated with any empirical evidence (Heal, 1994; Stich & Nichols, 1997). To others, it might indicate that even more data are required before settling the issue. Whatever is the explanation, philosophers should clarify why the traditional debate about mindreading is so silent when its empirical counterpart looks so flourishing.
3) *The silent invasion of IT*. The clash of IT on the traditional TT–ST debate lacked significant consequences. In some cases, IT has been criticized to be generic or implausible (e.g., Bohl, 2015; Spaulding, 2015). In many other cases, it has advanced its claims without outcry. In general, IT is by now perceived as another proposal in the field aside TT and ST. Still, IT purports to be a radical alternative to both of them. Philosophers should indicate what among the three is the most plausible candidate, or whether some forms of co-existence are also possible.
4) *The significance of the pluralist solution*. The latest discussion in the philosophical debate saw the appearance of a wide notion of pluralism about folk psychology. On this view, being a folk psychologist depends on having the capacity to deploy a variety of strategies to make sense of others' actions that include (but are not limited to) the abilities to attribute mental states, behavioral dispositions and traits, social roles and stereotypes as well as the capacity to embed observed actions within social norms and scripts (Andrews, 2012; Fiebich & Coltheart, 2015; Maibom, 2007). It remains unclear whether these are all specific forms of psychological understanding or whether they are part of some more general understanding of human action. Philosophers should clarify the boundaries (if any) of our psychological comprehension, and whether they constitute a domain different from the general comprehension of our own form of life.

REFERENCES

Andrews, K. (2012). *Do Apes Read Minds?: Toward a New Folk Psychology*. Cambridge, MA: MIT Press.

Apperly, I.A. (2008). Beyond Simulation-theory and Theory-theory: why social cognitive neuroscience should use its own concepts to study "theory of mind". *Cognition*, 107(1), 266-283.

Apperly, I.A. (2009). Alternative routes to perspective-taking: Imagination and rule-use may be better than simulation and theorising. *British Journal of Developmental Psychology*, 27(3), 545-553.

Apperly, I.A., & Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological Review*, 116(4), 953-970.

Baillargeon, R., Scott, R.M., & Bian, L. (2016). Psychological reasoning in infancy. Annual Review of Psychology, 67, 159-186.

Baillargeon, R., Scott, R.M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110-118.

Banovsky, J. (2016). Theories, structures and simulations in the research of early mentalizing. *Cognitive Systems Research*, 40, 129-143.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "Theory of Mind"?. *Cognition*, 21(1), 37-46.

Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, 1(4), 557-560.

Bloom, P., & German, T.P. (2000). Two reasons to abandon the false belief task as a test of Theory of Mind. *Cognition*, 77(1), 25-31.

Bohl, V. (2015). Continuing debates on direct social perception: Some notes on Gallagher's analysis of "the new hybrids". *Consciousness and Cognition*, 36, 466-471.

Botterill, G. (1996). Folk psychology and theoretical status. In P. Carruthers & P.K. Smith (Eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 184-199.

Buckner, C. (2014). The Semantic Problem(s) with Research on Animal Mind-Reading. *Mind & Language*, 29(5), 566-589.

Butterfill, S.A., & Apperly, I.A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28, 606-637.

Carruthers, P. (1996). Simulation and self-knowledge: a defence of theory-theory. In P. Carruthers & P.K. Smith (Eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 22-38.

Carruthers, P., & Smith, P.K. (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Currie, G., & Ravenscroft, I. (2004). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford-New York: Clarendon Press.

Davies, M.K., & Stone, T. (Eds.) (1995). *Folk Psychology: The Theory of Mind* Debate (1st ed.). Oxford: Blackwell.

de Bruin, L., & Newen, A. (2012). An association account of false belief understanding. *Cognition*, 123(2), 240-259.

De Jaegher, H. (2009). Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition*, 18(2), 535-542; discussion 543-550.

Dennett, D.C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1(4), 568-570.

Fenici, M. (2012). Embodied social cognition and embedded theory of mind. *Biolinguistics*, 6(3-4), 276-307.

Fenici, M. (2014). A simple explanation of apparent early mindreading: infants' sensitivity to goals and gaze direction. *Phenomenology and the Cognitive Sciences*, 14(3), 497-515.

Fenici, M. (2015). Social cognitive abilities in infancy: Is mindreading the best explanation?. *Philosophical Psychology*, 28(3), 387-411.

Fenici, M. (2017). What is the role of experience in children's success in the false belief test: Maturation, facilitation, attunement or induction?. *Mind & Language*, 32(3), 308-337.
Fenici, M., & Zawidzki, T. W. (2016). Action understanding in infancy: Do infant interpreters attribute enduring mental states or track relational properties of transient bouts of behavior?. *Studia Philosophica Estonica*, 9(2), 237-257.
Fiebich, A., & Coltheart, M. (2015). Various Ways to Understand Other Minds: Towards a Pluralistic Approach to the Explanation of Social Understanding. *Mind & Language*, 30(3), 235-258.
Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, 8(4), 465-486.
Gallagher, S. (2001). The practice of mind: Theory, simulation, or interaction?. *Journal of Consciousness Studies*, 5(7), 83-108.
Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17(2), 535-543.
Gallagher, S. (2011). Strong interaction and self-agency. *Humana.Mente*, 15, 55-76.
Gallagher, S. (2012). Neurons, neonates and narrative: From embodied resonance toempathic understanding. In A. Foolen, U. Lüdtke, J. Zlatev, & T.P. Racine (Eds.), *Moving Ourselves, Moving Others*. Amsterdam: John Benjamins, 167-196.
Gallagher, S., & Hutto, D.D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T.P. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity*. Amsterdam: John Benjamins, 17-38.
Gallese, V., & Goldman, A.I. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493-501.
Garfield, J.L., Peterson, C.C., & Perry, T. (2001). Social cognition, language acquisition and the development of the Theory of Mind. *Mind & Language*, 16(5), 494-541.
Goldman, A.I. (2006). *Simulating Minds: the Philosophy, Psychology, and Neuroscience of Mindreading*. New York: Oxford University Press.
Gopnik, A., & Meltzoff, A.N. (1996). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.
Gopnik, A., & Wellman, H.M. (1994). The Theory Theory. In L.A. Hirschfeld & S.A. Gelman, *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press, 257-293.
Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, 1(2), 158-171.
Harman, G. (1978). Studying the chimpanzee's Theory of Mind. *Behavioral and Brain Sciences*, 1(4), 576-577.
Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, Mind and Logic*. Cambridge: Cambridge University Press, 135-150.
Heal, J. (1994). Simulation vs. theory-theory: what is at issue? In C. Peacocke (Ed.), *Objectivity, Simulation, and the Unity of Consciousness*. Oxford: Oxford University Press, 129-144.
Husserl, E. (1982). *Ideas Pertaining to a Pure Phenomenology and to a a Phenomenological Philosophy. First Book: General Introduction to a Pure Phenomenology* (Engl. Transl. by F. Kersten). Dordrecht: Kluwer Academic Publishers.
Hutto, D.D. (2004). The limits of spectatorial folk psychology. *Mind & Language*, 19(5), 548-573.
Hutto, D.D. (2008). *Folk Psychological Narratives*. Cambridge, MA: MIT Press.
Hutto, D.D. (2017). Basic social cognition without mindreading: minding minds without attributing contents. *Synthese*, 194(3), 827-846.
Krueger, U. & Overgaard, S. (2012). Seeing subjectivity: Defending a perceptual account of other minds. In S. Miguens & G. Preyer (Eds.), *Consciousness and Subjectivity*, 47. Heusenstamm: Ontos Verlag, 239-262.
Leslie, A.M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, 9(10), 459-462.

Maibom, H. (2007). Social systems. *Philosophical Psychology*, 20(5), 557-578.

McGann, M., & De Jaegher, H. (2009). Self-other contingencies: Enacting social perception. *Phenomenology and the Cognitive Sciences*, 8(4), 417-437.

McGeer, V. (2007). The regulative dimension of folk-psychology. In D.D. Hutto & M. Ratcliffe (Eds.), *Folk-psychology Reassessed*. Dordrecht: Springer, 137-156.

Meltzoff, A.N. (2002). Imitation as a mechanism of social cognition: Origins of empathy, theory of mind, and the representation of action. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development*. Oxford: Blackwell, 6-25.

Meltzoff, A.N. (2005). Imitation and Other Minds: The "Like Me" Hypothesis. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation: From Neuroscience to Social Science*. Cambridge, MA: MIT Press, 55-77.

Mitchell, P., Currie, G., & Ziegler, F. (2009). *Two routes to perspective: Simulation and rule-use as approaches to mentalizing. British Journal of Developmental Psychology*, 27(3), 513-543.

Nichols, S., & Stich, S. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. New York: Oxford University Press.

Overgaard, S. (2015). The unobservability thesis. *Synthese*, 1-18. DOI: 10.1007/s11229-015-0804-3.

Overgaard, S., & Zahavi, D. (2009). Understanding (Other) Minds: Wittgenstein's Phenomenological Contribution. In E. Zamuner & D. Levy (Eds.), *Wittgenstein's Enduring Arguments*. London: Routledge, 60-86.

Overton, W.F. (2015). Processes, Relations, and Relational-Developmental-Systems. In R.M. Lerner (Ed.), *Handbook of Child Psychology and Developmental Science*, 1. Hoboken, NJ: John Wiley & Sons, 9-62.

Perner, J. (1991). *Understanding the Representational Mind. Cambridge*, MA: The MIT Press.

Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition*, 145, 77-88.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: how deep?. *Science*, 308(5719), 214–216.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a Theory of Mind?. *Behavioral and Brain Sciences*, 1(4), 515-526.

Rakoczy, H. (2012). Do infants have a theory of mind?. *British Journal of Developmental Psychology*, 30(1), 59-74.

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, 34(3), 265-293.

Scheler, M. (1954). *The nature of sympathy*. London: Routledge & Kegan Paul.

Spaulding, S. (2015). On Direct Social Perception. *Consciousness and Cognition*, 36, 472-482.

Stich, S., & Nichols, S. (1997). Cognitive Penetrability, Rationality and Restricted Simulation. *Mind & Language*, 12(3-4), 297-326.

Wellman, H.M. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72(3), 655-684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128.

Zahavi, D. (2008). Simulation, projection and empathy. *Consciousness and Cognition*, 17(2), 514-522.

Zahavi, D. (2011). Empathy and Direct Social Perception: A Phenomenological Proposal. *Review of Philosophy and Psychology*, 2(3), 541-558.

ALESSANDRA BUCCELLA
University of Pittsburgh
alb319@pitt.edu

NATURALIZING QUALIA

abstract

Hill (2014) argues that perceptual qualia, i.e. the ways in which things look from a viewpoint, are physical properties of objects. They are relational in nature, that is, they are functions of objects' intrinsic properties, viewpoints, and observers. Hill also claims that his kind of representationalism is the only view capable of "naturalizing qualia". After discussing a worry with Hill's account, I put forward an alternative, which is just as "naturalization-friendly". I build upon Chirimuuta's color adverbialism (2015), and I argue that we would better serve the "naturalizing project" if we abandoned representationalism and preferred a broadly adverbialist view of perceptual qualia.

keywords

qualia, naturalism, adverbialism, perceptual experience, appearance, reality

Phenomenology and Mind, n. 12 - 2017, pp. 152-161
DOI: 10.13128/Phe_Mi-21114
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Perceptual experiences are the filter through which the world reveals itself to us. All we *know* about the world, all we can *do* in the world, how it *feels* to be in the world: it's all informed by what experiences "tell us".

Philosophical theories aim at giving an account of what perceptual experiences are, and what their relationship with the world out there is. In particular, one theory seems to provide a quite elegant explanation of the relationship between our experiences of the world and the world itself: representationalism. The main idea is simple: perceptual experiences have *contents* which *represent* objects as having certain properties, which may or may not be the properties objects really have. Representationalists (e.g. Dretske, 1988; Fodor, 1987) think that, in order to *explain behavior,* we must understand perception as "organizing" the world for us by representing it as being in a certain way. The idea is the following. It is intuitively plausible to say that we act on the basis of how we experience the world to be, and that sometimes we fail at what we do precisely because we take the world to be in a way it is not. Representationalists then argue that this fact can only be explained if we postulate the existence of something that "mediates" between us and the environment as it really is, namely the representational content of perceptual experiences.

Representationalism has been found particularly attractive not only by philosophers who are interested in explaining perception-guided behavior, but also by those committed to an attempt to "naturalize" perceptual experience itself and conscious experience more generally.[1] Although a proper definition of 'naturalism' in philosophy of mind is hard to provide,[2] I take both Christopher Hill (my main target in this paper) and myself to share the following definition. To naturalize experience is (1) to explain the fact that we are consciously "in touch" with the world by means of our perceptual (and cognitive) capacities in a way that is maximally coherent with neurosciences, psychology, and other empirical sciences of the mind, and (2) to avoid commitments to entities which don't clearly belong to the physical world. For many naturalists who accept the rough definition above, (accurately/veridically) representing the world is simply the evolved function of the perceptual system, and therefore

[1] E.g. Dretske (1995), Tye (1995, 2000), Dennett (1990, 1991), Burge (2010), Hill (2014, chs. 11-12).

[2] Starting from the "early" naturalists such as John Dewey, or Roy Wood Sellars, authors who call themselves naturalists disagree with respect to how strong the commitment to coherence with the science should be. Moreover, the distinction between methodological and metaphysical naturalists is often overlooked and not sufficiently spelled out by naturalists themselves.

representationalism is the most straightforward route to the naturalization of experience.[3] On the other hand, philosophers who are less committed to the so-called "naturalizing project" have one main worry. Nagel (1974) puts the point in the following way. Usually, explaining something in naturalistic terms means explaining it in a way that is public, shareable, and accessible by many people. This entails describing the phenomenon we want to explain in increasingly objective terms, so that the description can "go beyond" our idiosyncratic experience of the phenomenon. However, it is hard for a naturalist to capture *every* component of conscious experience – and perceptual experience in particular – in such objective terms. According to Nagel and those who followed his lead,[4] conscious experience has an intrinsically subjective component.

Traditionally, such an intrinsically subjective aspect of conscious experiences (not limited to the perceptual case) is called *phenomenal character*, and it is captured intuitively by the idea of "what-it-is-like" to have a certain experience. Take the example of having a dull pain in your ankle.[5] You want to describe the experience to someone else, perhaps a neuroscientist who studies pain experiences. The scientist may have put electrodes all around your skull, and may be measuring blood pressure, heartbeat, and other stress indicators in order to understand how your body is reacting to pain. Additionally, you may try really hard and describe how it feels like to experience pain by using terms and phrases in a shared language as precisely as you can. Together, all these elements will help the scientist reach a naturalized explanation of pain experiences, that is, an explanation that uses objective and public descriptions in order to explain what pain is. However, Nagel would point out that neither words and phrases in a shared language, nor physiological indicators, seem to be the kind of things that can "disclose" *what this particular pain-experience is like for you*, from a first-person perspective.[6] The notion of a *quale* (plural: *qualia*) can be helpful in order to systematize this idea of "what-it-is-likeness", and characterize more precisely the *explanandum* for which naturalists should find a place in their theories of conscious experience.

The subjective component of perceptual experiences is constituted by *perceptual qualia*. Perceptual qualia are defined by Kim (2006) as "the ways that things look, seem, and appear to conscious observers" (p. 225). The particular brown-ness of the table I am sitting at, the particular softness of my cat's fur when I touch it, the particular sharpness of her meowing, etc. These all count as perceptual qualia: they make up the intrinsically qualitative aspect of *my* perceptual experiences.

Hill (2014, chs. 11-12) takes on board the challenge of giving a plausible account of perceptual qualia in a naturalistic framework, or, in his own words, "to bring qualia into the physical fold" (p. 197).[7] In fact, Hill thinks that representationalism is the *only* theory that can succeed in the hard enterprise of naturalizing qualia,[8] and he tries to articulate how exactly such a

[3] See, for example, Burge (2010).

[4] Apart from Nagel (1974), people who have highlighted the presence of an intrinsically subjective component of conscious experience are, among others, Jackson (1982, 1986) and Chalmers (1996, 2007).

[5] An analogous example was presented by Robert Howton in a lecture. I thank him for putting things this way, which I found very helpful.

[6] For a more thorough explanation of the difference between first-person and third-person descriptions of conscious experience, see Chalmers (2001).

[7] Other representationalists who chose the "naturalizing route", though in a less explicit way, are Shoemaker (1975, 1981) and Kriegel (2002).

[8] "As far as I can tell, there is only one theory of experience that provides a satisfactory way of dealing with this problem [i.e. the problem of placing qualia in the physical world] – representationalism" (2014, p. 201). Hill quickly dismisses naïve realist-type theories in the light of the "mysteriousness" of the notion of direct acquaintance, especially in the context of an attempt to explain qualia and justify something like a genuine appearance-reality

naturalization of qualia could go.

In this paper, I shall motivate my disagreement with Hill. While I grant that Hill does succeed in giving a naturalistic account of qualia, my disagreement regards the claim that his view provides the *only* good naturalistic account of qualia. In particular, I will argue that a kind of *adverbialist* account can do equally well. By questioning the very categories employed in the traditional debate about qualia, the kind of adverbialist[9] position I articulate opens a different, promising path for the naturalizing project while avoiding a problematic commitment of Hill's view. I will first sketch the main points of Hill's view, by at the same time emphasizing what I take to be its most controversial commitment. Secondly, I will sketch an adverbialist solution which avoids Hill's counter-intuitive commitment regarding the nature of qualia while at the same time being "naturalization-friendly".

2. Perceptual Qualia: What Are They?

Earlier on I have given the definition of qualia endorsed by Hill. He understands the notion of *quale* as the notion of *look,* or *appearance*, in a *phenomenal* sense. This notion is spelled out by Hill by collapsing it into the notion of "comparative" look:[10]

> When one says that an object looks small to an observer, using "looks small" in this phenomenological sense, one is not claiming that the observer's perceptual experience supports the judgment that the object really is small. One is not saying that the observer's experience represents the object as small. Rather one is drawing an analogy between the observer's current visual experience and the visual experiences he has when is viewing objects that are reasonably close at hand and really are small (2014, p. 198).

Take the *particular smallness* that we experience while observing a truck on the street very far away from us. The truck, even though it is big, *phenomenally looks small* (or looks$_P$ small) to me. There is a sense in which this phenomenal smallness is the same phenomenal smallness that other objects look as having. For instance, a miniature toy truck looks$_P$ small as well, with the only difference that the latter *is* small, whereas the former *is* big. The objective sizes of the real truck and of the toy truck are very different; however, when the real truck is far away and the toy truck is in our hands, they both look small to us *in the same way.*

If this is plausible enough, however, naturalistic accounts of perceptual experience face what we may call the "problem of qualia": how to explain in a naturalistic framework the fact that the real truck and the toy truck, despite having completely different objective sizes, *look* of the same size *to me* in particular circumstances? *Where* is the smallness the two trucks seem to share? Is it in our head? Is it in the world? Is it nowhere at all? Hill has a proposal.

distinction in a naturalistic framework. He also dismisses the group of so-called doxastic theories, which hold that awareness of qualia depends on the subject having some sort of doxastic attitude (e.g. a belief or a judgment) towards the perceived scene. Against doxastic views, Hill appeals to phenomenology, and claims that doxastic views don't accommodate all the properties of experiential awareness (2014, pp. 207-209).

[9] In fact, Hill provides an argument for rejecting adverbialism as a viable alternative in the "naturalizing qualia" project (2014, p. 208). I will briefly address the objection after the form of adverbialism I favor has been introduced.

[10] The notion of phenomenal/comparative look is meant to be kept distinct from the epistemic notion of look, where an observer is actually disposed to judge that an object is small and uses the "looks"-claim as evidence for that judgment.

3. Hill on Qualia and The Appearance/ Reality Distinction

Hill considers qualia as the ways things *phenomenally* look to observers. In particular, since he wants to “locate” qualia in the physical world, he suggests that qualia are a special kind of property: they are phenomenal properties objects have in virtue of how they appear to a certain observer, at a certain time, from a certain angle. The small quale that I experience when I look at (a) the real truck from a distance and (b) the toy truck in my hand are nothing but two distinct physical properties, each belonging to one of the two objects. Both are properties of looking small to an observer in circumstances C(a) and C(b). These are *physical* properties because they can be seen as functions with multiple variables, all of which can be assigned precise values empirically. Hill (2014) identifies phenomenal properties with what he calls *Thouless properties:*[11]

> What then is the general character of a Thouless property? An example is the property P that an external object possesses just in case it is producing a retinal image with a height H such that the result of applying f1, . . . , fn to H and other relevant quantities has the numerical value N. P is a Thouless size. Alternatively, a Thouless size can be seen as a property P* that an external object possesses just in case it is subtending a visual angle V such that the result of applying the computable functions f1*, . . . , fn* to V and other relevant quantities has the numerical value N*. P and P* are both legitimate physical properties of external objects, though they aren’t properties that a physicist would want to mention in giving an inventory of the characteristics that play a role in physical laws (2014, pp. 233-234).

Qualia, therefore, are characterized by Hill as physical properties of objects. The representationalist theory says that they are represented by the perceptual system in a certain way, just like every other property. Additionally, representationalism claims that how a phenomenal property appears to an observer constitutively depends on the way it is represented. Hill is a representationalist, thus he commits himself to the idea that there is a distinction to make regarding qualia: there are qualia-as-they-appear, that is, *representations of* Thouless properties, and qualia-as-they-really-are, that is, Thouless properties themselves. What we experience when an object looks$_P$ small to us is how the quale is represented, that is, the quale-as-it-appears.

This distinction allows Hill to claim, on one hand, that qualia belong to the physical world, and, on the other hand, that there is a sense in which it is legitimate to say that a truck seen at a distance looks small just in the same way as a miniature toy truck looks small in our hands. The two objects do not share the same Thouless property: we could say that the real truck at a distance *looks small*$_A$, while the toy truck *looks small*$_B$. However, the two Thouless properties are *represented* by our visual system *in the same way*.

The move of distinguishing qualia-as-they-are from qualia-as-they-appear seems to succeed in accommodating qualia in a naturalistic account of perception: qualia are physical properties, genuinely part of the natural world. A real truck and a toy truck do not share the same apparent size, since their apparent sizes are the results of different Thouless functions with different values assigned to the variables. Yet, our visual system represents the two Thouless properties in the same way. In other words, the real truck and the toy truck only *look as if they looked of the same size*. The sense in which that particular smallness is shared by the two appearances is due to the way the two appearances appear.

[11] Robert Thouless was a Scottish psychophysicist who in the 1930s modeled the perception of apparent sizes as a “compromise” between objective sizes and angular sizes. His idea became part of the “mainstream” thanks to the work of the famous vision scientist David Marr in the 1970s (see Marr, 1982).

One does not need to be a hard-core representationalist, neither does she have to be a philosopher, in order to see virtues in Hill's account. Hill seems to correctly capture the way in which we commonly characterize the phenomenology of our visual experiences. Take his description of the famous drawing by Mach (figure 1):

> I am in my study, sitting in a chair. [...] my hand looks much larger than the chimney; it looks slightly larger than the books; and it looks about the same size as my feet. In short, my current visual experience is very much like the one that Mach recorded in his famous drawing of how the world looked to him from the vantage point of a chair in his library (2014, pp. 225-226).

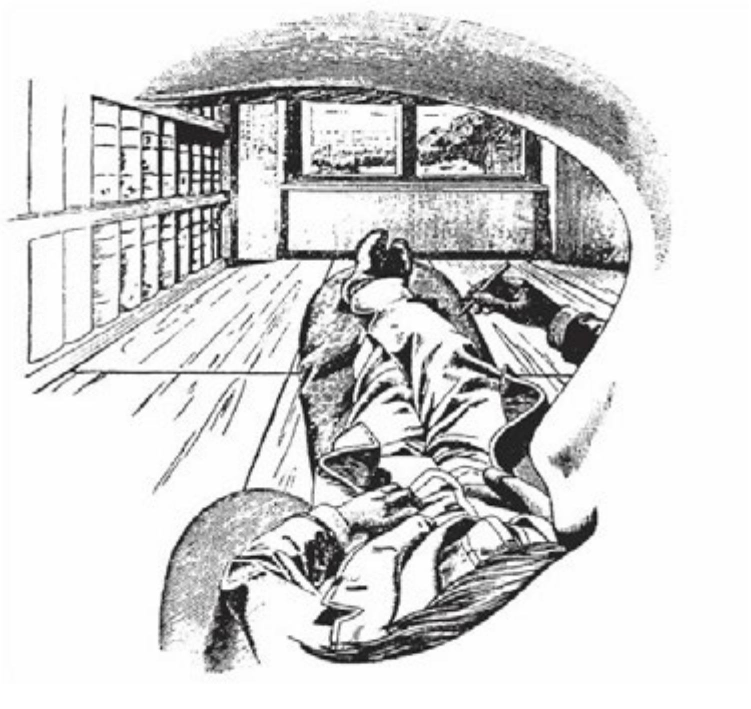

Figure 1: Mach's drawing. Hill (2014), p. 226.

Hill argues that, although the distinction between qualia-as-they-appear and qualia-as-they-are is strongly counter-intuitive, we just have to bite the bullet and commit to it, since it is a direct consequence of the representationalist view of perceptual experience, which in turn is the only view that can naturalize qualia, i.e. that can include qualia in the physical world. In short, if we substitute 'qualia' with 'appearance' – a move that Hill probably considers innocent, given his tendency to use the verb 'to appear' interchangeably with 'to look'[12] – Hill accepts that appearances have themselves appearances, or, in other words, that there are *appearances of appearances*: "A representational theory makes it possible to draw an appearance/reality distinction with respect to qualia. On the one hand, there are qualia-as-

[12] See Hill (2014), p. 198.

they-are-represented-by-experiential-representations. On the other hand, there are qualia-as-they-are-in-themselves" (2014, p. 210).

Now, it seems to me that the main advantage of Hill's introduction of Thouless properties was to bring qualia back into the physical world, thus dissolving the "mystery" lingering around them since Nagel (1974) and Jackson (1982) wrote their seminal papers on the topic. However, if Thouless properties aren't enough to account for everything that raised the problem of qualia in the first place – Thouless properties still cannot explain the specific way in which they are represented in experience – then it may be argued that the move isn't really worth making.

As Hill himself admits, there is still something that is left out by the identification of qualia with Thouless properties: namely, the way those properties look phenomenally to us.

Hill introduced Thouless properties into the picture so that no explanation of how we go from objective, mind-independent properties to properties that are dependent on our perspective in the world and other features of context. However, even after introducing Thouless properties, Hill needs to provide an explanation of *why* those properties are represented by our perceptual system so that they can phenomenally appear differently from occasion to occasion. In other words, he moved the explanatory burden from the "original" distinction between objective properties and perspective-dependent appearances to a distinction within the domain of appearances.

I happen to be one of those people for whom the notion of appearance of an appearance only brings further and unneeded complexity into an already complex picture, without thereby providing a big explanatory advantage. In fact, the notion of appearances of appearances, where appearances *simpliciter* already mark the difference between how things seem to us subjectively and how things really are, is just redundant, and I would rather do without it.

In other words, my objection to Hill is advocating for parsimony: if representationalism needs to draw an appearance/reality distinction also within the domain of phenomenology (i.e. with respect to qualia), then the theory postulates a further level of complexity in the structure of experience. Consequently, it might be wise to explore alternatives that avoid postulating the level of appearances of appearances without losing much in terms of explanatory power. The adverbialist picture I will now sketch offers a compelling way to avoid the redundant notion of appearance of an appearance, and still shows clear commitment towards the naturalizing project.

4. A Broadly Adverbialist Alternative

The kind of naturalization-friendly adverbialism I start from is defended by Chirimuuta (2015) with a specific focus on color. My goal is to show that what Chirimuuta argues about colors (a) can be extended to other features of the environment traditionally understood as properties of objects, such as size or shape, and (b) can provide a somehow radical solution to the "problem" of qualia by promoting a different way to think about what it is for an object to appear in a certain way to a perceiver.

The key thesis of Chirimuuta's color adverbialism is the following: colors are not properties of objects. Alternatively, colors are the *ways* in which the perceptual systems of certain animals, humans included, interact with the external world. According to Chirimuuta, colors are modifications of the *interaction* between an animal and its environment. This idea is spelled out with the help of the contrast between two hypotheses regarding the function(s) of color vision in the life of an animal with the appropriate visual system. The two models are called *coloring-in* and *coloring-for*, and are in turn tied to two different ways of conceiving the function of perception in general (pp. 69-99).

The *coloring-in* model of color perception falls into the broader "correspondence model" of perception, according to which the main function of perception is to *detect* invariant, mind-independent features of the environment. Perceptual systems have the function of making the perceiver's subjective experience *correspond* to what is "out there" in the most accurate way possible. According to this model, the visual scene is "colored in" as if it were a page of a coloring book.

It is quite easy to see how this model would be consistent with Hill's view of qualia. Perception works by representing the mind-independent world in a certain way. Phenomenal properties like looking yellow, or looking small, are mind-independent properties of the objects out there, that are represented by a visual system that aims at a correspondence between the subjective aspect of experience and the objects and properties out there. Consider again the two trucks. The real truck's phenomenal property is *looking small*$_A$, whereas the toy truck's phenomenal property is *looking small*$_B$. In order to achieve accuracy, the system "depicts" as many details of the scene as possible, just like a photograph (or Mach's drawing above) would do. The more detailed the representation, the more (real) features of the world are picked out, the more adaptive an animal's behavior based on that representation will be.

On the other hand, the rival *coloring-for* model fits in a framework that takes perception to be more concerned with utility than with correspondence/accuracy. In particular, this model suggests that perception aims at providing the organism with information that is *useful* for it in order to interact with its environment in an efficient way, and this may or may not entail presenting the subject with accurate representations corresponding to mind-independent objects and their features. This second model takes colors to be essential to the very construction of the perceived scene. For an animal equipped with cone cells in the retina, seeing a scene in colors isn't just an additional "embellishment" whose role is to improve the accuracy of the representation. 'Coloring for' means coloring *for utility*. Color vision presents features of the environment in a way that makes them immediately salient for the organism, depending on what the needs of that particular organism are.

For creatures with color vision, the scene is constructed *as colored, starting from colors*, and all the other features, such as size, shape, distance, etc. are constructed derivatively. In humans, for example, Chirimuuta points out that V4, i.e. the part of the visual cortex that was once identified as the "color center" due to the sensitivity of that neural population to wavelength information, has recently been found to have a much more diverse range of functions, such as perception of depth (stereopsis), perception of shape, or visual attention (p. 72).

By presenting a great deal of empirical work (pp. 78-98) regarding what she lists as the "functions of color vision" (p. 77), Chirimuuta argues that perception scientists are much more sympathetic with the coloring-for model, that is, the model according to which "color vision is a way of seeing *things* – flowers, tables, ladybirds – not, in the first instance, a way of seeing the *colors*" (p. 69). I take this to be at least a good reason to think that philosophical theories committed to the naturalizing project should be attracted to the utility-based model, too.

In the light of her arguments in favor of the coloring-for model, Chirimuuta then concludes that a certain type of adverbialism is a very plausible view of color ontology. According to "traditional" adverbialism, color properties are *ways* in which we perceive objects, adverbial modifications of our experience of objects. Canaries aren't yellow; they are experienced yellow-*ly*. If color perception is for utility instead of accuracy, the experience of seeing colors is part of what it is to be related to the world in a way that is useful, i.e. that promotes the animal's survival and reproduction.

What if we extended the account beyond colors? Just like we can say that a canary doesn't possess the color property 'yellow', I suggest that we think of all the other ways in which objects can perceptually appear to us as *not* being properties of objects, either. Analogously to colors, the shapes, sizes, etc. that we experience perceptually, that is, the perceptual shape-qualia or size-qualia, are nothing but modifications of the interaction between the subject and the environment, where such interaction is interest-relative, utility-based, and action-guiding.

Think again about the two trucks: we see the distant real truck and the miniature truck small-*ily* not because either of them has a particular property of 'looking small'. Rather, 'small-ily' is just the way in which we experience objects, whenever it is appropriate or useful for creatures like us to do so.

Hill (2014, p. 208) rejects adverbialist theories of experiential awareness – i.e. awareness of qualia – on the basis of the observation that, for the adverbialist, qualia would only be ways in which we are aware of something *else* (e.g. an object), and this is at odds with our actual experience. Clearly, Hill states, our experience is *of* qualia: when we experience pain, our experience is *of* the pain. Therefore, the adverbialist must allow for a non-experiential awareness of qualia, and this in turn forces her to account for this awareness either in terms of direct acquaintance (collapsing the theory into an acquaintance theory, which Hill already rejected), or in terms of judgment (making the theory a doxastic one, which Hill also rejects). However, let me briefly clarify why I think this objection is misguided, especially in the light of the radical "paradigm change" promoted by the kind of adverbialism I endorse.

First, from the fact that, according to adverbialism, qualia aren't themselves properties we can be aware of in the traditional sense, it does not follow that the adverbialist must concede a form of non-experiential awareness of qualia. In fact, an adverbialist like Chirimuuta would probably say the following: we should revise the way we understand the very structure of experience. Experiential awareness *in general* is not meant to be structured as classical representationalists want it to be, i.e. in the form of an attribution of properties to substances.[13] By itself, experience doesn't make us aware *of* anything. Rather, experiential awareness shapes our interaction with the environment so that we can act in it. Experiential awareness is *itself* the way in which a subject and an environment interact, and asking ourselves which properties we are experientially aware *of* while we experience is asking the wrong question. Adverbialism doesn't say that qualia are ways in which we experience something else. It says that qualia are the way in which we experience, period.

At the beginning, I stated that my plan for this paper was to, first, reject Hill's claim that representationalism is the only view capable of naturalizing qualia, and, second, propose a credible alternative. Once we refute a "paradigm" telling us that the perceptual system is a detection device, and we accept that it is more of an *interaction* device, suddenly we will notice that the need to identify the way things look with special kinds of properties possessed by objects loses most of its urgency. If the perceptual system doesn't aim at accuracy/veridicality in itself, but only insofar as accuracy is compatible with what is useful to the creature, why should we think that a representational model of the kind Hill, Burge, and others defend is the best way to explain how perceptual experience has a "grip" on the world? Why think that the structure of perceptual contents is necessarily one in which a property is attributed to a particular?[14]

The naturalistic adverbialism I sketched invites us to abandon the idea of a correspondence between the world and the representations produced by the perceptual system. Moreover, I urge that we stop thinking that, if objects have properties independently of our experience, experience *itself* has to "tell us" that objects have some kind of properties. If we abandon representationalism, we don't need anything that plays the role Thouless properties play in Hill's theory. In the framework I defend, we wouldn't have to say that the distant truck looks small because the property of looking small is a physical attribute the truck possesses

[13] See, in this respect, Burge's notion of "typing" (2010, p. 380).

[14] I am indebted to Alison Springle for articulating both the implications of a rejection of the accuracy-based model of perceptual content and the possibilities that this rejection opens.

in certain circumstances. Rather, the adverbialist claims that the property of looking small doesn't belong to the truck, but smallness is a modification of a particular "interaction-episode" between us and the world.

I am aware that more needs to be said in order for naturalistic adverbialism to stand as a solid alternative to Hill's proposal in a broad range of cases. Moreover, an adequate defense of this kind of adverbialism would require more detailed explanation of some core-notions, such as the one of "interaction episode". However, I believe that such further work can be done, and it will yield successful results for the naturalizing project more broadly. Naturalistic adverbialism promotes a radical change of paradigm regarding how we should think of the very role of perception, and of how experience relates to the world. In this framework, to naturalize qualia means to think of them as modifiers of the utility-based interaction between a creature and its environment.

REFERENCES

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

Chalmers, D. (2001). The First-Person and Third-Person Views (Part I). [Web log post]. http://consc.net/notes/first-third.html.

Chalmers, D. (2007). Phenomenal Concepts and the Explanatory Gap. In T. Alter & W. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press, 167-194.

Chirimuuta, M. (2015). *Outside Color*. Cambridge, MA: MIT Press.

Burge, T. (2010). *Origins of Objectivity*. New York: Oxford University Press.

Danks, D. (2014). *Unifying the Mind*. Cambridge, MA: MIT Press.

Dennett, D. (1990). Quining Qualia. In W. Lycan (Ed.), *Mind and Cognition*, Oxford: Blackwell, 519-548.

Dennett, D. (1991). *Consciousness Explained*. Boston, MA: Little, Brown and Company.

Dretske, F. (1988). *Explaining Behavior – Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Fodor, J. (1981). *Representations*. Cambridge, MA: MIT Press.

Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. (1994). *The Elm and the Expert*. Cambridge, MA: MIT Press.

Hill, C. (2014). *Meaning, Mind, and Knowledge.* Oxford: Oxford University Press.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, 32 (127), 127-136.

Jackson, F. (1986). What Mary didn't Know. *Journal of Philosophy*, 83 (5), 291-295.

Kim, J. (2006). *Philosophy of Mind* (2nd ed.). Boulder, CO: Westview Press.

Kriegel, U. (2002). Phenomenal Content. *Erkenntnis*, 57(2), 175-218.

Marr, D. (1982). *Vision.* San Francisco, CA: W.H. Freeman and Company.

Millikan, R. (1984). *Language, Thought and other Biological Categories*. Cambridge, MA: MIT Press.

Nagel, T. (1974). What Is it Like to Be a Bat?. *Philosophical Review*, 435-450.

Shoemaker, S. (1975). Functionalism and Qualia. *Philosophical Studies*, 27, 291-315.

Shoemaker, S. (1981). The Inverted Spectrum. *Journal of Philosophy*, 74, 357-381.

Sprevak, M. (2013). Fictionalism about Neural Representations. *The Monist*, 96(4), 539-560.

Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.

MARCO VIOLA
IUSS Pavia and Vita-Salute San Raffaele University
marco.viola@iusspavia.it

CARVING MIND AT BRAIN'S JOINTS. THE DEBATE ON COGNITIVE ONTOLOGY[1]

abstract

Since neuroimaging methods allow researchers to study the human brain at work, the vexed mind-brain problem ceased to be just a metaphysical issue, and became a practical concern for Cognitive Neuroscientists: how could they carve mind and brain into distinct entities, and what is the relation between these two sets? In this paper, I discuss the classical model of one-to-one mappings between mental and neural entities, inherited from phrenology, and make its assumptions explicit. I then examine the shortcomings of this "new phrenology", and explore two solutions to them: the first accepts many-to-many mappings, whereas the second proposes a radically rethinking of the relata of this correspondence.

keywords

philosophy of neuroscience, philosophy of psychology, cognitive ontology, one-to-one mapping, Mind-Body problem

Phenomenology and Mind, n. 12 - 2017, pp. 162-172
DOI: 10.13128/Phe_Mi-21115
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

In some cases, when thorny epistemological or foundational issues concerning some specific[1] science are addressed, the boundaries between philosophy and (that specific) science are blurred, and philosophical contributions cannot be distinguished from scientific discourse anymore – thus realizing a sort of Quinean ideal of continuity.
This is currently happening in the ongoing debate on cognitive ontology, where neuroscientists (e.g. Pessoa, 2014; Poldrack, 2010) and philosophers (e.g. the contributions in the special issue edited by Janssen *et al.*, 2017) are discussing the proper role of neuroscientific evidence in determining what mental entities there are. In this paper I sketch a brief reconstruction of this debate, focusing on the controversial assumption of one-to-one mapping between mental and neural entities. I will address the shortcomings of this assumption, and briefly describe an alternative framework that does without it. I will also discuss some attempts to save the one-to-one mapping hypothesis by revising the *relata*, i.e. by proposing that mappings occur between different entities from those that are usually conceived.

1. Basic Ontological Assumptions of Cognitive Neuroscience

In some senses, the cognitive ontology debate is but the heir of the vexed Mind-Body problem. What kind of relationship holds between "the mental" and "the physical" is one of the most debated topics throughout the history of Western philosophy. However, as modern tools and techniques for studying the human brain and mind were developed, the Mind-Body problem has been (re)framed as a methodological question concerning the proper relation between psychology and neuroscience. Inheriting phrenologists Gall and Spurzheim's bold thesis that some mental faculties are localized in some areas of the brain, filtered and refined by figures such as Broca, twentieth century neuropsychologists appealed to the co-occurrence between cognitive impairments and brain lesions (observed post-mortem) to establish some link between mental processes and neural areas. They also looked for counter-intuitive cases in which two (allegedly) unified systems could be selectively impaired in order to establish that they are independent on functional and (usually, though not always) structural grounds. The most famous case was that of memory: once thought to be a single capacity, the observation of

[1] I am grateful to Joe Dewhurst, Anneli Janssen and Colin Klein for their valuable comments, as well as for our many stimulating discussions. I am also indebted to three anonymous referees who helped me to ameliorate the intelligibility of the paper.

brain lesioned patients led to its being *split* into different kinds, e.g. "declarative" *vs* "implicit" memory (see for instance Squire, 1992).

However, by the end of the twentieth century, several techniques were developed, that allowed scientists to study living human brains at work, either by measuring electrical signals on the scalp (EEG, MEG) or by assessing the metabolic activity of cortical areas (PET, fMRI). More recent techniques also allow researchers to produce temporary interferences on the activity of some brain areas (TMS, tDCS). Each of these techniques represents a new window[2] through which the psychologist can peek at the brain at work. By accepting some bridge-principles, she can then use this new evidential source to add further constraints to her cognitive theories – each of them positing slightly different carvings-up of the mind (that is, different taxonomies of mental entities; see below). And so many did, thus giving rise to Cognitive Neuroscience.

A common feature of cognitive neuroscience is that neuroscience is deemed a legitimate arbiter for refining cognitive ontology, i.e. for choosing the right set of mental entities. This features rely on the following assumptions:

(1) the mind can be decomposed into distinct mental entities;
(2) the brain can be decomposed into distinct neural entities;
(3) some systematic correspondence holds between mental and neural entities.[3]

Assumptions (1) and (2) have both been challenged on the ground that mental and neural processes are too tightly interconnected to be studied in isolation (see e.g. van Orden and Paap, 1997). While not completely settled, a certain degree of decomposability of mind seems reasonably warranted by the successes of Cognitive Psychology, whereas the decomposability of the brain seems reasonable in virtue of the fact that neuropsychology has proved that the brain cortex is not equipotential. Moreover, albeit mechanistic explanations (regarded by most researchers as the proper kind of explanations in cognitive neuroscience) often begin with decomposition into parts (Bechtel & Richardson, 1993), they need not be insensitive to mutual interactions between such parts. Quite on the contrary, sophisticated mechanistic frameworks involve both a *de*composition and a *re*composition of phenomena, as well as a detailed characterization of the overall context (Bechtel, 2009).

Mind and brain can also be carved up independently of one another. This may result in several conflicting taxonomies. For example, the faculties posited by Thomas Reid differ from the systems described in modern psychology textbooks, and both are different from the taxonomies of folk psychology, which nonetheless underlie much of our everyday language about mind and behavior. Brains too can be carved along many joints: for instance, brain cartographers can distinguish among lobes, on the basis of gross spatial properties; or they can distinguish brain regions, according to either a single or a mix of properties such as their histology or their connectivity profiles (Mundale, 2009). Assumption (3) offers a possible way out from the underdetermination of both kinds of ontology, providing a (further) criterion for choosing one out of these many possible carvings: namely, it prescribes that we should choose

[2] Though they are very opaque windows. While we cannot address these topics here, it is worth remembering that each technique has its limitations, and that the assumptions necessary in order to interpret the data are far from being theoretically neutral. Rather than conceiving of them as tools for *seeing* brain activity, it is thus more prudent to conceive of neuroimaging techniques as tools for *inferring* brain activity. See e.g. Roskies (2008) and Klein (2010).

[3] Notice that, while obviously easier to reconcile with a materialist metaphysics, assumption (3) does not entail it: all that is required is some kind of psycho-neural bridge principle. See Nathan & Del Pinal (2016).

those that warrant systematic mappings between entities of each domain.[4]

Whilst assumptions (1-3) are arguably shared by every Cognitive Neuroscientist, they are often construed in different ways. In the following sections I distinguish and compare between two such ways, giving rise to two different ontological frameworks.

2. The Neo-Phrenological Framework of Cognitive Neuroscience

During the early years, and until recent criticisms have cast shadows upon them, the ontological desiderata underlying most studies in cognitive neuroscience were efficaciously stated by neuroscientists Price & Friston (2005), who claim that we should aim for "a systematic definition of structure–function relations whereby structures predict functions and functions predict structures" (p. 263). They stress that both functions and structures can be described at multiple levels of abstraction, and propose that the best level of abstraction is that which allows us to assign a single function to each structure (see also Rathkopf, 2013). Within this framework, assumptions (1-3) are thus specified as such:

> (1a) the mind can be decomposed into distinct mental entities $(m_1, m_2, \dots m_n)$, i.e. *mental functions*;
> (2a) the brain can be decomposed into distinct neural entities $(n_1, n_2, \dots n_n)$, i.e. *neural structures*;
> (3a) a *one-to-one mapping* holds between each mental entity and a given neural entity $(m_1 \leftrightarrow n_1, m_2 \leftrightarrow n_2, \dots m_n \leftrightarrow n_n)$.

This framework has been derogatorily dubbed *The New Phrenology* by some critics (notably, this is the title of Uttal's 2001 book). Indeed, this label is not totally unreasonable: the practice of mapping mental entities to some neural entity was introduced by the phrenologists, and continued through the twentieth century thanks to the work of physicians such as Broca and Ferrier, who in turn passed it to modern scientists. Therefore, I shall call this ontological framework "neo-phrenology".

However, modern neo-phrenologists differ from their ancestors in that they have thoroughly refined their ontology; they no longer try to associate the thickness of bumps on the skull to some of the disputable mental faculties posited by Gall, but rather seek to map inner neural regions to some mental function, i.e. usually psychological constructs that best fit the models of Cognitive Psychologists (Zawidzki & Bechtel, 2004).

Still, in our current ontology, virtually any attempt to map mental functions onto neural structures revealed a many-to-many mapping. Price and Friston are obviously aware that current theories are far from being as well ordered as they desire. But their claim is not meant to describe the current ontology. Rather, it is meant to play a heuristic function (Bechtel & McCauley, 1999), i.e. to prescribe how future ontologies should look (*fig. 1*).

[4] Usually, reductionist philosophers assume that the relation between psychology and neuroscience is merely or mostly bottom-up, that is, it is only/mainly neuroscience that influences the psychological categories (e.g. Bickle, 2006). However, historically brain cartographies have been widely driven by considerations concerning the function, i.e. ultimately the psychological role of brain regions (see Hatfield, 2000). In drawing his notorious brain maps, Brodmann relied on histological criteria because "tissue elements of uniform specific structure, whether they are limited to a large or small cortical field or diffusely distributed over the whole cortex, must also have a uniform physiological function, and thus [...] such elements are to be regarded as not only morphologically but also functionally equivalent" (2006, p. 5). Functional significance, along with anatomical criteria, is still at play in modern mappings (notably, Glasser *et al.*, 2016).

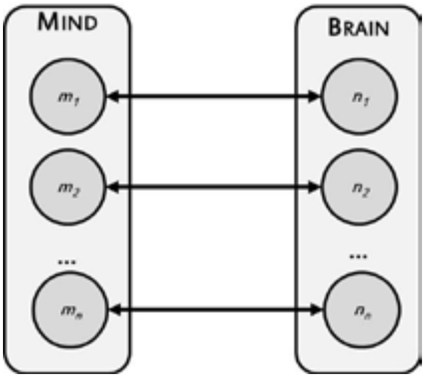

Figure 1. A stylized representation of the ideal ontology envisaged by the neo-phrenology: both Mind and Brain are decomposed into a set of discrete entities, each one standing in a one-to-one mapping with a member of the other set.

The moral of their story then is that, whenever there is no function-structure one-to-one relation, either the function or the structure (or both) are to be reformed or discarded. To put it into Lakatos's (1970) terms, rather than giving up the core assumptions (1a-3a), we can "put the blame" on one of the following auxiliary assumptions:

(4) the correct set of *mental functions is* $M = \{m_1, m_2, \ldots m_n\}$;
(5) the correct set of *brain structures* is $N = \{n_1, n_2, \ldots n_n\}$.

Either (4) or (5) or both are then replaced by either one or both the following auxiliary assumptions:

(4*) the correct set of *mental functions is* $M^* = \{m^*_1, m^*_2, \ldots m^*_n\}$;
(5*) the correct set of *brain structures* is $N^* = \{n^*_1, n^*_2, \ldots n^*_n\}$,

where 'the correct set' means 'the set that is better at preserving (3)' (*fig. 2*).

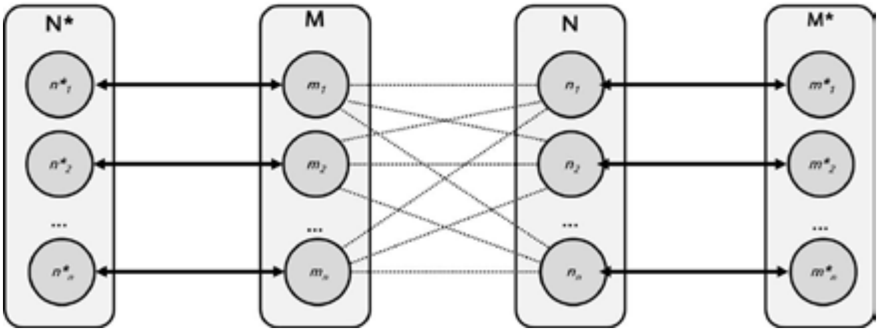

Figure 2. Whenever one-to-one mappings cannot be established between the members of a set of mental entities *M* and those of a set of neural entities *N*, neo-phrenology tells us to discard either (or both), in favor of different sets *M** and/or *N**, in order to reestablish one-to-one mappings between mental and neural entities.

For instance, the left posterior lateral fusiform gyrus has been associated with processing of visual words, and was thus called Visual Word Form Area. However, as Price and Friston report, this area is also engaged in processing the visual attributes of animals, and in tactile-visual integration. Thus, they propose that a better functional label for that area is one that characterizes its working at a more abstract level: *sensorimotor integration*. Similarly, since the left inferofrontal gyrus (Broca's area) is found to be activated in many domains outside the linguistic one, Tettamanti & Weniger (2006) propose to reconceive its function as a "supramodal hierarchial processor".

Yet, as Klein (2012) observes, a functional label as vague as 'sensorimotor integration' is not very informative, since "at some level of abstraction, that's what nearly all of the cortex does" (p. 5). Thus, rather than some too-abstract-to-be-purposeful definition, Klein argues that functional labels for brain regions are only meaningful when they are considered in the light of some given context. He then construes this context as neural context, i.e. the set of areas that are co-activated. In the end, he comes up with a prudent endorsement of an ongoing shift of emphasis about the level at which functions and structures ought to be mapped: from single regions to many regions gathered into functional networks (Bressler & Menon, 2010).

Such a shift was also fostered by the development of Multi-Variate Pattern Analysis (MVPA), statistical techniques that allow scientists to compare activation patterns spread across the whole brain, instead of single regions (Haxby *et al.*, 2014). The take-home message of such techniques is that the functional activity of the brain is better understood when complex arrays of activity are taken into account. Most of these MVPA employ machine learning techniques to perform what has come to be known as brain reading: in order to train a classifier to predict which cognitive function is performed from the corresponding neural activation, researchers "feed" it with functionally labelled patterns. After that, classifiers are shown unlabeled patterns of neural activation, and are asked to "guess" which one (among a given set) is the correct functional label, based on the similarity to the training set; a task in which they often succeed far above the chance level, sometimes even across different subjects. By applying such a method to a large database, Lenartowicz *et al.* (2010) investigated whether some cognitive constructs from the domain of control functions were mapped consistently to some specific and distinguishable activation pattern. As the classifier struggled to discriminate the pattern associated with "task switching" from those associated with some other constructs, the researchers suggest replacing "task switching" with some other constructs that map more neatly with some discriminable neural basis.

3. A Post-Phrenological Framework

Notwithstanding the reforming attempts of some researchers, many-to-many function-structure mappings are still ubiquitous. While Bechtel & Mundale (1999) claimed that Multiple Realizability vanished in the light of the empirical success of comparative Neuroscience, many mental functions turned out to be *degenerate*, i.e. they can be implemented by multiple brain structures – which, according to Figdor (2010), might sometimes count as multiple realizability. Meanwhile, while not *equipotent*, all brain regions are found to exhibit *pluripotency*, i.e. to be involved in the implementation of several (apparently) different mental functions. Degeneracy and pluripotency contradict the one-to-one mapping assumption (3a), and therefore neo-phrenology regards them as two anomalies to deal with by revising the ontology. Yet, despite many years of reformistic efforts, including the shift from regions toward networks, degeneracy and pluripotency seem to be here to stay. The neuroscientist Pessoa (2014) pessimistically upholds that

> the attempt to map structure to function on a one-to-one manner in terms of networks will be fraught with similar difficulties as the one based on brain regions [...] – the

> problem is simply passed along to a higher level. Thus, two distinct networks may generate similar behavioral profiles ([...] many-to-one); a given network will also participate in several behaviors (one-to-many) (p. 408).

Thus, some thinkers bite the bullet and try to sketch an ontological framework that does without one-to-one mappings. The most mature formulation up to date is presented in Anderson's 2014 book *After Phrenology* (see also Barrett & Saptute, 2013). Stressing the often-neglected phenomenon of neural plasticity, both at long and at short timescales, Anderson argues that brain structures have no such things as "intrinsic functions". Rather, they get their functional significance depending on their structural characteristic, and on the partnership they institute with other neural structures, gathering into transient neural assemblies. Structures that usually play some role are thus commonly redeployed for other functions – a principle that Anderson dubs *neural reuse*.

Since neural reuse can hardly if ever be made consistent with one-to-one mapping, Anderson sketches a new protocol for linking mind and brain. The relation between brain and mind, he claims, could be established by measuring the functional dispositions of each neural structure (be it a region or a network), i.e. the likelihood that such structure gets (significantly) activated when some task is undertaken that bears the label of a given function or cognitive domain. In other words, (3a) is rejected and (3) is rather construed as (3b):

> (3b) for each structure *s*, and for each function (or domain) $f_1, f_2, ... f_n$, there is a probability *P* that *s* is engaged ($P(s,f_1), P(s,f_2), ... P(s,f_n)$).

By exploiting available databases of neuroimaging data, it is possible to provide robust measurements of each structure's disposition to be engaged in various functions/domains, obtaining what Anderson calls a *functional fingerprint*. Statistical dimension techniques could then be used to "dig out" similarities, and cluster related functions into what he calls *Neuroscientifically Relevant Psychological (NRP) factors* – which Anderson presents as an analogue of "personality traits" for neural structures.

An interesting feature of this approach is that, because the only direction of ontological revision is from brain data to mental categories, it might spur radical revisions in psychological taxonomies, leading to notice similarities and distinctions that could have been otherwise foreshadowed by our folk psychological prejudices.

Groundbreaking as it is, this post-phrenological framework has its weak spots too. For instance, it is very conservative with respect to neural ontology (McCaffrey & Machery, 2016), i.e. it does not offer incentives to find new and more interesting neural entities (see below). Moreover, whenever different functions are inadvertently given the same functional label (which is frequent across different labs; see Sullivan, 2016), Anderson's big data approach seems ill suited to recognize the distinction (Kaplan & Craver, 2016). Nonetheless, since it is still in its infancy, no doubt post-phrenology will arguably address these and other shortcomings. Would neo-phrenology be completely supplanted by post-phrenology, or can the old framework still be saved somehow?

4. What Reforms (If Any) Can Save Neo-Phrenology?

While many are pessimistic about the fate of one-to-one mapping (3a), I think that it is worth attempting to rescue it, in order not to throw the baby out with the bath water. As stressed by Bechtel & McCauley (1999), even gross function-structure mappings, inasmuch as they approximate the truth, may provide the basis for posing further research questions. For instance, while a claim such as "vision happens in the striate cortex" is largely imprecise, it was a starting point for further finer-grained models. Thus, once properly refined, one-to-one mapping might still represent a viable heuristic for prompting ontological revisions.

Whilst much emphasis has been placed upon revising mental entities, revising the neural entities might also be possible. Indeed, while both lesion studies and neuroimaging biased researchers toward assuming contiguous brain regions as the basic entities, due to their spatial resolutions, it is possible that two (or more) functionally coherent and intertwined neural populations can co-exist within what is currently classified as "the same" brain region.

McCaffrey (2015) stresses that various areas labelled as multifunctional (i.e. pluripotent) might differ in their mechanistic organization, thus requiring various explanatory strategies: some areas contain different structures, and thus require the abovementioned "divide-and-conquer" approach; others conserve the same role in different tasks, thus qualifying for a more abstract redefinition of their function such as that advocated by Price & Friston (2005, see §3); but for yet other areas, context-sensitive mappings are unavoidable.

Therefore, perhaps replacing some specific set of either or both of mental function and/or of neural structure (i.e., replacing either or both assumptions 4a and 5a with 4b and 5b) might be insufficient. A more productive approach may be that to stop seeking one-to-one mappings *between structures and functions*, and rather pick some wholly different kind of *neural* and/or *mental entities*. After all, this is how phrenology turned into *neo*-phrenology. For instance, after having stressed that a same set of brain regions could support several functions, Pessoa (2014) speculates that different functions can be due to differences in the strengths of the connections between them, or by the different time course of their activation.

Similarly, in Viola & Zanin (forthcoming) I proposed another kind of reform in neural ontology, i.e. dropping

> (2a) the brain can be decomposed into distinct neural entities (n_1,n_2, … n_n), i.e. *neural structures*,

and replacing it with

> (2b) the brain can be decomposed into distinct neural entities, i.e. *neural processes* (p_1,p_2, … p_n),

and further specifying that

> (6) each neural process *p* is defined as the activity of some (set of) neural structure(s) *n* when acting according some way of working *w (n=s,w).*

The notion of a "way of working" is introduced in order to account for the intuition that the same part (or set of parts) of a mechanism can implement multiple functions (Bechtel & Abrahamsen, 2005). This proposal thus implies that the functions of brain mechanisms, rather than their parts, should be the *relata* of one-to-one mappings.

Admittedly, "way of working" is a vague notion. Given that, paraphrasing Price and Friston, we want to be able to predict functions from structures *plus something* (and vice versa), we cannot discriminate distinct ways of working on the basis of their contribution to mental functions, since that would entail a circularity. Rather, we need some purely neural marker for discriminating ways of working. In Viola & Zanin (forthcoming) I tentatively adopted the suggestion advanced by Siegel *et al.* (2012, p. 121) that "frequency-specific correlated oscillations in distributed cortical networks may provide indices, or 'fingerprints', of the network interactions that underlie cognitive processes". Therefore, I proposed that different oscillatory patterns represent distinct ways of working.

The expected outcome of such framework might be represented as in *fig. 3*:

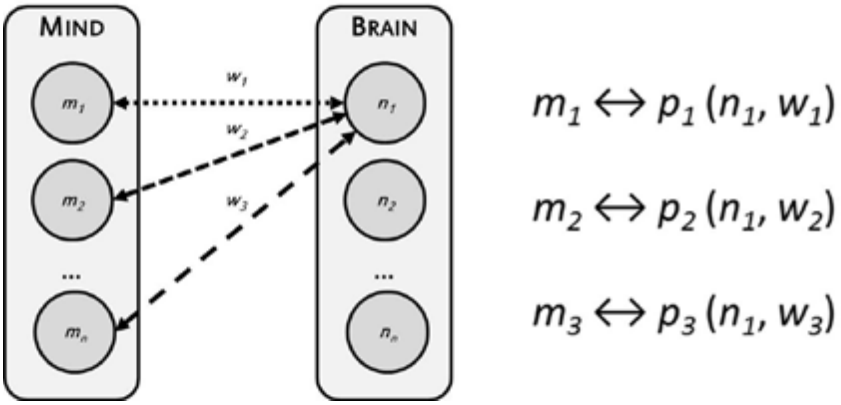
MIND
BRAIN
m_1
m_2
...
m_n
n_1
n_2
...
n_n
w_1
w_2
w_3
$m_1 \leftrightarrow p_1(n_1, w_1)$
$m_2 \leftrightarrow p_2(n_1, w_2)$
$m_3 \leftrightarrow p_3(n_1, w_3)$

Figure 3. The revision of the neo-phrenology proposed in Viola & Zanin (forthcoming). The *m*s, *n*s,*w*s, and *ps* represent, respectively, mental functions, neural structures (or sets of neural structures) and ways of workings, which I propose to construe in term of neural oscillations. Mental functions cannot be put into one-to-one correspondences with (sets of) neural structures anymore; rather, (sets of) neural structures become one of the two properties that define neural processes, along with a specific way of working.

Such a strategy allows us to reconcile pluripotency and one-to-one mapping. However, notwithstanding this specific proposal, I am more concerned with demonstrating that, notwithstanding its anomalies, neo-phrenology (of some evolved version, say neo-phrenology*) can still play a role in Cognitive Neuroscience.

Ultimately, as suggested by Anderson and Pessoa (among others), it is possible that the Brain-Mind relation turns out to be so complex that no one-to-one mapping whatsoever will be tenable for organizing existent knowledge. In other words, it is not implausible that one-to-one mappings will eventually vanish from the context of justification. Even so, hypothesizing and testing some sophisticated one-to-one mapping might still play some role in the context of discovery, e.g. by challenging the researchers to reshape the taxonomical landscapes of mind and brain.

REFERENCES

Anderson, M.L. (2014). *After phrenology*. Cambridge, MA: MIT Press.

Barrett, L.F., & Satpute, A.B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current opinion in neurobiology*, 23(3), 361-372.

Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, 22(5), 543-564.

Bechtel, W. & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421-441.

Bechtel, W. & McCauley, R.N. (1999). Heuristic identity theory (or back to the future): The mind-body problem against the background of research strategies in cognitive neuroscience. In Hahn, M., & Stoness, S.C. (Eds.), *Proceedings of the 21st annual meeting of the cognitive science society*. Mahwah, NJ: Erlbaum, 67-72.

Bechtel, W. & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of science*, 66(2), 175-207.
Bechtel, W. & Richardson, R.C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Boston, MA: MIT Press.
Bickle, J. (2006). Reducing mind to molecular pathways: Explicating the reductionism implicit in current cellular and molecular neuroscience. *Synthese*, 151(3), 411-434.
Bressler, S.L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences*, 14(6), 277-290.
Brodmann, K. (2006). *Localization in the Cerebral Cortex* (Eng. Transl. by L. J. Garvey.). New York, NJ: Springer. (Original work published in 1909).
Figdor, C. (2010). Neuroscience and the multiple realization of cognitive functions. *Philosophy of Science*, 77(3), 419-456.
Glasser, M.F., Coalson, T., Robinson, E., Hacker, C., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jankinson, M., Smith, S.M., & Van Essen, D.C. (2016). A Multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615), 171-178.
Hatfield, G. (2000). The brain's "new" science: Psychology, neurophysiology, and constraint. *Philosophy of Science*, 67, S388-S403.
Janssen, A., Klein, C., & Slors, M. (2017). What is a cognitive ontology, anyway?. *Philosophical Explorations*, 20(2), 123-128.
Kaplan, D.M., & Craver, C.F. (2016). A registration problem for functional fingerprinting. *Behavioral and Brain Sciences*, 39, 15-16.
Klein, C. (2010). Images are not the evidence in neuroimaging. *The British Journal for the Philosophy of Science*, 61(2), 265-278.
Klein, C. (2012). Cognitive ontology and region-versus network-oriented analyses. *Philosophy of Science*, 79(5), 952-960.
Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. In Lakatos, I., Musgrave, A. (Eds.), *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, 91-195.
Lenartowicz, A., Kalar, D.J., Congdon, E., & Poldrack, R.A. (2010). Towards an ontology of cognitive control. *Topics in Cognitive Science*, 2(4), 678-692.
McCaffrey, J.B. (2015). The brain's heterogeneous functional landscape. *Philosophy of Science*, 82(5), 1010-1022.
McCaffrey, J.B. & Machery, E. (2016). The reification objection to bottom-up cognitive ontology revision. *Behavioral and Brain Sciences*, 39, 16-18.
Mundale, J. (2009). Epistemic Preliminaries: Normative Priorities and Neuropsychological Kinds. *Humana Mente*, 11, 1-9.
Nathan, M.J., & Del Pinal, G. (2016). Mapping the mind: bridge laws and the psycho-neural interface. *Synthese*, 193(2), 637-657.
Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of life reviews*, 11(3), 400-435.
Poldrack, R.A. (2010). Mapping mental function to brain structure: how can cognitive neuroimaging succeed?. *Perspectives on Psychological Science*, 5(6), 753-761.
Rathkopf, C.A. (2013). Localization and intrinsic function. *Philosophy of Science*, 80(1), 1-21.
Roskies, A.L. (2008). Neuroimaging and inferential distance. *Neuroethics*, 1(1), 19-30.
Siegel, M., Donner, T.H., & Engel, A.K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, 13(2), 121-134.
Squire, L.R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of cognitive neuroscience*, 4(3), 232-243.
Sullivan, J. (2016). Neuroscientific kinds through the lens of scientific practice. In C. Kendig

(Ed.), *Natural Kinds and Classification in Scientific Practice*. New York: Routledge, 47-56.
Tettamanti, M. & Weniger, D. (2006). Broca's area: a supramodal hierarchical processor?. *Cortex*, 42(4), 491-494.
Uttal, W.R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT press.
van Orden, G.C., & Paap, K.R. (1997). Functional neuroimages fail to discover pieces of mind in the parts of the brain. *Philosophy of Science*, 64, S85-S94.
Viola, M. & Zanin, E. (forthcoming). The standard ontological framework of cognitive neuroscience: Some lessons from Broca's area. *Philosophical Psychology*, DOI: 10.1080/09515089.2017.1322193.
Zawidzki, T. & Bechtel, W. (2005). Gall's legacy revisited. Decomposition and localization in cognitive neuroscience. In Emering, C.E. & Johnson, D.M. (Eds.), *The mind as a scientific object: Between brain and culture*. New York: Oxford University Press, 293-318.

JOANA RIGATO
Champalimaud Center for the Unknown - Lisbon
joana.rigato@neuro.fchampalimaud.org

LOOKING FOR EMERGENCE IN PHYSICS[1]

abstract

Despite its recent popularity, Emergence is still a field where philosophers and physicists often talk past each other. In fact, while philosophical discussions focus mostly on ontological emergence, physical theory is inherently limited to the epistemological level and the impossibility of its conclusions to provide direct evidence for ontological claims is often underestimated. Nevertheless, the emergentist philosopher's case against reductionist theories of how the different levels of reality are related to each other can still gain from the assessment of paradigmatic examples of discontinuity between models in physics, even though their implications must be handled with care.

keywords

emergentism, ontological emergence, epistemological emergence, reductionism, physics, singular limits

Phenomenology and Mind, n. 12 - 2017, pp. 174-183
DOI: 10.13128/Phe_Mi-21116
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

LOOKING FOR EMERGENCE IN PHYSICS

1. Introduction

Emergentism, in its various forms, is the view according to which there are features of reality (properties, objects or laws) that are irreducible to the lower-level basis from which they emerge, in the sense that they are more than just the result of the combination of the system's parts and their interactions. These features are paradoxically (or so it seems) both *dependent* on and *autonomous* from their emergence base, i.e. from the lower-level that brings them about. They are dependent in the sense that they cannot exist unless the lower-level structure is in place, but they are autonomous insofar as that structure does not suffice for the novel features that arise to be fully explained and predicted. This inexplicability and unpredictability is, of course, in the eyes of the beholder and amounts to what is commonly known as epistemological emergence. Still, metaphysical or ontological emergence can be assumed to underlie this apparent underivability, in which case there is a radical discontinuity in the hierarchical organization of reality, whereby the causal effects of emergent phenomena in the world cannot even *in principle* be accounted for by the causal powers of the lower-level structure on which they depend (Kim, 1999; O'Connor & Wong, 2005).

This type of theory has been developed in very many areas, from different scientific branches to philosophy. In the latter field, its study ranges from metaphysics and the philosophy of mind (in which the focus has been mainly on the relationship between mental/conscious entities and their physiological substrate – Kim, 1993; Searle, 1992), to philosophy of science in general (where the focus is on whether certain theories are reducible to others or not – Bedau, 1997), to philosophy of physics (where emergence seems to be a good conceptual tool for explaining nonlinear phenomena – Andersen, 1972) and philosophy of biology (where the main interest is top-down causation in self-organizing systems – Arp, 2008).

What is common in all these fields is the opposition to the reductionist ideal of a Lego world where the elements of the bottom-most domain provide necessary and sufficient conditions for the phenomena taking place at higher levels of organization. According to reductionism, the lower-level properties and laws of a system determine its upper-levels properties and laws. The instantiation of the former necessitates the instantiation of the latter. This implies, as a consequence, that the scientific domain that explains the lower-level occurrences is *in principle* sufficient to explain the upper-level ones (Sober, 1999). Emergentist views all counter this

[1] Acknowledgments: I thank Margarida Telo da Gama, Robert Bishop and Pranab Das for helpful discussions and comments on earlier drafts of this paper.

perspective with various examples that are used as evidence for the existence of upper-level phenomena that challenge microphysical explanation, not only in practice, but allegedly in principle.

However, needless to say, there is no uniformity in the way the concept of emergence is used and in the candidates that are accepted as good examples of emergent phenomena (Bedau & Humphreys, 2008).

In this paper, I chose to focus on the debate about emergence going on in philosophy of physics, hoping to show its relevance to the general discussion.

2. Emergence in Physics

A commendable tendency in the past years in philosophy of science has been to develop accounts of emergence that move away from armchair metaphysics and anchor philosophical analyses in scientific theory and practice.

Philosopher and physicist Robert Bishop, for example, has been working in questions of Emergence and Complex Systems for many years. Alone (2005, 2009) or together with theoretical physicist Harald Atmanspacher (2006), Bishop developed an account of what he calls "contextual emergence", which is a relation between different levels of description (in its epistemological form) or between domains of reality (in its ontological form), whereby the description, properties or behaviors of the lower domain provide some necessary but no sufficient conditions for the novelty existing at the upper-level. The remaining conditions must be provided by the context, which includes the stability conditions of the emergent states and observables (i.e. the conditions that guarantee their existence and persistence), which are not given by lower-level descriptions. Bishop uses several examples as evidence for the ubiquity of contextual emergence, from the domain of quantum chemistry to that of human society. All of them have to do with scale transformations: how the laws of microphysics give rise to the laws and properties of the macro world.

Bishop is in good company, as voices have been rising in the attempt to tell philosophers and unexamined reductionists that real-world science does not actually have any models that drill down from many-body physics to some mythical microphysical state and that, in fact, productive scientific models largely ignore such thinking altogether. According to nobel laureate Robert Laughlin, for example, the idea that one might in principle deduce the goings-on in the domains of chemistry, biology and other special sciences from a complete knowledge of particle physics is totally unfounded and, even though reductionism is a belief that is central to much of physical research, "the safety that comes from acknowledging only the facts one likes is fundamentally incompatible with science. Sooner or later it must be swept away by the forces of history" (Laughlin & Pines, 2000, p. 264).

Let us now look at some examples of emergence in physics that have been put forward in the literature. The examples challenge the reductionist ideal from different fronts: first, the need for singular limits, for example in the transition from quantum to classical mechanics (section 2.1), is used to call into question the idea that a macro-physical state can be derived from a micro-physical equation. Second, the case of criticality, which is an example of universality (section 2.2), questions the assumption that a macro-physical state has a unique microphysical basis, which is commonly taken as a consequence of a reductionistic world. Finally, the phenomenon of liquidity (section 2.3) highlights the role of stability conditions which are often not provided by the underlying emergence base.

2.1. When a Macro State Cannot Be Derived From a Micro Equation

The transition from quantum to classical mechanics is a mysterious one. Mathematically, these two realms are separated by singular limits (mathematical expansions in which some quantities are assumed to tend either to zero or to infinity), which means that the transition between the formalism that describes the behavior of particles at the quantum level (the Hamiltonian dynamics) and the equations used in the field of many-body physics is discontinuous. The behavior before and after that transition is qualitatively different and has to be described by a totally distinct equation. And in order to move from one equation to the other, Planck's constant is assumed to tend to zero, which is a mathematical trick that departs from reality (where it is actually non-zero). Hence, between the classical and the quantum domains there is a radical epistemological gap which can be bridged only with the help of a formal artifact.

This can be illustrated with the example of molecular shape. Isomers are molecules that share identical chemical formulas but have different spatial arrangements which give them very different properties. According to Bishop, these are good candidates as examples of contextually emergent phenomena since the specific structure into which a certain quantum description (the so-called Hamiltonian) will evolve at the chemical level cannot be deduced from quantum mechanical data alone.

> Even though QM [i.e. quantum mechanics] contains necessary conditions in terms of nucleons, electrons and their properties, fundamental force laws and so forth, observables relevant for molecular structure do not exist in the domain of QM. For such observables to obtain, an additional context not given by QM must be specified (Bishop, 2009, p. 177).

Only with the help of heuristic formal procedures, like assuming the nucleus of the atom to be stationary and infinitely larger than the electron mass, can one derive the equation encoding molecular shape.[2] This means that the chemical context (the stability conditions of a "clamped nucleus" together with the ratio of the electron mass over the nucleus mass tending to zero) must be fed into the mathematical treatment of the quantum mechanical information. It is thanks to these constraints that come from "outside" the quantum realm that the quantum correlations between nuclei and electrons are broken and classical position and momentum observables, as well as molecular shape, can arise.

2.2. When a Macro State Is Compatible With Multiple Micro States

Several natural phenomena share with the previous case this feature of being theoretically dependent on mathematical tricks, such as the postulation of the infinite or null value of certain observables that we know to be finite. They are cases in which the appearance of the macro property is not a mere quantitative derivation from a smaller scale to a larger scale, but rather a qualitative transformation which can be explained and predicted (at least on the basis of the models and theories presently available to us) only through the artificial normalization of singular limits.

According to Robert Batterman, a leading figure in Philosophy of Physics, emergence happens precisely there where singular limits cause our theories to break down. And as a matter of fact, our most important physical theories are asymptotically related in pairs:

[2] This "clamped-nucleus" assumption is part of the so called Born-Oppenheimer "approximation". Mathematically, it corresponds to an asymptotic series expansion in which the parameter ε (= electron mass/ nuclear mass) diverges to zero, that is, the nuclear mass is assumed to be infinitely large with respect to the electron mass.

$\mathrm{Lim}_{1/c \to 0}$ (special relativity) → Newtonian mechanics
$\mathrm{Lim}_{\lambda \to 0}$ (wave optics) → ray optics
$\mathrm{Lim}_{h \to 0}$ (quantum mechanics) → classical mechanics.

This can be interpreted as an indicator of the inadequacy of our theories and models, or instead as a "source of information". Batterman has been arguing for the latter attitude for several years now:

> If it were not for the singularities that appear in our theories and models we would have no understanding of the emergence at different scales of distinct and apparently "protected" states of matter (2011, p. 1040).

The "protected" states of matter that Batterman is referring to are what Laughlin and Pines (2000) call "protectorates", which are stable states of matter which are not only mathematically underivable from more fundamental equations without the help of singular limits, but are also insensitive to changes at the micro-level. These protectorates are the units of the phenomenon physicists call "universality", which is what philosophers dub "multiple realizability". One example of such a phenomenon is thermodynamic criticality.
The critical point of a fluid is a state in which liquid and vapor can coexist, and it is determined by a specific temperature and pressure (which is different from fluid to fluid).[3] Surprisingly, once they reach their specific critical point, all fluids (as well as magnets) behave in an identical manner, even if their properties are radically different in other phases and even if the values of their critical points are as diverse as 1,040.85ºC/270 atm for sulfur and -239.95ºC/12.8 atm for hydrogen. This macroscopic similarity beyond microscopic differences has been mathematically accounted for by the renormalization group theory (Batterman, 2002, 2011, 2014). This mathematical technique (for which Kenneth Wilson won the Nobel Prize) shows how the molecular details that are specific to each fluid are irrelevant for the macroscopic behavior that it shares with all other fluids.[4] Batterman's conclusion (2014, p. 15) is that this concrete method for explicating a process whereby higher order patterns arise that are not derivable from micro-structure can be considered as evidence against reductionism.
In the words of Laughlin, who uses examples such as these in his battle against the reductionistic framework often used in physics, it is obvious at the eyes of solid-state physicists, chemists and biologists that nature is filled with phenomena that are insensitive to microphysical variability, the behavior of which is determined by higher organizing principles, and that we may confidently call emergent.[5] This is one of the reasons why "predicting protein functionality or the behavior of the human brain from [quantum mechanical] equations is patently absurd" (Laughlin & Pines, 2000, p. 260).

[3] As with the previous cases, this phenomenon too is described as the result of assuming a variable to be infinite: viz. the number of particles or the correlation lengths between them (the distance over which one particle can influence another).

[4] The process, developed by Kadanoff, Fisher, and Wilson (cf. Batterman, 2002), is based on an iterated transformation of the Hamiltonian of each system, by which as one gradually changes scale, more and more fine-grained information is lost and the resulting function ends up being the same for all the elements of the universality class in question (a value that is called a "fixed point").

[5] Laughlin and Pines (2000) provide various examples. Here are two more: "The Josephson quantum is exact because of the principle of continuous symmetry breaking. The quantum Hall effect is exact because of localization. Neither of these things can be deduced from microscopics and both are transcendent, in that they would continue to be true and to lead to exact results even if the Theory of Everything were changed" (p. 261).

2.3. When Macro Context is as Crucial as Micro Structure

But are emergent phenomena exhausted by cases such as these, where we find radical discontinuities between theories? What can physics tell us regarding properties that do not look so mysterious to us?

Liquidity, for example, is a very familiar property (or cluster of properties) which we may resist considering emergent. In spite of its *in practice* unpredictability and novelty with respect to the properties of the components of the liquid taken in isolation (Weisskopf, 1977), the macroscopic properties of liquids (like viscosity or surface tension) and their causal powers seem straightforwardly derivable from the laws governing chemical bonds and other microscopic states and events. What we see at the level of the liquid is not something *over and above* the goings-on at the level of the molecules and their interactions. So reducibility seems possible, almost unavoidable. However, condensed-matter physicists reply, it is not that simple.

The stability of liquids depends on temperature, which *cannot* be derived from the particle's interactions. Even though in the philosophical literature temperature is still cited as a good example of reduction (it is taken to be *nothing but* the mean translational kinetic energy of molecules in a system), it is actually considered to be a case of emergence by most condensed-matter physicists. Temperature is a property that arises out of two mathematical transitions (from particle mechanics to statistical mechanics, and from there to thermodynamics), the calculation of which depends upon mathematical limits (e.g. the thermodynamic limit, which assumes the container of a gas to be infinitely large) as well as on stability conditions which are not available in the underlying domain, such as thermodynamic equilibrium.

Hence the calculation of the macro properties of liquids cannot be made without first establishing the stability conditions upon which the liquid depends, that is, without taking into account the macro conditions that make it so that some laws of interaction rather than others apply.

Molecular shape and criticality are considered to be good candidates for emergence because of the irreducibility of their macro description to the underlying quantum properties. This happens also in the case of liquids, since they cannot exist unless there is a certain sort of symmetry breaking induced by temperature, which in turn depends on conditions that can be provided only at the macro level. Therefore, if molecular shape and criticality are emergent, liquids should be considered to be so as well.

3. From Epistemology to Ontology in Physics

What the aforementioned examples show is that the reductionist ideal of macro properties being derivable from microscopic features and laws is not grounded in scientific practice. Scale transformations are highly problematic and many aspects of reality seem to simply pop up when a certain threshold of complexity is crossed, which mathematically corresponds to unphysical singular limits.

But is this not a merely epistemic matter? Even if we cannot predict upper-level phenomena on the basis of our lower-level knowledge, this does not imply that we are dealing with ontologically irreducible features.

As a matter of fact, besides the misunderstandings caused by the lack of agreement regarding the definition of emergence to which I alluded briefly in the introduction, another major source of confusion, especially in the scientific community, is the lack of clarity concerning the distinction between epistemological and ontological forms of emergence. Epistemological emergence is a relation existing between theories or models of the world. Ontological emergence is a relation existing between objects or properties in the world. Even though the latter implies the former, the inverse is not true. Epistemological emergence is no guarantee for ontological emergence. The impossibility of reducing a certain theory, with which we explain a certain upper-level domain, to a lower-level theory may be due only to our lack of knowledge of the details and intricacies of that lower level, its parts and the relations obtaining between them. Our theories may be incomplete.

If the epistemic irreducibility we found in the cases described in the previous section were to express a deeper ontological irreducibility, that would mean that there is a spontaneous and unexplained symmetry breaking at a certain point in the evolution of the system, whereby new properties with new causal powers come about. Isomers with different boiling points and densities, critical points in which new visible phenomena such as opalescence take place (the fluid becomes opaque and colored), temperature with different effects on macroscopic bodies (such as melting). If our epistemic limits express true ontological irreducibility, these examples, as well as many others, which might be more or less familiar and more or less complex, all seem to be cases of causally new and irreducible, hence emergent, macro features.

However, it is very hard to apply the epistemological/ontological distinction to physics. Physics does not have the pretense of *knowing* reality. All physics does is designing models that are quantitative, predictive and falsifiable. Whether those models correspond to the actual objective truth is something physics cannot tell us. Such an instrumentalist approach, which is the physicist's default standpoint, can make it hard on the philosopher to extract useful information from physical theory and practice for her metaphysical speculations. Of course, philosophers of a realist inclination would consider it legitimate to move from epistemological facts about physical theorizing to facts about the world. However, they would be moving alone. Physicists would hardly approve of such an extrapolation, which would hence be missing the safety net provided by the scientific method and the credibility that comes with intersubjective agreement.

But philosophers do not give up easily on what might be a fruitful dialogue. Even if physics only provides us with models of the world, one can still hypothesize a laplacean demon, a universal and omniscient calculator, whose complete and truthful knowledge of a certain system might be sufficient to explain all its macro properties. Could such a calculator predict the formation of a certain molecule in a fluid with such and such initial and boundary conditions? Granted, the conceivability of this omniscient being is traditionally used as an argument for universal determinism, not ontological reductionism, the truth of which it assumes from the start. Still, this theoretical exercise serves as a way to flesh out what can otherwise seem too abstract a hypothesis and allow physicists to more easily explicate their case against it.

Unfortunately, they will likely find the idea of a laplacean demon to be inapplicable to current physical science for several reasons. First, because in the world of particles at the subatomic scale, classical physics does not apply. Only once a certain measurement is made, is the system in a well determinate state; before, in general, it is considered to be in what is called a superposition of states. There is no way of knowing with absolute certainty all the information that fully characterizes a physical system: if we know precisely where a particle is located – its position in a certain spatial coordinate –, we will miss all knowledge about its momentum (in the same coordinate), and vice-versa.[6] The laplacean demon, therefore, cannot do his job in the quantum realm.

Second, because the *in practice* impossibility of calculating all the information contained in any macro system, not to mention the whole universe, is considered by physicists to be an *in principle* impossibility. It is presently established that no computer can ever accurately solve the equations describing the total energy of a system with more than ten particles at the quantum level (Laughlin & Pines, 2000, p. 160), because the interactions, which grow with the

[6] This is, of course, an extreme example of the renowned Heisenberg uncertainty relations.

factorial of N (number of particles), are intractable. So to imagine a universal calculation of the evolution of an ideal "system of the world" just sounds plainly absurd.

Third, because the theories we have that describe and explain macro properties on the basis of molecular properties are statistical in nature. They do not express the sort of one-to-one causality relations we would like a laplacean demon to have access to. This means that a really carefully imagined omniscient being would have to have a theory set that is fully coherent across scales, which is something we are nowhere near to achieving and cannot even know is possible.

A philosopher may tend to react to such arguments with dismay and call attention again to the hypothetical nature of the laplacean demon that need not suffer from the physical limitations of our brains, theories and actual computers, but the dialogue with the physicist will likely have come to a halt.

It seems like the philosopher and the physicist are talking past each other. The former aims at inferences about the ontology underlying our theories, which the latter will never be able to provide. *In principle* reductionism is impossible to prove and so is non-derivability.

4. Weakening the Reductionist's Case

Even if we ignored the distance between the epistemological and the ontological levels, we still could not use as arguments what science might be able to reveal in the future, but only what it is able to verify right now. And even that is extremely difficult to generalize.

Every microphysical law, which is an abstract construct formulated by theoreticians in as simple and context-free a way as possible, is tacitly implying that its application depends on the absence of outside influences (influences from upper levels of organization). What happens in a laboratory, then, is the testing of such abstract physical laws in equally aseptic environments, carefully designed to exclude any disturbing factor. However, outside these controlled setups, things get very messy. Even though the results of the experiments often corroborate the laws we wish to test, they cannot confirm their applicability to real-case scenarios where the boundaries between organization levels are loose and causal interactions between them much more likely (Dupré, 2001).

Hence, it would be fallacious to infer from the results of experimental scientific research such a strong metaphysical assumption as reductionism, which would require us to be able to ascertain with profound detail what happens in increasingly complex and ever changing contexts. Evidence for reductionism should consist in the verification that the behavior of all complex systems (from chemical, to biological, to neurological, to psychological, to social), in real-case situations, can be fully explained by microphysical laws, which is something that cannot even be done at a molecular level.

While one single case attesting to the failure of the *universal* claim of reductionism would suffice to falsify it, the same argument cannot be applied to emergentism, which is committed only to the existence of *some* irreducible phenomena. We would not need to survey the whole natural realm in order to prove it true; one relevant case of irreducibility would be enough.

In this sense, even if the epistemological emergence of many-body properties and the radical mathematical discontinuity between theories at different levels cannot prove the truth of ontological emergentism, they do come in handy as *circumstantial* evidence against the ontological reducibility of the macro to the micro.

In short, given the impossibility of using physics to prove their ontological claims, what both reductionists and emergentists must do is try to make the case for the higher implausibility of their rival position. That is much easier for a weaker, existential claim (emergentism) than for a strong, universal one (reductionism). And while the empirical evidence we have surveyed in section 2 cannot prove the truth of ontological emergence, it can shed strong doubts on its alternative and thus make the case for emergentism stronger.

5. Conclusion

Let us sum up. All the examples used in this paper consist in systemic properties that are *qualitatively* different from the properties of the parts. They can be calculated once we know the stability conditions that allow them to persist, but the singular limits that separate the theories that describe them render it impossible to explain the whole only on the basis of the parts. Nevertheless, the fact that we cannot know whether this epistemic irreducibility corresponds to an ontological gap rather than to mere limitations of our models prevents us from being able to assert conclusively whether these are cases of ontological emergence or not. In the end, the move from the epistemic to the ontological level of analysis is a matter of personal preference and intuition. Physics is silent about what is really *there* and so all it can do to help the emergentist's case is tell her that ontological emergence is not an absurd anti-scientific hypothesis. It is actually plausible, if our theories are true, since the way our models relate to each other is exactly what one should expect if ontological emergence were the case.
Brian MacLauglin has famously said:

> Given the advent of quantum mechanics and these other scientific theories, there seems not a scintilla of evidence that there are emergent causal powers or laws (1992, p. 23).

As we have seen, this statement is highly questionable. Despite the advent of quantum mechanics and other scientific theories that allow us to explain the behavior of the smallest portions of matter we know, there is much more than a few "scintillas of evidence" that there are emergent causal powers and laws in the world. And they might be much more common than usually supposed, even though reductionism cannot be disproven.

REFERENCES

Anderson, P.W. (1972). More is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science. In M.A. Bedau & P. Humphreys (Eds.) (2008). *Emergence. Contemporary Readings in Philosophy and Science*. Cambridge, MA: MIT Press, 221-229.

Arp, R. (2008). Emergence in Biology. *Cosmos and History: The Journal of Natural and Social Philosophy*, 4, 260-285.

Batterman, R.W. (2002). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. New York: Oxford University Press.

Batterman, R.W. (2011). Emergence, Singularities, and Symmetry Breaking. *Foundations of Physics*, 41, 1031-1050.

Batterman, R.W. (2014). Reduction and multiple realizability. www.robertbatterman.org.

Bedau, M. (1997). Weak Emergence. *Noûs*, 3, 375-399.

Bedau, M.A. & Humphreys, P. (2008). *Emergence. Contemporary Readings in Philosophy and Science*. Cambridge, MA: MIT Press.

Bishop, R.C. (2005). Patching physics and chemistry together. *Philosophy of Science*, 72, 710-722.

Bishop, R.C. (2009). Whence chemistry? Reductionism and neoreductionism. *Studies in History and Philosophy of Modern Physics*, 41, 171-177.

Bishop, R.C. & Atmanspacher, H. (2006). Contextual emergence in the description of properties. *Foundations of Physics*, 36, 1753-1777.

Dupré, H. (2001). *Human nature and the limits of science*. Oxford: Clarendon Press.

Kim, J. (1993). *Supervenience and Mind: Selected Essays*. Cambridge: Cambridge University Press.

Kim, J. (1999). Making Sense of Emergence. *Philosophical Studies*, 95, 3-36.

Laughlin, R.B. & Pines, D. (2000). The Theory of Everything. In M.A. Bedau & P. Humphreys (Eds.) (2008). *Emergence. Contemporary Readings in Philosophy and Science*. Cambridge, MA: MIT Press, 259-268.

McLaughlin, B. (1992). The Rise and Fall of British Emergentism. In M.A. Bedau & P. Humphreys

(Eds.) (2008). *Emergence. Contemporary Readings in Philosophy and Science*. Cambridge, MA: MIT Press, 19-59.
O'Connor, T. & Wong, H.Y. (2005). The Metaphysics of Emergence. *Nôus*, 39, 658-678.
Searle, J.R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Sober, E. (1999). The Multiple Realizability Argument against Reductionism. *Philosophy of Science*, 66, 542-564.
Weisskopf, V.F. (1977). About Liquids. *Transactions of the New York Academy of Sciences*, 38, 202-218.

ANDREA BLOMQVIST
University of Sheffield
ablomqvist1@sheffield.ac.uk

DIRECT SOCIAL PERCEPTION OF EMOTIONS IN CLOSE RELATIONS[1]

abstract

Drawing on a pluralist approach to mindreading, I explore Direct Social Perception with respect to perceiving the emotional states of people that we are close to, such as spouses, friends, and family. I argue that in general, emotions are embodied and can be perceived directly. I further claim that perceptual content includes concepts. That is, I argue against a non-conceptual view of emotion recognition, claiming instead that we learn emotional concepts by attending to certain expressive patterns of emotions. This view implicates that we can directly perceive both basic and non-basic emotions of people we are close to.

keywords

direct social perception, embodied emotions, mindreading, pattern recognition

Phenomenology and Mind, n. 12 - 2017, pp. 184-195
DOI: 10.13128/Phe_Mi-21117
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Direct Social Perception (DSP) is the idea that we can directly perceive others' mental states. Proponents of DSP claim that some mental states – usually motor intentions and emotions, but also sometimes beliefs (Gallagher and Hutto, 2008) – are embodied, and can therefore be directly accessed by others using perception (Gallagher, 2008; Gallagher, 2015; Krueger and Overgaard, 2012; Newen, Welpinghus, and Juckel, 2015; Spaulding, 2015; Zahavi, 2011). This paper will exclusively deal with emotions – in particular the non-basic emotions[2] of people that we know well.

DSP is often contrasted with inferentialist mindreading theories such as Theory Theory (TT) and Simulation Theory (ST).[3] Both of these theories claim that others' mental states are hidden from us; we need to use a theory or simulation to gain access to others' mental states, and we can only non-inferentially gain access to our own mental states. According to TT, we do this by using folk psychological rules such as "if A wants p and believes that doing q will bring about p, then ceteris paribus, A will do q" (Borg, 2007). When observing someone reaching for a bottle, I attribute the desire that they want the bottle to them, and the belief that reaching for it is the best way of achieving their goal. Similarly, if I observe someone smiling as they greet a friend, I may attribute happiness to them. Thus, by observing behavior, I infer that an agent has certain mental states. According to ST, we understand others' mental states by simulating their emotions or thoughts offline in ourselves and then projecting the simulated state to them (Goldman, 2006). That is, we put ourselves "in their shoes" to understand them.

In this paper, I will argue for a pluralist approach to mindreading.[4] That is, rather than claiming that the above theories are incompatible, I will claim that we use different strategies in different contexts depending on which is the least effortful to use (Fiebich and Coltheart,

[1] I am grateful to Will Hornett who on numerous occasions gave me insightful feedback on my manuscript as well as took time to discuss it with me – this significantly improved the paper. I am also grateful to my supervisor Luca Barlassina and the two anonymous referees who reviewed the manuscript and provided useful comments.

[2] I use 'non-basic' and 'complex' emotions interchangeably, where I take them to be any emotion that is not basic. Basic emotions include happiness, surprise, anger, sadness, fear, and disgust (Ekman, Friesen, and Ellsworth, 1972). Examples of non-basic/complex emotions are jealousy, sorrow, pride, etc.

[3] See for example Gopnik and Wellman (1992) for a defense of a version of TT, and Goldman (2006) for a defense of a version of ST.

[4] For pluralist accounts of mindreading, see Fiebich and Coltheart (2015); Gallagher (2015); and Fiebich, Gallagher, and Hutto (2016).

2015, p. 249). I will focus on the context[5] of interaction in close relations and argue that we use DSP in face-to-face interactions with people that we know well (e.g. spouses, friends, family).[6] My account will thus focus on interactions between people who are well acquainted, something which is not traditionally discussed by TT and ST (Fiebich and Coltheart, 2015, p. 236). There is a strong reason for the mindreading literature to focus on these cases, since we actually spend a significant amount of time mindreading people that we are close to, rather than strangers (think for example about how frequently spouses interact and how mindreading is important for the relationship to work smoothly). In these kinds of close relations, I will further claim, we can not only come to perceive the basic emotions, but we also perceive non-basic (sometimes called "dispositional") emotions, such as jealousy and pride.[7] I will argue that this is a process that is based on pattern recognition as suggested by Newen, Welpinghus, and Juckel (2015). However, I will depart from their view of the process as non-conceptual (Newen and Schlicht, 2009), and instead argue for a conceptual account based on Siegel's theory of rich perception and Wu's theory of attention (Siegel, 2006; Wu, 2008). This paper will thus develop an account of how we gain the ability to directly perceive non-basic emotional states of people we are close to.

In Section 1, I will outline the basic claims of DSP, considering how we can come to perceive emotional states by perceiving embodied emotions. Following Green, I will claim that we can perceive a mental state by perceiving part of it (Green, 2010). In Section 2, I will argue that perception is rich, and that we do not only perceive low-level properties like shapes and colors, but we can also perceive high-level states such as emotions, provided that we possess the right emotional concepts (Siegel, 2006; 2010). To explain how these concepts are acquired, I will draw on Wu's account of attention (Wu, 2008). In Section 3, I will consider Newen and Schlicht's argument that emotion recognition is based on non-conceptual pattern recognition (Newen and Schlicht, 2009). Here, I will be sympathetic to the idea that pattern recognition is necessary for emotion recognition, but argue that this process is conceptual. I will back this up by empirical data showing that people with dementia have impaired emotion recognition because they lack emotional concepts. Finally, in Section 4, I will show how rich perception, attention, and pattern recognition can enable us to directly perceive the non-basic emotions of people that we are close to.

1. Directly Perceiving Embodied Emotions

DSP theorists claim that mindreading is perceptual rather than cognitive and that we can directly perceive certain mental states of other people. DSP, they argue, is a non-inferential kind of mindreading that stands in contrast to the inferential mindreading proposed by theories such as TT and ST. Commonly, DSP theorists argue that we have direct access to the basic emotional states and motor intentions of other people (Gallagher, 2008; Gallagher, 2015; Krueger and Overgaard, 2012; Newen, Welpinghus, and Juckel, 2015; Spaulding, 2015; Zahavi, 2011). On these accounts, perceiving the emotional state of another person is on a par

[5] This can be seen as a response to Bohl (2015) who raises the worry for DSP theorists that they have not specified in which contexts DSP rather than other mindreading strategies are used.

[6] I wish to clarify two readings of this statement that I do not intend. The first reading is that DSP is *only* used in close relations, and hence not used to recognize the emotions of strangers. This is false, since we can directly perceive basic emotions of strangers as well. The second reading is that in close relations, we only use DSP, as opposed to any other mindreading strategies. Although I find this more plausible, it is likely that we sometimes have to resort to other strategies as well, for example when we are not able to directly perceive that a partner is hurt without them saying so. My main claim in this paper is that we most often use DSP, as opposed to other mindreading strategies, in close relations. Thanks to an anonymous referee for pushing me to clarify this.

[7] See Spaulding (2015) for an account of DSP whereby only occurrent basic emotions and not dispositional emotions are perceived.

with perceiving properties such as being a table or a chair (Siegel, 2006).[8] When perceiving the property of being a table, there is no extra step of inferring or judging that it is a table. Similarly, when I see a person smiling, I do not need to infer that they are happy. Rather, I see the happiness directly in their expression since the emotion is embodied in the expression.

It is important to note that DSP is not the same as behaviorism.[9] Whereas behaviorism claims that emotional states can be *reduced* to expressive states, DSP claims that *expressions partly constitute emotional states*.[10] Thus, happiness cannot be reduced simply to smiling, but rather smiling is a constitutive part of happiness.[11] Other parts include physiological responses (such as increased heart rate), phenomenal experience, cognitive features (such as attitudes and shifts of attention and perception), and an intentional object (Newen, Welpinghus, and Juckel, 2015: pp. 192-193).[12]

DSP does not need to be committed to the view that emotional expressions are necessarily partly constitutive of emotional states – one could be sad without crying. When this is the case, a pluralist theorist could claim that we cannot *directly* access the emotional state of the other person since it is not expressed, but it is still possible to access it via an inferentialist strategy. However, there are good reasons to think that these emotions in many cases are directly perceived, as there are characteristic expressions of embodied emotions that we readily recognize. There has been evidence showing that we can identify both emotional states and intentions from perceiving someone's posture or the way in which an action is executed.[13] Studies have shown that we can correctly categorize emotions when seeing point-light or full-light displays of both moving and static bodies (Atkinson *et al*., 2004; Heberlein *et al*., 2004). In a full-light display, an actor's body – but not facial expression – is visible to a subject, whereas in a point-light display, the actor wears luminescent straps on selected parts of their body (such as wrists, shoulders and knees) such that only the straps are visible as light points. In both these cases, subjects were able to tell what emotion the actor was expressing.

Further, there is a close relationship between the expression and the phenomenal experience of emotions, such that our emotional experiences are also affected by our expressions. This lends support to the idea of embodied emotions. For example, studies of people with Moebius syndrome support that facial expressions affect the phenomenology of emotions. Moebius syndrome is a congenital condition whereby a person's facial muscles are paralyzed and their eyes cannot move laterally. According to verbal accounts from people with Moebius syndrome, some feel like they "liv[e] in their head" and do not recognize themselves as *feeling* happy or *feeling* sad, but rather just as "think[ing] happy or think[ing] sad" (Cole, 1998). It seems as though they do not experience the phenomenology of happiness, and part of the reason for this could be that they are unable to express happiness by smiling. Turning to how emotional states are expressed by bodily actions, studies have also found that the intensity of the emotional experience is reduced when people are not able to express their emotion using body language. This can be seen in individuals who have suffered spinal cord injuries (Chwalisz *et al*., 1988; Hohmann, 1966; Laird, 2007; Mack *et al*., 2005). Studies have also

[8] Spaulding characterizes the directness with which we perceive emotions to be on a par with how we perceive objects, rather than properties. See Spaulding (2015, p. 3).

[9] See Krueger and Overgaard (2012) for an argument against DSP being a kind of behaviorism.

[10] Whether or not DSP needs expressions to be partly constitutive of emotions might be a controversial issue, though it is usually assumed and argued for in the literature (Krueger and Overgaard, 2012). Here, I will set this issue aside since it is outside the scope of this paper. Thanks to an anonymous reviewer for bringing this to my attention.

[11] I am here using 'constitutive part' in the same sense as Green (2010) and Krueger and Overgaard (2012).

[12] I will not argue for the inclusion of these features here. For discussion, see Newen, Welpinghus, and Juckel (2015).

[13] For work on decoding intentions from movement kinematics, see Ansuini *et al*. (2016); Cavallo *et al*. (2016).

manipulated facial expressions, posture, and gestures to produce a corresponding change in emotional phenomenology (Davis *et al.*, 2009; Laird, 2007; Niedenthal, 2007). These studies suggest that there is a causal link between the emotional expression, the physiology and the phenomenology of the emotion. Such a tight causal link is a good reason to think of these features as making up a distinct system and, as such, being constitutive parts of that system. We would readily take this to be the case when considering other systems, such as a computer. In this case, there are causal connections between the internal working parts, such that the activation of one will cause the activation of another, and we further take all the internal parts to be constitutive of the computer as a whole. Causality and partial constitution are thus not mutually exclusive here.

Let us now turn to the question of how we can perceive the emotion as an entity by perceiving one part of it. There is a worry that we cannot perceive others' emotions since we cannot perceive the physiological or cognitive states of another's emotion. A response to this remark is that we can perceive entities by perceiving characteristic parts of them (Green, 2010; Krueger and Overgaard, 2012). Expressions are characteristic parts of emotions, and perceiving a characteristic part of an entity is sufficient for perceiving that entity.[14] Bodily expression is part of an emotion, e.g. crying is part of being sad. By perceiving someone crying, we can thus perceive their emotional state of sadness.

There are two worries pulling in opposite directions that one might have with regards to perceiving emotional states by perceiving characteristic expressions. The first worry is that one might think that since emotions are embodied, we should be able to perceive any and all emotions in anyone. This does not seem to be the case, as only basic emotions are perceived cross-culturally (Ekman, Friesen, and Ellsworth, 1972). The second worry is that we should only be able to perceive certain emotional states, since not all emotional states have characteristic expressions. Both of these worries can be resolved by clarifying what 'characteristic' means. It is useful to draw a distinction between perceiving the emotions of strangers by perceiving *characteristic expressions of basic emotions*, and perceiving the non-basic emotions of people we are close to by perceiving those *particular people's characteristic expressions.* In the case of strangers, since we have little exposure to their expressions, we are only able to directly perceive basic emotions that are individuated in characteristic ways, e.g. smiling to express happiness. Empirical research has shown that we are able to recognize these expressions both within our culture and cross-culturally, lending support to the idea that these expressions are universal (Ekman, Friesen, and Ellsworth, 1972). 'Characteristic' here thus refers to what is generally (or universally) characteristic of an emotion for *any* given person. With a person we are close to, however, the expression need not be characteristic of how people *in general* express the emotion, but only of how *that person in particular* expresses the emotion (this has been dubbed *individual-typical expressions* – Glazer, forthcoming). If I know a person who always expresses surprise by frowning, I will come to recognize that, *for that person in particular*, frowning is a characteristic component of surprise. How we learn this will be clarified in Section 4. In the next section, I will argue that we also need a view of perceptual content as conceptually rich in order for mindreading to be a perceptual process.

2. Rich Perception

This section will focus on Siegel's account of rich perceptual content, and how we come to acquire the concepts necessary to perceive e.g. anger *as anger*, rather than just as a pattern of recognizable features. Rich perception is the idea that the content of perception is not limited

[14] I will not dwell on this point since it has been extensively discussed by Green (2010) and Krueger and Overgaard (2012). See these papers for more detailed arguments.

to representing low-level features such as shapes and colors, but is instead extended to high-level features which can include emotions and intentions. Siegel argues that, in perception, we also represent properties such as (but not limited to) 'being a natural kind'.[15] Thus, we perceive a tiger *as a tiger*, rather than as a bundle of features which are inferred be a tiger.

Siegel's argument runs from the phenomenology of perception to the representational content of the visual phenomenal experience. Suppose that you have never seen a pine tree before, and are hired to cut down all the pine trees in a forest containing trees of various kinds. It is pointed out to you what pine trees look like. After some time, your ability to tell pine trees from other trees improves, and pine trees gradually become more salient to you. Siegel claims that there is a phenomenological difference in the visual experience before and after you were able to successfully pick out pine trees. This phenomenological difference is due to a sensory difference between the two experiences, which is in turn due to the fact that the two experiences differ in content. That is, the content of the visual experience before learning what trees are pine trees is different to the content of the visual experience after having learnt which trees are pine trees.[16] The best explanation for this change in experience is that pine trees are represented *as the kind 'pine tree'* when we become familiar with them. Since all the low-level properties, such as the color and the shape of the trees, are the same in both experiences, low-level representational content is not enough to explain the phenomenal contrast before and after acquiring the concept PINE TREE. That is, improved pattern recognition of low-level features is not enough to account for the different ways in which we experience a pine tree before we are familiar with it and after we are familiar with it; the concept PINE TREE needs to be acquired for the phenomenological shift to occur. I will now turn to the issue of how we come to recognize the pattern of features that make up a pine tree, since this is important for acquiring the concept.

In my view, we come to acquire the concept partly *by means of* becoming better at picking out certain features, i.e. a pattern. In order to see how emotional concepts get into the content of perception, I draw on Wu's argument concerning the relationship between action and attention, and how to solve the Many-Many Problem (Wu, 2008; 2011). The Many-Many Problem is the problem of how we decide to perform a particular action (e.g. cutting down a pine tree), given that we are first faced with many perceptual inputs and need to decide which are relevant to our goal, and then faced with many possible actions that can be performed to reach the goal. By using intention to guide attention, we can solve this problem and focus on the relevant perceptual features of the world in order to act in the best way possible. An agent must first locate the attention to select a specific target, e.g. a glass that they wish to drink from. They must then parse the attention to focus on the specific properties of the object that are relevant to their goal. It is not relevant whether the glass has a particular color, but it is relevant whether it has a handle since that renders it more pick-up-able. Wu argues that concepts play a role in the attentional parsing of a glass. That is, lacking the concept GLASS, I would not be able to attend to the features relevant for my using the glass.[17] Acquiring the concept GLASS allows me to see the glass *as a glass*. There is a clear link here to Siegel's example of seeing a pine tree *as a pine tree* and its features then becoming more salient. When it is pointed out to me what a pine tree looks like, and I have the intention of cutting down all pine trees, my intention can guide my attention such that certain characteristic features of

[15] See Siegel (2006; 2010) for a full defense of this argument.

[16] See the literature on perceptual learning for empirical evidence supporting this thesis (e.g. Connolly, 2017).

[17] Wu allows for the possibility that we can still accidentally use the glass in the right way (Wu, 2008, p. 1019).

the pine tree become more salient to me.[18] My skill of recognizing pine trees thus improves as I become better at picking out the pine tree pattern, allowing me to acquire the concept PINE TREE.

In what follows, I will give an account of the pattern recognition we, as humans, develop. I will consider why we should take this to be a conceptual process rather than a non-conceptual one, by looking at an account that holds pattern recognition to be non-conceptual.

3. Against Non-Conceptual Recognition of Emotions

Newen and Schlicht developed a non-conceptual account of pattern recognition of emotions, and Newen, Welpinghus, and Juckel argued that we recognize emotions by recognizing individuated patterns of emotional states (Newen and Schlicht, 2009; Newen, Welpinghus, and Juckel, 2015). I agree with Newen, Welpinghus, and Juckel's characterization of pattern recognition of emotions, but I will argue that we *do* require the concept of e.g. anger to recognize the expressive pattern *as anger*.

According to Newen, Welpinghus, and Juckel, emotional states are individuated as patterns of characteristic features. This can most easily be seen in basic emotions which are said to be recognized cross-culturally (Ekman, Friesen, and Ellsworth, 1972). Plausibly, the reason why they are recognized cross-culturally is because they are all individuated as patterns of characteristic features (which may or may not be hardwired). In any given case of pattern recognition, Newen, Welpinghus, and Juckel claim that a feature F (such as a facial expression) is constitutive of a pattern P if it is part of at least one set of features which is minimally sufficient for a token pattern to belong to type T. The minimally sufficient conditions are then jointly sufficient for the emotional episode to be classified as e.g. the type anger. If one or more of these features were removed, the token emotional episode would no longer be classified as one of anger. It is plausible that we are able to develop this reliable way of recognizing emotional patterns because we are very often exposed to these particular emotional expressions. The mechanism thus has a large database of expressions to draw on in order to establish the characteristic patterns of basic emotions. In the final section, I will show how this feature of pattern recognition is also important for recognizing non-basic emotions in people we are close to.

Before doing so, let us discuss why Newen and Schlicht's claim that there can be a non-conceptual understanding of others' emotions is to be rejected (Newen and Schlicht, 2009, p. 232). This understanding is supposed to be underpinned by mirror neuron processes.[19] Mirror neurons are neurons that fire both when an action is executed by an agent, and when the agent sees another executing the same action. For emotions, research has shown that the same brain areas are activated when a person feels disgust, as when they see someone else looking disgusted (Wicker *et al.*, 2003). The thought is that we can non-conceptually perceive another person's emotional state using the mirror neuron system.[20] However, as argued above, perception is a conceptual process and the understanding of emotions cannot be explained only in terms of non-conceptual processes. To further strengthen this point, I will show that studies on subjects with dementia support that concepts are necessary both for using objects correctly and for recognizing emotions.[21] These similarities also support my application of Wu's account of attention to emotion recognition.

[18] Importantly, the first intention cannot involve the concept PINE TREE, since this has not been learnt yet. Instead, I suggest that the content of the intention would be something like 'to cut down all trees looking thusly', where 'thusly' refers to whatever features one's instructor pointed out.

[19] For work on mirror neurons, see Gallese *et al.*, 1996; Iacoboni, 2005.

[20] Mirror neurons have often been taken as providing evidence in favor of Simulation Theory (Goldman and Gallese, 1998; Goldman, 2006). Newen and Schlicht (2009) argue against this.

[21] Wu makes a similar argument to show that concepts are necessary for correct use of objects (Wu, 2008).

In order to possess concepts, semantic memory is needed to store these concepts. It is thus plausible to think that a person with impaired semantic memory would lose some of their concepts, and therefore not be able to use the objects corresponding to these concepts. This is indeed the case in semantic dementia, as demonstrated in a study by Hodges (2000). Subjects were tested on their ability to use everyday objects, as well as naming the objects and associating them with their use (e.g. a cork screw is used to open a bottle). To test whether the subjects could use the objects, the subjects' grasp, orientation and movement were tested. Failure to use the object in a correct manner correlated with semantic disabilities. In experiments testing subjects without dementia, it has also been shown that their grasping of objects is impaired when having to deploy their semantic memory in a distractor task (Creem and Proffitt, 2001). The claim that concepts are needed for emotion recognition can be supported in the same way. Studies have shown that impaired semantic memory also has an impact on emotion recognition. Subjects with dementia were tested on emotion recognition, and performed significantly worse than control groups without dementia (Keane *et al*., 2002; Lavenu *et al*., 1999; Rosen *et al*., 2004). This supports that the perceiving of emotions is indeed conceptual if we are to see an emotion *as a particular emotion*. That is, in order to see sadness *as sadness*, an agent needs to possess the concept SADNESS. I still grant that there could be a non-conceptual way of registering emotions, but it would be wrong to call this non-conceptual *understanding* as Newen and Schlicht do (Newen and Schlicht, 2009, p. 234). In order to understand the emotion of another person, we need to do more than simply recognizing a pattern which allows us to distinguish *this pattern* from *that pattern.* Being able to distinguish a mere pattern of sadness from a mere pattern of anger does not entail that an agent understands what distinguishes the *emotion sadness* from the *emotion anger*. I thus take it that the understanding of emotions using DSP is a conceptual process where concepts are the content of perception. Next, I will tie this together with how we develop a direct perception of complex emotions in close relations.

4. Directly Perceiving Complex Emotions in Close Relations

Finally, I wish to show that we can directly perceive complex emotional states of people we know well. In a way, we become experts at recognizing the individuated patterns of emotions in close relations.[22] Again, I will draw on Wu's account of attention to show how we can come to recognize these more complex patterns.

First, *prima facie* it seems like we can recognize a wider range of emotions in people that we are close to, such as spouses, family or friends.[23] I will focus the discussion around friendship, but my account extends to any close relationship. When seeing a stranger's smile, you might recognize the basic emotion it expresses (happiness), but you will not be able to distinguish that way of smiling from how the person smiles when proud; your pattern recognition is not fine-grained enough. This is because you are not familiar with that person's *particular expressions*, i.e. how *they in particular* smile when being happy as opposed to proud. You are able, however, to recognize particular ways in which a close friend smiles. You can easily distinguish how they smile from being bemused from how they smile from being happy or proud. That is, your skill at recognizing their particular expressions has improved. I will show how intention-guided attention plays a crucial role in this.

It could be thought that our pattern recognition improves only in virtue of *exposure* to the friend. This seems like a plausible suggestion since we tend to spend more time with friends

[22] That non-conceptual mindreading is used in family relations is suggested by Newen and Schlicht (2009) in a footnote, but there is no further development of this account.

[23] To my knowledge, there has not yet been any research on emotion recognition in close relationships, but there is research showing that we are better at recognizing emotions in cultures we are more frequently exposed to. This fits in well with my account. See Elfenbein and Ambady (2003) and Elfenbein *et al*. (2007).

than with other people. However, consider the following case. Samira works as a vet. In commuting to work, Samira is every day exposed to the same people on the tram. She is thus every day exposed to these people's expressions, and one might think that this is all that is needed for her to be able to develop direct perception of their complex emotions. Still, I think there is a good reason to think that her pattern recognition and ability to directly perceive the emotions of her fellow commuters are restricted in this case, and do not improve further. This is, at least in part, because Samira lacks the right intention in these cases.[24] As with the pine trees, an intention is needed to guide the attention to pick out the patterns relevant to a complex emotion. Normally, we do not have the intention to get to know our fellow commuters, and therefore our attention is not guided towards improving pattern recognition. However, if Samira were particularly interested in a person she saw on the commute every day, and had the intention of getting to know them, it is possible that she would come to be able to directly perceive some emotional states of that person. It is unlikely that she would come to be able to directly perceive many complex emotional states, however, since she is only exposed to the person in one context. If she were exposed to the person in more contexts – such as during a family dinner, when receiving some good news, when a friend makes an unfair remark, etc. – she would likely experience a wider range of that person's emotions. The range of emotions expressed by a person during a commute is likely to be smaller since the context is somehow monotonous.

In friendship, on the other hand, we are exposed to the friend in many contexts, and we do have the intention to e.g. get to know them better. It is important to note that the intention here is not as explicit as 'to learn to read the facial expressions better', but neither is the intention in Wu's example of picking up a cup 'to pick up the cup in the best way possible' – rather, the intention is just 'to be a good friend' or 'to drink from the cup'. Since part of what it means to be a good friend is to be able to respond to the other person's emotional states in an appropriate way, intending to be a good friend will guide attention in the right way for direct social perception to develop. Since we are exposed to a wide range of emotions in close relations, we can also come to directly perceive complex emotions such as jealousy. This follows directly from my claims about pattern recognition. Basic emotions are individuated by patterns that can be recognized cross-culturally, whereas jealousy might be expressed differently in different people, but crucially in *characteristic ways in particular people*. Having observed jealousy in a friend on multiple occasions, I can learn to recognize that pattern in the same way as I can recognize happiness in most strangers on the street. We can thus directly perceive the basic emotions of most people, but we additionally directly perceive complex emotions of people we are close to.

Finally, once we are able to perceive complex emotions of people we are close to, other mindreading strategies become redundant in face-to-face interaction. If I can directly perceive that my friend is proud, it is redundant to use another mindreading strategy that requires more steps. Since emotions are embodied, the information needed is already there in perception. Other mindreading strategies are still needed in close relations in other cases, such as if a friend texts me. Since I am not able to directly perceive their emotional state from the words I am reading, I will have to resort to another mindreading strategy.[25]

[24] The case is probably more complicated than as outlined here. The way in which we interact with people we are close to – e.g. how we confide in each other and how we openly avow feelings that we might otherwise keep to ourselves – could also contribute to the development of direct perception. My thanks to an anonymous reviewer for pointing this out.

[25] This account could be seen as an explanation for why we also *feel* closer to people that we know well. It seems plausible that actually having direct perception of someone's mental states might also affect your relationship with

5. Conclusion

In this paper, I provided an account of DSP, which hinges on the ideas of rich perception, attention, and pattern recognition. I first distinguished DSP from other accounts of mindreading, such as TT and ST, claiming that the main difference between these camps is that DSP is a non-inferential kind of mindreading. I then argued that emotional expressions partly constitute emotional states. For DSP to be a viable theory, we need to be able to perceive emotional states by seeing emotional expressions. I supported this by showing that emotional expressions affect the phenomenology of emotions. For example, people with Moebius syndrome or people who are paralyzed and therefore not able to bodily express their emotions, experience emotions differently. Expressions should thus be considered as partly constitutive of emotional states. Since we can perceive an entity by perceiving a part of it, I suggested that, by perceiving part of an emotion (i.e. its expression), we can perceive the emotion. I then considered the nature of perception, and how emotions can be part of perceptual states. Siegel argues that concepts feature in our perception, and that there is a contrast between experiencing a pine tree before having acquired the concept PINE TREE and experiencing a pine tree after having acquired the concept. Certain features of the pine tree become salient to us when we learn the concept, and we become able to recognize the pine tree *as a pine tree.* Similarly, when studying someone's face, we learn to recognize the patterns that individuate particular emotions. It is necessary both that we learn this pattern and that we possess the concept of a particular emotion, such as anger, to recognize it *as anger*. Drawing on studies which show that people with dementia are not able to recognize emotions for which they lack a concept, I maintained that the perception of emotions is a conceptual process, contra Newen and Schlicht. Finally, I argued that DSP is particularly important in close relations. Adopting Wu's theory of attention, I showed how we can come to recognize complex emotions of people that we are close to by intention-guided attention. In virtue of intending to get to know a person better, certain features become salient to us such that we are able to recognize these as patterns of complex emotions. Theoretical considerations suggest that DSP is particularly important for close relationships, but future empirical research is also needed to validate this claim.

REFERENCES

Ansuini, C., Cavallo, A., Campus, C., Quarona, D., Koul, A., & Becchio, C. (2016). Are we real when we fake? Attunement to object weight in natural and pantomimed grasping movements. *Frontiers in Human Neuroscience*, 10(471).

Atkinson, A., Dittrich, W., Gammell, A, and Young, A. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33, 717-746.

Bohl, V. (2015). Continuing debates on direct social perception: Some notes on Gallagher's analysis of "the new hybrids". *Consciousness and Cognition*, 36, 466-471.

Borg, E. (2007). If mirror neurons are the answer, what was the question?. *Journal of Consciousness Studies*, 14, 5-19;

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific reports*, 6(37036), 1-8.

Chwalisz, K., Diener, E., & Gallagher, D. (1988). Automatic arousal feedback and emotional experience: evidence from spinal cord injured. *Journal of Personality and Social Psychology*, 54(5), 820-828.

Cole, J. (2010). Agency with impairments of movement. In D. Schmicking and S. Gallagher (Eds.), *Handbook of Phenomenology and Cognitive Science.* Dordrecht: Springer, 655-670.

them such that you feel like you have a closer connection to them. This would need to be further developed in another paper.

Colombetti, G., & Roberts, T. (2015). Extending the extended mind: the case for extended affectivity. *Philosophical Studies*, 172(5), 1243-1263.

Connolly, K. (2017). Perceptual Learning. *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (Ed.). https://plato.stanford.edu/archives/sum2017/entries/perceptual-learning.

Creem, S., Proffitt, D. (2001). Grasping objects by their handles: A necessary interaction between cognition and action. *Journal of Experimental Philosophy: Human Perception and Performance,* 27, 218-228.

Davis, J., Senghas, A., & Ochsner, K. (2009). How does facial feedback modulate emotional experience?. *Journal of Research in Personality*, 43(5), 822-829.

Ekman, P., Friesen, W.V., & Ellsworth, P. (1972). *Emotions in the Human Face*. Oxford: Pergamon Press.

Elfenbein, H.A., & Ambady, N. (2003). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *Journal of Personality and Social Psychology*, 85, 276-290.

Elfenbein, H.A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, 7(1), 131-146.

Fiebich, A., & Coltheart, M. (2015). Various ways to understand other minds: Towards a pluralistic approach to the explanation of social understanding*. Mind and Language*, 30(3), 235-258.

Fiebich, A., Gallagher, S., & Hutto, D. (2016). Pluralism, interaction and the ontogeny of social cognition. In J. Kiverstein (Ed.), *The Routledge Handbook Philosophy of the Social Mind*. London: Routledge, 208-221.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17(2), 535-543.

Gallagher, S. (2015). The new hybrids: Continuing debates on social perception. *Consciousness and Cognition,* 36, 452-465.

Gallagher, S., & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. Racine, C. Sinha & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: John Benjamins Publishing Company, 17-38.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593-609.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493-501.

Glazer, T. (forthcoming). Looking angry and sounding sad: The perceptual analysis of emotional expression. *Synthese*, 1-25.

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading.* New York: Oxford University Press.

Gopnik, A., & Wellman, H.M. (1992). Why the child's theory of mind really is a theory. *Mind and Language*, 7 (1-2), 145-171.

Green, M. (2010). Perceiving Emotions. *Aristotelian Society Supplementary,* 84(1), 45-61.

Heberlein, A., Adolphs, R., Tranel, D., & Damasio, H. (2004). Cortical regions for judgements of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience*, 16(7), 1143-1158.

Hodges, J., Bozeat, S., Lambon Ralph, M., Patterson, K., Spatt, J. (2000). The role of conceptual knowledge in object use: Evidence from Semantic Dementia. *Brain,* 123, 1913-1925.

Hohmann, G.W. (1966). Some effects of spinal cord lesion on experienced emotional feelings. *Psychophysiology*, 3(2), 143-156.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLOS Biology*, 3(3), 529-535.

Keane, J., Calder, A., J., Hodges, J. R., & Young, A.W. (2002). Face and emotion processing in frontal variant frontotemporal dementia. *Neuropsychologia*, 40(6), 655-665.

Krueger, J., Overgaard, S. (2012). Seeing subjectivity: defending a perceptual account of other minds. *ProtoSociology*, 47, 239-262.

Laird, J.D. (2007). *Feelings: The perception of self.* Oxford: Oxford University Press.

Lavenu, I., Pasquier, F. (2005). Perception of emotion on faces in frontotemporal Dementia and Alzheimer's Disease: A longitudinal study. *Dementia and Geriatric Cognitive Disorders*, 19, 37-41.

Mack, H., Birbaumer, N., Kaps, H., Badke, A., & Kaiser, J. (2005). Motion and emotion: Emotion processing in quadriplegic patients and athletes. *Zeitschrift Fur Medizinische Psychologie*, 14(4), 159-166.

Newen, A., & Schlicht, T. (2009). Understanding other minds: A criticism of Goldman's simulation theory and an outline of the person model theory. *Grazer Philosophische Studien*, 79(1), 209-242.

Newen, A., Welpinghus, A., & Juckel, G. (2015). Emotion recognition as pattern recognition: The relevance of perception. *Mind and Language*, 30(2), 187-208.

Niedenthal, P. (2007). Embodying emotion. *Science*, 316(5827), 1002-1005.

Rosen, H., Pace-Savitsky, K., Perry, R., Kramer, J., Miller, B., & Levenson, R. (2004). Recognition of emotion in the frontal and temporal variants of frontotemporal Dementia. *Dementia and Geriatric Cognitive Disorders*, 17, 277-281.

Siegel, S. (2006). Which properties are represented in perception? In T. S. Gendler and J. Hawthorne (Eds.), *Perceptual Experience*. New York: Oxford University Press, 418-503.

Siegel, S. (2010). *The Contents of Visual Experience*. New York: Oxford University Press.

Spaulding, S. (2015). On Direct Social Perception. *Consciousness and Cognition*, 36, 472-482.

Wicker, B., Keysers, K., Plailly, J., Royet, P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, 40(3), 655-666.

Wieser, M., & Brosch, T. (2012). Faces in context: a review and systematization of contextual influences on affective face processing. *Frontiers in Psychology*, 3(471).

Wu, W. (2008). Visual attention, conceptual content, and doing it right. *Mind*, 117(468), 1003-1033.

Wu, W. (2011). Confronting the Many-Many Problem: Attention and agentive control. *Noûs*, 45(1), 50-76.

Zahavi, D. (2011). Empathy and Direct Social Perception: A phenomenological proposal. *Review of Philosophy and Psychology*, 2(3), 541-558.

NICCOLÒ NEGRO
University of Milan
niccolo.negro@studenti.unimi.it

ME, YOU, AND THE MEASUREMENT. FOUNDING A SCIENCE OF CONSCIOUSNESS ON THE SECOND PERSON PERSPECTIVE

abstract

Modern science was born when physicists started studying phenomena by recruiting mathematical explanatory frameworks. Since this appears to be the direction followed in recent studies on consciousness, philosophers have to analyze the justification of this third-person methods of explaining a phenomenon that is supposed to be entirely subjective. In this paper I argue that this kind of justification could be found in a certain interpretation of the second-person perspective and I briefly sketch how one of the most promising contemporary theory of consciousness (IIT) could fit with such an interpretation.

keywords

second-person perspective, Integrated Information Theory, science of consciousness, embodied simulation theory

Phenomenology and Mind, n. 12 - 2017, pp. 196-206
DOI: 10.13128/Phe_Mi-21118
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

ME, YOU, AND THE MEASUREMENT

Introduction

More than forty years have passed since David H. Hubel (1974) identified the need for a Copernicus in neurobiology. Copernicus rethought the Earth's position within the Universe and, in doing so, he forced the entire intellectual community to revise the philosophical assumptions that used to characterize the scientific inquiry of the time.

More recently, Giulio Tononi (2003) referred to Galileo Galilei in his attempt to outline a scientific theory of consciousness. In fact, Galileo was the one who fathered the modern conception of science, by employing mathematical tools in order to arrange empirical observations in a predictive and quantitative framework. Thanks to Copernicus, we attributed a new position in the Universe to human being; thanks to Galileo, we discovered how to deal with the world around us.

In this article, I will argue that, in order to justify the claim that consciousness can be investigated scientifically, it's important to get rid of out-dated assumptions about subjectivity and in particular of the idea that subjective experience is entirely evident only to the subject itself. Contrary to this view, subjective experience seems to be nothing but a form of our being living creatures in a biological and social environment, and cannot be understood apart from it. After that, I will argue that a Science of Consciousness (SOC) requires a measurement and a mathematical vocabulary: Integrated Information Theory (IIT) can be a promising starting point from which a quantitative framework, in which we are supposed to set our data, can flourish.

The main idea informing this paper is that in addition to the first-person perspective and the third-person one, another one is needed: something like a second-person perspective. In other words, before invoking Galileo, we need a Copernican revolution in consciousness studies, which would allow us to rethink the current conception of subjectivity.

In §1 I argue against the idea of a first-person SOC, in §2 and §3 I claim that there must be a "bridge" between first-person perspective (1PP) and third-person perspective (3PP), and that such a bridge can be built upon the very basis of human social interactions, namely the second-person perspective (2PP). In conclusion, in §4 I explore the possibilities that IIT opens for a SOC, by considering its future prospects and problems.

1. First Person Data and the Science of Consciousness

This section addresses the assumption that subjective experience is a private phenomenon, directly accessible only to the bearer of the experience. This line of thought has been strongly defended by David Chalmers, who maintains that the task for a SOC is to integrate two different classes of data into a scientific scheme:

> As anyone who has listened to music knows, there is also a distinctive quality of subjective experience associated with listening to music. A science of music that explained the various third-person data just listed but that did not explain the first-person data of musical experience would be a seriously incomplete science of music. A complete science of musical experience must explain both sorts of phenomena, preferably within an integrated framework (Chalmers, 2004, p. 1112).

According to Chalmers, the 1PP data are irreducible to the 3PP data. Consciousness is composed of properties of the experience that are by their own nature intrinsically qualitative and private: they are not measurable with a quantitative analysis from an external point of view. Thus, the existence of such properties, famously called *qualia*, seems to rule out, *a priori*, the possibility of a traditionally characterized SOC. However, rather than engaging in this metaphysical debate, this section seeks to demonstrate how certain metaphysical assumption can in fact damage a scientific enterprise.

In Chalmers' view, phenomenal concepts like 'pain' are not reducible to functional analysis – for every functioning that could be found at the neuronal level, it can be asked why such a functioning should be associated with a qualitative feeling. In other words, according to Chalmers, traditional science adopts a reductive form of explanation in terms of functions and dispositions, but these instruments will never be enough to explain phenomenal consciousness. Therefore, a SOC must take a non-reductive form, which considers first-person data as fundamental.

I am not claiming that Chalmers denies the feasibility of a SOC. Rather, my aim is to reformulate his proposal, with special emphasis on the idea that first-person data are *irreducible* to third-person data:

> A science of consciousness will not reduce first-person data to third-person data, but it will articulate the systematic connections between them. Where there is systematic covariation between two classes of data, we can expect systematic principles to underlie and explain the covariation. In the case of consciousness, we can expect systematic bridging principles to underlie and explain the covariation between third-person data and first-person data. A theory of consciousness will ultimately be a theory of these principles (Chalmers, 2004, p. 1113).

Chalmers' assumptions lead him to the idea that, if we want to construct a SOC, we need to *correlate* 1PP data with 3PP data, and for this reason he seems particularly interested in the so-called *Neural Correlates of Consciousness* research (Chalmers, 1996; 1998; 2000).

The first problem with this view is that it requires a non-standard concept of science. Science is typically taken to look for explanations rather than correlations. The scientific method, since Galileo, requires framing observations in a mathematically characterized explanatory model. Disregarding this (historically successful) model because of a metaphysical assumption about the nature of subjectivity could end up being a risky move.

The second problem concerns the idea of considering first-person data as fundamental. As a matter of fact, subjective experience is what we want to *be explained* by a SOC, it is not what *explains*. Introducing first-person data into a SOC would bring to the *explanans* what should

stay on the *explanandum*'s level. This remark helps clarify what should be the real *explanandum* for a SOC. As far as science deals with general phenomena, and not with individual events, we are not looking for an explanation of *my* particular subjective experience. Instead, we are concerned with the possibility that a subjective point of view can exist in the first instance.
A successful SOC will provide us with general laws explaining how a subjective and intrinsic point of view arises from an organism biologically and physically constituted like us. Furthermore, it must explain how it is possible that my experiences always seem to have a quality that differs from yours. In other words, we need a SOC that explains consciousness from the intrinsic perspective of an organism. This ambitious project can be carried out only if our metaphysical presuppositions do not impede the empirical work.

In the following sections, I will explore the hypothesis that our experiential states are, by their quality, open to others' experiences. Not that you feel *my* pain, but you can mirror my pain directly in your body. You can experience it without feeling it.

Again, it seems that the notion of subjectivity, with which we attempt to construct a SOC, must be rooted in the intersubjective realm and unfolded towards the social environment a living being develops in. Rather than advocating for a scientific revolution, I will argue, here, that we need to rethink the conception that locks experience in the private realm of the subject.

2. In Need of a Bridge

The need to build a bridge between the 1PP and the 3PP is not a novel claim. The hypothesis I am proposing is, at least for the basic intuitions underpinning it, akin to the approach developed by Daniel Dennett and called *heterophenomenology* (Dennett, 2003; 2007):

> Let me begin, then, with something of a bird's-eye view of what I take heterophenomenology to be: a bridge – the bridge – between the subjectivity of human consciousness and the natural sciences (Dennett, 2007, p. 249).

Dennett does not deny that there is a contrast between the point of view of the subject and the point of view of the observer. Rather, he simply posits that there are no *a priori* reasons for not conveying the 1PP into the domain of natural science, as traditionally characterized. Investigating the value of heterophenomenology as a methodology for contemporary cognitive science is far beyond the scope of this paper, and I shall set aside the discussion about the different methodologies which have been explored to study subjective experience.[1]
It is worth noting here, though, that such a conveyance (from 1PP to 3PP) is not ruled out *a priori*. Besides, its *possibility* is strongly suggested by an empirical hypothesis that I will outline in the next section.

Looking more closely at heterophenomenology, it can be said that this method aims to single out the structure of the experience by using interviews and self-reports from the subject, in order to anchor those subjective experiences in *something* that can be detected and confirmed in replicable experiments. We arrive at this *something*, according to Dennett, through the equipment of the natural sciences. In this respect, heterophenomenology counts as a third-person methodology:

> So heterophenomenology could just as well have been called – by me – first-person science of consciousness or the second-person method of gathering data. I chose instead to stress its continuity with the objective standards of the natural sciences:

[1] Related to this issue, it is worth noting the role of neurophenomenology and the second-person techniques that this school of thought is carrying out. See Varela (1996), Thompson (2007), Olivares *et al*. (2015).

> "intersubjectively available contents which can be investigated as to truth and falsity" as Alva Noë puts it (Dennett, 2007, p. 252).

It is worth emphasizing that, according to the view presented in this paper, the 2PP is more than a method for gathering data. In the construction of a SOC, we do not observe that a subject *S* is in pain because she is reporting that. Rather, observers can agree upon the qualitative aspect of *S*'s experience because that experience is open to an experiential realm, which is shared by interacting subjects. As far as we are concerned with the construction of a SOC, this conception paves the way for a scientific understanding of the subjective experience. If subjectivity is spread in an intersubjective field, as a consequence, it becomes possible to study it by using intersubjective methods. Natural sciences are just one option, as it can be seen as a high-level social interaction whose contents are intersubjectively available.
I shall highlight that the approach I present is complementary to the heterophenomenological method. The view presented here, unlike Dennett's, underlines the *nature* of the data of a SOC, namely that such data are essentially disposed to be studied intersubjectively. The bridge between the 1PP and the 3PP that the heterophenomenological approach aims to build cannot be understood without stressing this aspect.
Before anchoring the subjective experience in a 3PP characterized by a mathematical framework, we need to go one step further and assess the questions as to (i) whether the idea of founding our SOC on the 2PP is scientifically justified or not and (ii) whether this 2PP will be necessary to study consciousness, once such third-person method is achieved.

3. Second Person Perspective as the Basis of Social Interactions

As Michael Pauen (2012) underlines, the 2PP is a perspective on a perspective: by observing something from a 3PP, we observe something that can be publically accessed, but this is not the case for the 2PP observation. The object I am analyzing from the 2PP has the same status of the observing subject:

> For the same reason, this perspective is not merely subjective [...] that is, it is not about the epistemic subject's own thoughts, feelings, and desires. Again, it's about other people's thoughts, feelings, and desires. This distinguishes the second from the first person perspective. Given that the second person perspective is a relation between the epistemic subject and one or more other subjects, it seems most appropriate to describe it as "intersubjective". This means that it is not primarily a scientific perspective. First and foremost it's a social perspective – even if it plays an important role in science and the humanities, too (Pauen, 2012, p. 22).

How can such an intersubjective relation be useful for the purpose of a SOC? As Pauen points out – quoting Loar (1997) and Papineau (1998) - the very possibility to employ phenomenal concepts such as 'pain' is enabled by the 2PP. According to this view, these concepts refer to an experience of the subject that can be simulated by other subjects of the linguistic community. There are two notions which are worthy of further enquiry: reference and simulation. Without delving into the wide field of the theory of reference (and meaning), I will adopt here a kind of Wittgensteinian point of view according to which the meaning of a linguistic item is given by the role, shared by a community of speakers, that that item plays in a certain linguistic game. When you use the word 'pain', I know what *you* mean because of *my* experiences of feeling pain. But this does not entail that each subject refers to her own private concept of pain. The meaning itself is a matter of intersubjective appraisal and validation. However, we still need an explanation of how a phenomenal concept can be intersubjectively shared, even though it is used by appealing to *my* own experiences, rather than *ours.* To see this, we need to address the problem posed by the notion of simulation.

Recent neuroscientific studies (Singer *et al.*, 2004; Ryu *et al.*, 2008) have discovered that some brain areas are activated both when a subject imagines an emotion or a perception, and when she actually *feels* that emotion or *has* that perception. Pauen seems to take these findings as proofs for the fact that an actual simulation is going on when I am listening to you telling me about your pain, as though my imagination process brings about an empathic relation. Such interpretation does not seem to be fully satisfactory: imagination is quite an active and conscious process, and it seems inappropriate to describe what is going on in the comprehension of concepts like 'pain'. The notion of simulation that is needed here is more primitive and sub-personal. Thus, the simulation process does not need to be consciously recruited by the subject, but, rather, it must be something that happens to a subject embedded in a social environment.

The main point of this section is that the notion of simulation that can contribute constructing a SOC could be derived from the Embodied Simulation Theory (EST) as proposed by Vittorio Gallese (Gallese, 2005, 2016; Gallese & Sinigaglia, 2011). According to this view, the producer of representational content is not the brain *per se*, but rather the brain-body system. Thus, intersubjectivity is nothing but a form of intercorporeality. Simulation is not a "resemblance process" that reproduces or copies a mental state. Rather, it is a functional process depending on the possibility to reuse certain brain structures for different purposes:

> According to the alternative view, simulation as reuse, there is mental simulation just in case the same mental state or process that is used for one purpose is reused for another purpose (Gallese, 2016, p. 303).

Such a proposal goes beyond the role of mirror neurons[2] and deals with the possibility for a subject to simulate another subject's mental state. Such a simulation derives from the fact that the same brain structure is activated both when a subject experiences a certain state and when she sees someone else experiencing that very same state. As Gallese puts it:

> ES theory posits that the MM [Mirror Mechanism] counts as implementing mental simulation processes primarily because brain and cognitive resources typically used for one purpose are reused for another purpose. For instance, the activation of parieto-premotor cortical networks, which typically serve the purpose of representing and accomplishing a single motor outcome (such as grasping something), [...] might also serve the purpose of attributing the same motor goal or motor intention to others. The same holds for emotions and sensations. Within the anterior insula the same voxels typically underpinning the subjective experience of disgust also activate when attributing disgust to others (Gallese, 2016, p. 304).

What is crucial is that our understanding of others is represented via the bodily format: we can map the others onto ourselves in a non meta-representational, sub-personal and pre-intentional way. By collecting suggestions from Pauen's and Gallese's works, we can extend EST to phenomenal concepts and speculate that at least certain kinds of concepts are in fact embodied. For example, in Broca's area there is a wide overlap between structures activated both in speech production and in speech perception. Moreover, Mirror Mechanism is activated when certain actions are described verbally (Gallese & Glenberg, 2011). It can be suggested,

[2] For an introduction, see Rizzolatti & Sinigaglia (2008).

at this point, that concepts like ‘pain’ can also be mapped in our body, because the same brain regions that are responsible for the quality of *my* feeling of pain are responsible for my comprehension of *someone else’s* pain in several intersubjective practices.

If this hypothesis were correct (and we need further empirical evidence regarding the influence of language on the Mirror Mechanism, as well as the role of the body in language processing), it would suggest that the bodily format of concept representation is the mechanism underpinning that intersubjective consonance that, in the present proposal, makes the bridge between the 1PP and the 3PP. Of course, my experience remains in a certain sense private because it is *my* body interacting with an environment. It is *my* point of view. However, others can, in principle, map this experience onto their selves, and this mapping seems to be the enabling condition for the intersubjective formation of phenomenal concepts we use in intersubjective practices. Since science is one of such contexts, our SOC will have to deal with concepts that are both “private” from an intrinsic point of view and “public” from the point of view of a social interaction between subjects. Once again, the final goal of a SOC is to account for the possibility of a point of view to arise, given a certain physical substrate and a dynamic interaction of this substrate with a biological and social environment. What I suggest is that science must rely on the degree of public availability of experiences happening to this point of view. As further evidence, I shall consider the work of the developmental psychologist Vasudevi Reddy.[3] Importantly, Reddy (2003) shows that early infants interpret others as attending beings during reciprocal interactions, even before they develop awareness of themselves and others as psychological entities.

Reddy’s suggestion shows that the mutual engagement I-Other is perceived, rather than conceived. Felt, rather than inferred:

> A second-person approach to self-and-other awareness is suggested as an embodied bridge across the alleged gap between first-person experience and third-person observation (Reddy, 2003, p. 401).

Evidence from developmental psychology appears to confirm that the development of our subjectivity derives from intersubjective practices. The idea is that 2PP is the perspective from which we cognitively form, in a later time, a conception of the Self. In conclusion, it can be argued that the empirical findings discussed in this section strongly suggest that the metaphysical conception of subjectivity as a private realm does not seem to be adequately supported.

Before proposing a 3PP method to study consciousness, it is worth noting that 2PP needs to be transcended. Indeed, 2PP is, in the present proposal, a conceptual tool that allows us to grasp the intersubjective nature of subjectivity and to open the possibility of studying consciousness scientifically. Once one acknowledges that, the possibility to draw a complete theory of consciousness based exclusively on the 3PP is not excluded and this is in fact the result we should expect from a mature SOC.

[3] I owe this suggestion to an anonymous referee, whom I sincerely thank.

4. How Can We Measure Experience?

The last step requires bringing subjective experience into a structure defined by its functional properties,[4] so that a particular aspect of an experience can be defined by its power to affect the state of the organism. Once defined in such a way, it will be easier to set phenomenal concepts in a mathematical framework that will provide us with measures and predictions. The crucial point is that this functional definition is possible because of the intersubjective nature of the phenomenal concepts discussed above: 'pain' is a concept that a community of speakers employs not because of what it *is*, but of what it *does*. We use that concept because of how the phenomenon picked up by the concept affects our *bodies.* Generally speaking, phenomenal concepts deal with bodily states, not mental states. For a SOC, this aforementioned "*does*" has to be treated mathematically. One could object and argue that a numerical-mathematical description of consciousness would provide us with a method for individuating conscious experience in a way that is independent from the observer's point of view. This would be, in itself, an objection to the idea that experience is intersubjectively constituted. A possible answer to this remark consists in clarifying that maintaining that phenomenal experience is intersubjective does not mean that there is no experience without the second-person consonance. In the case of locked-in patients, there can be subjective experience without the ability to show that that experience is actually there within an intersubjective domain. Such evidence does not show that subjective experience is essentially private, but, rather, that there is the possibility for an emergence of a subjective point of view, even in such cases where the mechanisms underlying the *expressions* of that experience are impaired. This shows once again that the 2PP is not enough for constructing a complete theory of consciousness. We need to shift to the 3PP in order to study consciousness both as a dynamic phenomenon of an organism interacting with an environment and as a brain's capacity in off-line situations.

In this section, I will suggest that IIT is a convincing method and a well-informed epistemological theory of measuring experience. IIT is chosen among several other alternatives because the practical measures that are based on IIT have shown empirical effectiveness.[5] IIT[6] poses, as I remarked several times in this paper, that consciousness exists from a subjective point of view and claims that phenomenal properties of experience can be explained by the informational properties of a system.

Anyways, information here is not defined in the classical Shannonian sense, according to which it would be the measure of the entropy of a system, but as "difference that makes a difference"[7] in a system. In IIT, information about the cause/effect repertoire of a basic mechanism is measured by recruiting the *minimum* function between cause information (CI) and effect information (EI) repertoires. Roughly speaking, a mechanism in a state specifies a set of possible previous states (CI) and a set of possible outcome states (EI). Appealing to minimum function guarantees that that current state is available from the intrinsic point of view of the system, since it works as an "information bottleneck". Whenever a system cannot be partitioned without a loss of information, it exists from an intrinsic point of view, according to the key notion of integration, the second pillar of the theory.

This analysis aims to explain the emergence of an intrinsic point of view of a system from the basic elements constituting it. Nevertheless, we still need an explanation of how to measure experience. IIT defines concepts as Maximally Irreducible Cause/Effect Repertoire (MICE),

[4] For 'functional property' I mean a property defined by its function, namely the capability to affect and be affected by other properties in its own network.

[5] The most important method is the so-called perturbational approach developed by Massimini *et al.* (2009).

[6] See Tononi (2003; 2008; 2012), Tononi, Oizumi, & Albantakis (2014), Tononi & Massimini (2013).

[7] The expression, used in Tononi (2008) and Tononi & Massimini (2013), derives from Bateson (1972).

namely the cause-effect repertoire that generates a maximum of integrated information among different subsystems. This C/E repertoire is maximally integrated inasmuch as, if it is segregated, it will bring the maximum loss of information within the system. As Tononi puts it:

> an element (or set of elements) implementing the concept "table", when ON, specifies 'backward' the maximally irreducible set of inputs that could have caused its turning ON (e.g. seeing, touching, imagining a table); 'forward', it specifies the set of outputs that would be the effects of its turning ON (e.g. thinking of sitting at, writing over, pounding on a table) (Tononi, 2012, p. 302).

A particular concept, according to IIT, can be set in a structure that defines a temporal fragment of our experience, forming a Maximally Irreducible Conceptual Structure (MICS), the *locus* which specifies the maximum of *conceptual* integrated information. The MICS, in IIT, is a *quale*, the minimum building block of our phenomenal consciousness.
With this theoretical apparatus, IIT claims to be able to measure φ, namely the *quantity* of an experience, how much such an experience is present to the subject, while the shape of the concept constellation in the "qualia space" would specify the quality of the experience. The problem, here, is to understand how this qualia space is characterized. It is defined as a multidimensional space whose axes are constituted by each possible state of the system. Along each axis, the probability of each state is distributed, constituting a point in the space. As a consequence, vectors connecting different points in the space realize the informational relationship between possible states. The geometrical shape of these connected points in the multidimensional space constitutes a complex *quale,* a particular quality of experience.
I am not concerned, here, with the consistency of this particular proposal, nor with its ontological claim about the fundamentality of consciousness – which is, nevertheless, controversial, since we are dealing with an emergent property of a physical substratum.[8]
The most significant point is that IIT provides the possibility of a geometrical treatment of conscious experience, a necessary step for whoever aims to construct a scientific theory of consciousness.

Conclusion

Several problems arise from IIT, and we cannot expect to explain consciousness entirely by adopting this theory. Firstly, there is a computational problem: in order to calculate φ, one is supposed to partition a system in all possible ways, and this seems an unfeasible procedure in a complex system like the human brain. Second, IIT does not explain how mental agency works. In other words, it does not explain how informational patterns, which give rise to experience, are also able to guide behavior and self-report (this problem could evoke the ancient problem of access/phenomenal consciousness).[9]
My aim here has not been to defend or to interpret IIT. Rather, it has been to acknowledge that the epistemological status of a SOC has eventually to account for a measure and a mathematical analysis of the phenomenon at stake. In doing this, IIT seems to be one of the most promising theories in the field,[10] and a future SOC will likely develop from its outlook.

[8] To see how IIT can bring about panpsychist implications, see Tononi & Koch (2015).
[9] See Block (1995).
[10] For a critical review, see Seth *et al.* (2006).

Future directions in consciousness studies should be concerned with:

i) the evolutionary (both in ontogenesis and in phylogenesis) development of a subjective point of view from a physical body which is embedded in an environment involving interactions with other living creatures;
ii) the gradation of consciousness in the sleep-wakefulness cycle;
iii) the relation between on-line consciousness (active when the organism is engaged in direct interactions with the world) and off-line consciousness (e.g. during NREM sleep, locked-in patients, paralysis, etc.);
iv) the feasibility of a measure of consciousness, in terms of both content and level of consciousness, that allows for empirically testable predictions.

It can be argued that when these empirical findings are organised in a theoretical framework, which considers conscious experience as intersubjectively constituted, the time will be ripe for the construction of a mature and mathematically informed science of consciousness.

REFERENCES

Bateson, G. (1972). *Steps to an ecology of mind*. Chicago, University of Chicago Press.

Block, N. (1995). On a confusion about a function of consciousness, *Behavioral and Brain Sciences*, 18, 227-287.

Chalmers, D.J. (1996). *The conscious mind: in search of a fundamental theory*. New York, Oxford University Press.

Chalmers, D.J. (1998). On the Search for the Neural Correlate of Consciousness, in S. Hameroff, A. Kaszniak, & A. Scott (Eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, Cambridge, MIT Press, 219-230.

Chalmers, D.J. (2000). What is a Neural Correlate of Consciousness?. In T. Metzinger (Ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, Cambridge, MIT Press, 17-39.

Chalmers, D.J. (2004). How can we construct a science of consciousness?. In M. Gazzaniga (Ed.), *The Cognitive Neuroscience III*, Cambridge, MIT Press, 1111-1120.

Dennett, D.C. (2003). Who's on first? Heterophenomenology explained, *Journal of Consciousness Studies*, 10(9), 19-30.

Dennett, D.C. (2007). Heterophenomenology reconsidered. *Phenomenology and the Cognitive Science*, 6(1-2), 247-270.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience, *Phenomenology and the Cognitive Science,* 4(1), 23-48.

Gallese, V. (2016). Finding the body in the brain. In B. McLaughlin & H.K. Kornblith (Eds.), *Goldman and His Critics*. West Sussex, John Wiley & Sons, 297-314.

Gallese, V. & Sinigaglia, C. (2011). What is so special with Embodied Simulation. *Trends in Cognitive Science*, 15(11), 512-519.

Glenberg, A. & Gallese, V. (2011). Action-based language: A theory of language acquisition production and comprehension. *Cortex*, 48(7), 905-922.

Hubel, D. (1974). Neurobiology: A science in need of a Copernicus, in J. Neyman (Ed.), *The Heritage of Copernicus*, Cambridge, MIT Press, 243-260.

Loar, B. (1997). Phenomenal States. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness. Philosophical debates*, Cambridge, MIT Press, 597-616.

Massimini, M., Boly, M., Casali, A., Rosanova, M., & Tononi, G. (2009). A perturbational approach for evaluating the brain's capacity for consciousness. *Progress in Brain Research*, 177, 201-214.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput Biol*, 10(5), DOI: 10.1371/journal.pcbi.1003588.

Olivares, F.A., Vargas, E., Fuentes C., Martinez-Pernía, D., & Canales-Johnson, A. (2015). Neurophenomenology revisited: second-person methods for the study of human consciousness. *Frontiers in Psychology*, 6, DOI: 10.3389/fpsyg.2015.00673.

Papineau, D. (1998). Mind the gap. *Philosophical Perspectives*, 12, 373-389.

Pauen, M. (2012). The second-person perspective. *Inquiry*, 55(1), 33-49.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirror in the brain*. Oxford: Oxford University Press.

Ryu, J., Borrmann, K., & Chauhuri, A. (2008). Imagine Jane and identify John: Face identity aftereffects induced by imagined faces. *PLoS One*, 3(5), DOI: 10.1371/journal.pone.0002195.

Reddy, V. (2003). On being the object of attention: implications for self- other consciousness. *TRENDS in Cognitive Sciences*, 7(9), 397-402.

Seth, A.K., Izhikevich, E., Reeke, G.N. & Edelman, G.M. (2006). Theories and measures of consciousness: an extended framework, *Proceedings of the National Academy of Sciences*, 103(28), DOI: 10.1073/pnas.0604347103.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., & Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), 1157-1162.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard: Belknap Press.

Tononi, G. (2003). *Galileo e il fotodiodo*. Bari: Laterza.

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, 215(3), 216-242.

Tononi, G. (2012). Integrated information theory of consciousness: an updated account. *Archives Italiennes de Biologie*, 150 (2-3), 290-326.

Tononi, G., & Massimini, M. (2013), *Nulla di più grande*. Milan: Baldini & Castoldi.

Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere?. *Philosophical Transactions of the Royal Society*, *B*, 370, DOI: 10.1098/rstb.2014.0167.

TIMOTHY A. BURNS
Loyola Marymount University
timothy.burns@lmu.edu

EMPATHY, SIMULATION, AND NEUROSCIENCE: A PHENOMENOLOGICAL CASE AGAINST SIMULATION-THEORY[1]

abstract

In recent years, some simulation theorists have claimed that the discovery of mirror neurons provides empirical support for the position that mind reading is, at some basic level, simulation. The purpose of this essay is to question that claim. I begin by providing brief context for the current mind reading debate and then developing an influential simulationist account of mind reading. I then draw on the works of Edmund Husserl and Edith Stein to develop an alternative, phenomenological account. In conclusion, I offer multiple objections against simulation theory and argue that the empirical evidence mirror neurons offer us does not necessarily support the view that empathy is simulation.

keywords

empathy, phenomenology, simulation-theory, mirror neurons, intersubjectivity, social-cognition

Phenomenology and Mind, n. 12 - 2017, pp. 208-216
DOI: 10.13128/Phe_Mi-21119
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

EMPATHY, SIMULATION, AND NEUROSCIENCE

1. Introduction

Phenomenology is marked, from its inception, by its engagement with the scientific thinking of its day. For example, Husserl's *Logical Investigations* contain meticulous criticisms of both psychologism and naturalism (2001). Indeed, it could be argued that his attempt to refute the naturalism that he diagnosed as prevalent in European science shaped the rest of his career from *Ideas I* to the *Crisis of the European Sciences*. Given the rich history that phenomenology has of engaging with the various sciences, it will be no wonder if this continued engagement shapes its future. In this paper, I will discuss one such current debate, namely the debate that revolves around the form of social cognition termed empathy/mind reading (I will use these terms interchangeably), and the significance of the relatively recent discovery of mirror neurons for this debate.

Most of the arguments concerning mind reading in the philosophical literature fall into one of three camps: simulation theory (ST), theory theory (TT), or direct perception, which, following Dan Zahavi (2011), I will call the phenomenological proposal (PP). Many simulationists claim the discovery of mirror neurons provides empirical support for the position that mind reading is, at some basic level, simulation. The purpose of this essay is to question that claim. I will begin by providing some brief context of the current mind reading debate. I will then draw on the works of Edmund Husserl and Edith Stein to develop PP. In conclusion, drawing upon and expanding the work of Husserl, Stein, and Scheler as well as recent phenomenological work in this area, I will argue that the empirical evidence mirror neurons offer us does not necessarily support the view that empathy is simulation.

2. Mind Reading

In philosophy of mind, "mind reading" is neither a parlor trick nor does it belong in the domain of dubious, self-proclaimed psychics. Simply put, it is the act of attributing mental states to other individuals. That human beings are able to read other's minds is an impressive ability, one of which we may be the sole possessors. Yes, many animals have mental states. However, as Alvin Goldman points out, it is one thing to *have* a mental state, and it is an entirely different thing to *represent* someone else as having a mental state (2006, p. 3).[2] In

[1] The author wishes to thank the editors of this special issue for their patience and diligence and the anonymous reviewers for their many insightful comments and critiques.

[2] There is a debate over the ability, or lack thereof, of animals to read minds. This topic is not within the scope of the current project. For an interesting recent account see Lurz (2011).

recent years there has been much literature produced on the question of how it is that we are able to accomplish this remarkable feat. On what basis do we attribute mental states, or minds more generally, to other beings? Answers to this question have tended to fall into two camps: TT and ST.

Those who subscribe to TT propose that we possess a certain folk-psychological theory about how minds – our own included – work. A folk psychological theory may be as simple as this. I have beliefs and desires about the world that motivate me to act in certain ways, so when I see persons behave in particular ways it is probably because of their beliefs and desires. Understanding another person is a matter of applying the theory in a way that will allow us to predict her future behavior and make her actions intelligible (see Dennett, 2011, pp. 87-106). When and how we come to possess such theories is a matter of some discussion (cf. Gopnik & Wellman, 1992), as is the issue of in what, exactly, the theory consists (see Baron-Cohen, 1995, esp. pp. 31-58).

On the other hand, those who subscribe to ST hold that, when we understand the mental lives of others, we do so by putting ourselves "in their mental shoes". Goldman defines ST in broad strokes when he writes, "[ST] says that ordinary people fix their targets' mental states by trying to replicate or emulate them" (2006, p. 4). In other words, understanding others is a matter of imitating them. Simulation theorists refer to different forms of simulation. For instance, there is explicit simulation, which is conceptually and linguistically mediated. There is also implicit simulation, which is meant to be non-conceptual, non-linguistic, and automatic. These theorists locate implicit forms of simulation at the subpersonal level. One of my criticisms of ST will be the use of 'simulation' to refer to both of these levels of description. However, I will return to this point in the conclusion. For the moment, ST is of particular interest because it is within ST camps that the word 'empathy' has recently resurfaced as a way to describe our mind reading abilities. Goldman has claimed that "mindreading is an extended form of empathy (where this term's emotive and caring connotations is bracketed)" (2006, p. 4). Karsten Stueber has gone so far as to identify simulations theorists as "today's equivalent of empathy theorists" (2006, p. ix). Given this current trend to identify the ST position on mind reading with empathy, the first goal of this essay is to explicate Goldman's influential simulation based account of empathy.

3. Simulation Theory

Simulation theorists hold that we understand others' mental states by trying to place ourselves in their "mental shoes". Gallese and Goldman write, "ST depicts mind-reading as incorporating an attempt to replicate, mimic, or impersonate the mental life of the target agent" (1998, p. 497). When I imitate the other's mental life, the goal is to achieve symmetry between my mental life and hers. Once I have simulated the other's mental state, I attribute it to her by projecting the mental state *I* have achieved into the other. So, in a decision prediction example of mind reading, Goldman writes, "According to ST the mind reader takes her *own* m-decision – a decision that occurs in the simulation mode, to be sure – and ascribes that type of state to the target" (2006, p. 40). A typical example would run as follows. I return my students their essays at the end of class. When I hand Eve her essay, she sighs, her shoulders slump, and she hangs her head. As she continues to stand there, her face reddens and her jaw clenches. She then stomps out of the room. In order to understand Eve's mental state I simulate her actions inwardly. The result is a mental state of disappointment and then anger. I then project these feelings into Eve. Having completed the "final stage" of my mind reading act, I now understand that Eve was disappointed and then angry.

It behooves us to notice that trying to put ourselves in the other's mental shoes, per Goldman, requires us to create "pretend states intended to match those of the target", which we then impute to the person we are observing (2005, p. 80). The imputation can be as simple as a two-step process: creation and imputation. Or it can involve a longer process in which I feed the pretend states into a psychological mechanism of my own – for instance a decision making mechanism – and allow this mechanism to work on the pretend states before attributing the results to the other (Goldman, 2005, p. 81). Either way, Goldman writes, "the distinctive idea of ST is that mind reading is subserved by pretense and attempted replication" (2005, p. 81).

Simulation theorists have heralded the discovery of mirror neurons as evidence in support of their theory. Mirror neurons are so called because it has been observed that the exact same neurons fire during "motor act" observation and performance (Gallese & Goldman, 1998, p. 495). Gallese and Goldman summarize the discovery, "Every time we are looking at someone performing an action, the same motor circuits that are recruited when we ourselves perform the action are concurrently activated" (1998, p. 495). In the study, motor acts that activated mirror neuron activity included grasping a tool so as to take possession of it with either the hand or the mouth and manipulating an object with a "precision grip", a grip between the thumb and the index finger of the same hand (Gallese & Goldman, 1998, pp. 493-494). Recent research has also shown that a listener's motor system is activated while perceiving gestures that speakers make (Ping, Goldin-Meadow, & Beilock, 2014). This is especially significant in simulation based theories of mind reading that rely on mirror neurons activity as empirical backing for their claims because it allows the agent of the observed actions to be candidates for the attribution of mental states such as belief, desire, and purpose of action. The argument is that if the same neural circuits are active in both the perception and performance of goal driven action, then it seems justifiable to attribute a mental state – e.g., desiring the realization of said goal – to the other person who is performing the action, and furthermore, when we do attribute a mental state to the observed agent, it is on the basis of having simulated her mental state (Gallese & Goldman, 1998, p. 494).[3]

4. The Phenomenological Proposal

At first blush, the neuro-scientific evidence that mirror neurons provide appears to be decisively in favor of a ST account of empathy. However, phenomenology proposes a different theory of empathy, one that is compatible with the scientific facts and nonetheless rejects ST. I will now offer a brief overview of the phenomenological theory of empathy.

In explaining PP, I will focus on the accounts of Edmund Husserl and Edith Stein with emphasis on their continuity rather than their subtle differences. Husserl and Stein insist that empathy is a unique form of experience. Husserl refers to empathy as a "special form of empirical experience (*empirischen Erfahrung*)" (2006, p. 82). Likewise, Stein defines empathy as "the experience (*Erfahrung*) of foreign consciousness in general" and wishes to divorce the term from any previous historical interpretation attached to it (1989, p. 11). Their analyses describe the object of empathic experiences and the mode of givenness of both the empathic experiencing and the empathized content. They agree in their insistence that careful analysis of the intentional structure of empathy will delineate it from closely related but diverse experiences. This will be important in the objections against ST that are to follow.

The object of the empathic experience (*Erfahrung*) is consciousness that belongs to an I that is not the empathizer's own. One ought to notice that their analyses do not seek to answer

[3] Since the publication of Gallese and Goldman (1998), Vittorio Gallese has developed his own account of social cognition, viz., Embodied Simulation (ES), that is distinct from Goldman's version of ST. See Gallese (2016) and Gallese and Sinigaglia (2011) for insights into the differences between ST and ES. See also Gallese and Cuccio (2015). Gallese (2001) seeks to revise and expand upon the concept of empathy, and speaks favorably of phenomenological accounts of empathy and intersubjectivity – notably of Stein and Merleau-Ponty. Furthermore, according to Overgaard and Zahavi (2012), Gallese considers ES to be a development of PP. Thus, the reader should note that the target of my argument is Goldman's ST, not ES.

the question, "Does one experience other subjects?". Their starting position is the experience of others. In other words, amongst the experiences that a conscious subject has, one finds experiences of other subjects, their experiences, and their conscious lives; the concept of empathy is meant to describe the intentional structures of these experiences. The ontological question is bracketed. A phenomenological theory of empathy is not a "proof" of the existence of other minds, and it does not purport to be.

Both Husserl and Stein seek the essence of empathy by comparing it to other experiences. In doing so, they draw a distinction between originary/primordial and non-originary/non-primordial experience. All experience, insofar as it is one's own, is originary. This may mean something as simple as that the experience is had from, or in, the first person perspective. However, that an experience is originary also says something about the way in which its object is present. For Husserl, external perception is the case, *par excellence*, of originary experience. It is the direct having of the perceptual object. In it, one has a really given access to the physical thing (Husserl, 1983, p. 5).

Non-originary experiences, on the other hand, are ones in which the object is not "itself there" in the same way. The object of non-originary experience is still present to consciousness; however, it is present in a different manner. Husserl tends to designate a quartet of experiences as ones that exemplify non-originary experience: memory, expectation, fantasy, and empathy. Take memory as an example. The object of memory is not present in the same manner as the object of external perception. The object of the memory is there, but not in an originary manner; it is there as remembered, *as having been previously*. It may be remembered as *having been* originarily present; for example, one may remember seeing a particularly beautiful sunset. The sunset was "itself there" in the prior experience, but its presence in memory is different in form from its presence in the perceptual experience.

Empathy possesses a unique intentional structure; it is originary experience with non-originary content, the object of which is the experience or consciousness of a foreign subject. Let us consider one of Stein's examples: "A friend tells me that he has lost his brother and I become aware of his pain" (1989, p. 6). The object of the empathic experience is the friend's pain. The content of the outer perception, considered as purely physical data, comprises the friend's face, his voice, and the position and posture of his body. The pain does not show up the way that these other things do. Nonetheless, the friend's pain is there for me. I do experience it. There is a primordial experiencing. However, I experience the pain *as his*. The content of the empathized experience is given in a non-originary mode – as belonging to another subject.

As a non-originary experience, Husserl classifies empathy as belonging to the broader category of apperception (2006, p. 83; 1989, p. 177). Apperception is the name given to something that is perceived with or alongside another perception; it thus encompasses the concepts of empty intention and horizon, which Husserl was beginning to develop around this time. All apperception is founded on originary perception. The case *par excellence* of apperception is that found in the visual perception of a physical object. In the experience of a physical thing, only the side of the object facing me is actually present to consciousness. It belongs to the essence of a physical thing, according to *Ideas I*, to only be able to only possibly be given in these "one sided adumbrations" (Husserl, 1983, p. 9). I nonetheless apperceive the averted sides of the object.

There is, however, an important distinction between the apperception involved in perception of a physical object and empathy as a form of apperception. The apperceived sides of an object may, in principle, come to originary givenness through a harmonious course of experience. This is impossible for the apperceived mental states of another subject. This is an eidetic law of consciousness. There is "no channel linking the empathized stream to the stream in which the empathizing itself belongs" (Husserl, 2006, p. 85). The essential difference between empty horizons that accompany the perception of physical objects and the apperception of another's consciousness is their mode of fulfillment. As regards the averted sides of a physical object, it is possible to bring them to intuitive presentation whereas that which is given through an empty horizon in empathy must, because of the nature of the object being appresented, remain empty. In other terms, the phenomenological claim is that what counts as perception of a human being is essentially different than what counts as perception of a physical object. There are an infinite number of perspectives one may take on a physical thing and thus an infinite number of empty intentions implied in the perception of a physical object. These may, because of the type of thing that it is, be brought to original givenness. What it means to perceive *the kind of thing* that a human is involves *both* the perception of body and the apperception of an inner life. The perception of a human is "an 'incomplete' one, being constantly *open*, since of this human being there and, especially of his interiority this perception expresses only a few things" (Husserl, 2006, p. 150). Perception of another is an open-ended project, one that is never fully accomplished. It remains in its essential peculiarity a conjunction of the founding presentation of the other's body and her appresented interiority.

5. Criticisms of Simulation Theory

With that brief understanding of the phenomenological account of empathy in hand, I now wish to draw your attention to one aspect of ST that will serve as the basis for my objections. The simulation account of mind reading involves my arriving at a mental state *of my own* and then attributing an identical mental state to the target subject. This cannot be denied. For instance, Gallese and Goldman write, "In the simulation scenario there is a distinctive matching or 'correspondence' between the mental activity of the simulator and the target" (1998, p. 497). Later in the same article, they identify mirror neuron activity as creating "in the observer a state that matches that of the target" (Gallese & Goldman, 1998, p. 498). In other words, according to ST, when I understand your mental state, I have one that matches it (see also de Vignemont & Singer, 2006, pp. 435-441).

Phenomenologists reject the idea that empathy requires that the empathizing subject must have a mental state that matches that of the target subject. It seems odd to require this. Consider again the example of the student receiving the grade on her essay. According to ST, in order for me to grasp her mental state – to understand that she is disappointed and then angry – I must simulate her states. However, when I see the student's reaction to her grade, I am neither disappointed nor angry. I simply experience her *as* disappointed and angry. Neither will it help to say that I have these mental states in the mode of simulation, as Goldman suggests (2006, p. 40). Such a suggestion does not align with the "things themselves". The simulation story is not true to the experience of the face to face encounter. In such a scenario, I simply perceive the student's mental states without the need to simulate either her bodily movements or her inner state. Consider another example. Suppose that my downstairs neighbor bangs on my door and, when I open it, he yells at me, complaining that the music is too loud and he cannot sleep. I see that he is angry. In order for me to grasp this, ST requires that I also be angry. However, I need not become angry in order to perceive my neighbor's anger. I may be understanding, apologetic, unconcerned, or angry depending on the situation, our history, and the level of concern that I have for others' feelings. The point is that, if I do react out of anger, it is just that, a reaction. There is no reason to assert that I must experience *as-if* anger in order to perceive another's anger.

Max Scheler identifies a further reason to reject ST. A simulation plus projection model fails to distinguish between empathy – as the experience of another's inner life – and other forms of social cognition, especially emotional contagion (*Gefühlsansteckung*) and sympathy (*Mitgefühl*) (Scheler, 1979, pp. 14-18). The simulation of another's mental state may very well be an example of emotional contagion. For example, if I am at an opera sung in Italian and cannot understand the plot, the emotions of the crowd may sweep over me and I may have joyful or sorrowful experiences *as-if* they were my own. However, emotional contagion by way of such simulation is generally distinguished from empathy because of its anonymous character. In emotional contagion, I do not know from whom I receive the emotional states, and in a real sense, I receive them from no one in particular but rather from the crowd itself.

Consider now the difference between empathy and sympathy. Sympathy is thought to involve an understanding of the other's mental state and a sharing of it. Suppose that my friend comes to me sobbing and tells me that his mother has passed away. I am first aware of his sadness in the pained expression on his face and the heaving of his chest. If I also happen to grieve over the loss of his mother, this is an emotional response above and beyond knowing or understanding that he is grieving. As Scheler writes, "[M]y having an experience similar to someone else's has nothing whatever to do with understanding him" (1979, p. 11). His grieving and my commiserating with him are separate facts and must be explained separately. "Fellow-feeling proper, actual 'participation', presents itself in the very phenomenon as a *re-action* to the state and value of the other's feelings" (Scheler, 1979, p. 14). Sympathy with another's grief, or joy, implies first an empathic understanding of the emotion and, second, a similar or pro-social feeling. However, a simulation based account of empathy fails to distinguish between these two because it requires that the empathizer share, in some form, emotional states with the target subject insofar as the states are simulated in the empathizer and then projected into the other.

Furthermore, it seems that requiring simulation of the other in order to understand his mental life runs the risk of begging the question. Scheler notes, "imitation, even as a mere 'tendency', already presupposes some kind of acquaintance with the other's experience and therefore cannot explain what it is here supposed to do" (1979, p. 10). Let us return to the example of the grieving friend. I can only have an emotional state similar to his if *I already know* what his emotional state *is.* I cannot simulate his grief unless I already know that he is grieving. Such an acquaintance with his inner state is provided by empathy, in the phenomenological sense. It cannot be based on simulation unless I am already familiar with the target subject's mental state; and then, the simulation theorist is caught in a circle.

The final objection against ST is that it does not provide us with knowledge of other's mental states. Simulation, as Zahavi and Overgaard write, "seems de facto to imprison me within my own mind and to prevent me from ever encountering *others*" (2012, p. 9). Simulation plus projection only ever *actually* arrives at self-understanding, not an understanding of others. I experience mental states of my own; I then project them into the other, making the assumption that this must be, or probably is, what she is experiencing. Thus, as Stein points out in criticizing a similar theory from Theodor Lipps, even as a genetic account of empathy, a simulation plus projection theory fails to explain what it sets out to explain (1989, pp. 22-24).

6. Conclusion

But, what about mirror neurons? Does the neuro-scientific evidence not point to evidence of simulation in brain activity at a basic level? The evidence seems to suggest that my neural circuits imitate goal-directed action when I perceive it. I claimed above that the phenomenological account was consistent with the scientific data, and this is the point I now wish to address. Referring to the evidence of mirror neuron activity, Stueber, despite being a simulation theorist himself, observes that "it must be noted that the evidence does not provide conclusive support for [any] theoretical paradigm" (2006, p. 117). He goes on to note ways in which proponents of TT may interpret mirror neuron activity in support of their own theory. It is crucial not to over-interpret empirical data.

The mere discovery of mirror neurons does not prove that our awareness of others' mental states is based on simulation. I find it less than conclusive for this reason. Simulating, emulating, or imitating another is a *personal level* activity. It is something that persons do or they do not.[4] It seems strange to refer to both my personal level activity of imitating someone's actions – as I may do when I am recounting to my son the tale of how a clown danced at the circus – and my unconscious brain activity with the same word, 'simulation', and expect that it should mean the same thing in both cases. And yet, this appears to be exactly how Goldman interprets the neuroscientific data. He writes, "Mental simulations might occur automatically, without intent, and then get used to form beliefs about mind-reading questions" (Goldman, 2006, p. 40). And, "[R]ecent cognitive science and cognitive neuroscience disclose striking instances of mental simulation that are largely automatic and unconscious" (Goldman, 2006, p. 49). In so doing, Goldman collapses two different levels of description that are relevant to our account of empathy – the personal and the subpersonal – by using 'simulation' to describe the processes at work in both. The fact that the same neural networks are active in action perception and in action performance is unsurprising at some level. Why should it not be the case that perception of goal oriented action involves neural stimulation of the same neural pathways as performance of goal oriented action? The mere presence of mirror neuron activity is not *prima facie* evidence in favor of ST. The phenomenological account of empathy can be true even in light of the neuro-scientific evidence that mirror neurons provides, and given the problems with ST, should be preferred.

REFERENCES

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Dennett, D. (1971). Intentional systems. *The Journal of Philosophy*, 68 (4), 87-106.

de Vignemont, F. & Singer, T. (2006). The empathic brain: How, when, and why?. *Trends in Cognitive Science*, 10 (10), 435-441.

Gallese, V. (2016). Finding the body in the brain: From simulation theory to embodied simulation. In B.P. McLaughlin & H. Kornblith (Eds.), *Goldman and his critics* (1st ed.). West Sussex, UK: John Wiley & Sons, 297-314.

Gallese, V. (2001). The 'shared manifold' hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies*, 8 (5-7), 33-50.

Gallese, V. & Cuccio, V. (2015). The paradigmatic body-Embodied simulation, Intersubjectivity, the bodily self, and language. In T. Metzinger & J.M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group. DOI: 10.15502/9783958570269.

Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Science*, 2 (12), 493-501.

Gallese, V. & Sinigaglia, C. (2011). What is so special about embodied simulation?. *Trends in Cognitive Science*, 15(11), 512-519.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* Oxford: Oxford University Press.

Goldman, A. (2005). Imitation, mind reading, and simulation. In Chater, N. & Hurley, S.L. (Eds.), *Perspectives on imitation: From neuroscience to social science*. Cambridge, MA: MIT Press, 79-91.

Gopnik, A. & Wellman, H. (1992). Why a child's theory of mind really 'is' a theory. *Mind and Language*, 7 (1-2), 145-171.

Husserl, E. (2006). *The basic problems of phenomenology: From the lectures, winter semester (1910-11)* (Engl. Transl. by I. Farin & J. G. Hart). Dordrecht: Springer.

[4] In a footnote to "Empathy Without Isomorphism" Overgaard and Zahavi make a similar observation regarding empathy and I am indebted to them for this insight. See note 12 in the above cited article.

Husserl, E. (2001). *Logical investigations* (D. Moran, Ed.) (Engl. Transl. by J.N. Findlay). London & New York: Routledge (Original work published 1900-1901).

Husserl, E. (1989). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: Second book* (Engl. Transl. by R. Rojcewicz & A. Schuwer). Edmund Husserl: Collected Works, 3. Dordrecht & London: Kluwer Academic Publishers.

Husserl, E. (1983). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: First book* (Engl. Transl. by F. Kersten). Edmund Husserl: Collected Works, 2. Dordrect & London: Kluwer Academic Publishers.

Lurz, R. (2011). *Mindreading animals: The debate over what animals know about other minds.* Cambridge, MA: MIT Press.

Overgaard, S. & Zahavi, D. (2012). Empathy without isomorphism: A phenomenological account. In J. Decety (Ed.), *Empathy: From bench to bedside*. Cambridge, MA: MIT Press, 3-20.

Ping, R., Goldin-Meadow, S., & Beilock, S.L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, 143(1), 195-204.

Scheler, M. (1979). *The nature of sympathy* (Engl. Transl. by P. Heath). London: Routledge & Kegan (Original work published 1923).

Stein, E. (1989). *On the problem of empathy* (Engl. Transl. by W. Stein) (3rd rev. ed.). Washington, DC: ICS Publications.

Stueber, K.R. (2006). *Rediscovering empathy: Agency, folk psychology, and the human sciences.* Cambridge, MA: MIT Press.

Zahavi, D. 2011. Empathy and Direct Social Perception: A Phenomenological Proposal. *Review of Philosophy and Psychology*, 2, 541-558.

JOHN JOSEPH DORSCH
University of Tübingen
johnjosephdorsch@gmail.com

ON EXPERIENCING MEANING: IRREDUCIBLE COGNITIVE PHENOMENOLOGY AND SINEWAVE SPEECH

abstract

Upon first hearing sinewaves, all that can be discerned are beeps and whistles. But after hearing the original speech, the beeps and whistles sound like speech. The difference between these two episodes undoubtedly involves an alteration in phenomenal character. O'Callaghan (2011) argues that this alteration is non-sensory, but he leaves open the possibility of attributing it to some other source, e.g. cognition. I discuss whether the alteration in phenomenal character involved in sinewave speech provides evidence for cognitive phenomenology. I defend both the existence of cognitive phenomenology and the phenomenal contrast method, as each concerns the case presented here.

keywords

irreducible cognitive phenomenology, sinewave speech, experiencing meaning

Phenomenology and Mind, n. 12 - 2017, pp. 218-227
DOI: 10.13128/Phe_Mi-21120
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

It helps to think of sinewave speech as a phenomenal contrast. In the naive case, you hear the sinewave as noise. In the informed case, you hear the sinewave as speech. So undoubtedly, an alteration in phenomenal character has taken place. How did this happen? In between the naive and the informed case, you heard the original speech, from which the sinewave is derived. After having heard the original speech, you hear the whistles and beeps as syllables and words. Since the audio stream does not change from one case to the next, one is motivated to consider whether the alteration in phenomenal character is attributable to some extrasensory faculty, such as cognition.[1]

In this paper, I discuss whether and to what degree sinewave speech provides evidence for cognitive phenomenology. Before proceeding, I wish to dismiss a possible misconception. If you wish to claim that sinewave speech involves an alteration in the phenomenal character of sensorial content, I believe there is only one possible path open to you, other than perhaps building a case based on shifting one's attentional focus, which will be discussed later. You need to argue that the difference between the two cases can be accounted for by appealing to mnemonic recall of the original speech. In other words, upon hearing the sinewave the second time, one recalls and replays the sounds of the original speech. This claim, however, does not correspond to the phenomenology of hearing the sinewave the second time. If you are like me, when you hear the sinewave in the informed case, it does not seem to you that you hear a mnemonic recall of the original speech; rather, it seems to you that you can actually hear the words of the original speech in the sinewave, whereas before you heard only beeps and whistles. Appealing to the phenomenology of hearing the sinewave the second time is not meant to deny the role that memory plays in hearing the sinewave as speech. The argument provided by concentrating on the phenomenology simply amounts to denying that one hears an internal, monological echo or mnemonic replay of the original speech.

I. Let us now discuss to what extent sinewaves provide evidence for irreducible cognitive phenomenology. To clarify, by 'irreducible cognitive phenomenology' I mean a phenomenology of cognition that is not reducible to the phenomenology of the senses, whether imagined or otherwise (see Chudnoff, 2015).

[1] I strongly suggest that readers experience sinewaves themselves. Some examples can be found at the following web-address: http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/SWS/ (accessed July 5th, 2017).

I would like to begin by performing an anatomy of both the naive and the informed cases. The method here is similar to the method developed by Siegel (2007), based on introspection and inference to the best explanation. Recall that the naive case involves hearing the sinewave before having heard the original speech, whilst the informed case involves hearing the sinewave after having heard the original speech. To conduct an anatomy of these cases, I find it helpful to concentrate on the distinguishing phenomenal character involved in each. In the naive case, merely beeps and whistles are heard. These sounds do not seem to refer to, indicate, or otherwise mean anything that the listener can yet parse. I will call the conscious content of hearing the beeps and whistles the "content" of the episode, which I take to be almost entirely comprised of sensorial elements. I say 'almost' because the content of auditory experiences is rarely ever meaningless and thus possesses cognitive elements.

If you hear a screech and subsequent bang from outside your window, you may think you just heard a car accident. The screech referred[2] to the brake drums applying pressure to the wheels, while the bang referred to the sudden impact of the car crashing into something solid. In this case, the content of the auditory experience had meaning, that is to say, the content of the experience referred to something in a meaningful way. In the case of sinewaves, the listener has never heard these sounds before, so she does not know what the sounds refer to, which is an altogether different claim than saying that the content is meaningless and without referents. Rather, the content of hearing the sinewave in the naive case possesses, as of yet, unknown referents. If you are like me, when you first hear the sinewave you wonder about what you just heard, to what it might refer. So it can be said that the content of the auditory experience in the naive case is mostly comprised of sensorial elements and partially comprised of cognitive elements due to the content's possession of, as of yet, unknown referents. In this sense, unknown referents can be thought of as placeholders for possible, forthcoming referents.

I will now continue the anatomy by examining the informed case. Recall that the informed case involves hearing the sinewave a second time after having heard the original speech. Now the sinewave seems to sound wholly different. What before was a beep has become a syllable, and the fade in the whistle's pitch has begun sounding like a word. Nothing has changed about the sinewave, and yet the sinewave seems to sound like a digitized human voice, so that what the listener hears is the language of the original speech.

Dissecting this episode reveals the same sensorial content as before; however, I will argue that the sensorial content of the informed case is arranged differently than the naive case due to how the listener focuses her attention. If you are like me, when you hear the sinewave the second time, you find yourself attending to the modulation in pitch and tone more acutely. O'Callaghan makes this point clear when he discusses the role phonetics plays in language learning (O'Callaghan, 2011). As learning to hear the phonemes of a foreign language requires learning to shift attention to the sounds that segment and distinguish the phonemes, hearing the sinewave after having heard the original speech demands attending to certain minute shifts in pitch and tone. Learning to hear the phonemes of a foreign language is similar to learning to hear the phoneme-esques of sinewave speech. Having heard the original speech, one is provided with a kind of map, with which the strange auditory landscape can be navigated. This leads me to claim that, although the sensorial content is the same, it has a slightly different form in the informed case because the listener has shifted her attentional focus to those sounds that distinguish the phoneme-esques of the sinewave. But can the

[2] This use of 'refer' is non-standard. I mean that one infers from the sound that some particular event is the cause of that sound; so in this loose sense, the sound refers to the event.

alteration in phenomenal character from naive to informed case be explained completely by an attentional shift?

II. The attentional shift helps explain how it is possible for a listener to hear one and the same audio stream and yet undergo two different experiences. In the naive case, the listener does not know the referents of the sounds – she does not even know how to interpret the sounds she hears. Upon hearing the original speech, she experiences the dynamics needed for parsing the sounds. Hereafter, she attends to the sinewave with enhanced focus, due to having already experienced the dynamics of the original speech. The sensorial elements of the content can now be arranged, so that the sinewave sounds similar to the original. However, a question now arises about how this concerns the cognitive elements of the content, i.e. the (un)known referents.

In order to address this question, the "referent-resolution" of the informed case needs to be examined. By 'referent-resolution' I mean resolving the referents of the sounds into words, so, for example, resolving a whistle into a word. Recall that in the naive case, the cognitive content consists of unknown referents, which is wholly different from consisting of no referents, since the unknown referents act as placeholders for possible, forthcoming referents. In the informed case, the question mark of the unknown referents is resolved: the listener knows what the sounds of the sinewave refer to. It is clear that this referent-resolution is, in part, due to the listener's attentional shift, but the question is not whether the attentional shift brings about the referent-resolution, which it certainly does; rather, the question is whether the referent-resolution can be *reduced* to the shift in attentional focus.

To begin, I need to clarify what experiential states and phenomenal properties are. Any state is an experiential state if it makes sense to ask what it is like for an agent to seem[3] to be in that state. Imagine that you are looking at a red apple. It makes sense to ask what it is like for you to seem to be in this state, so this state is an experiential state. You might describe what it is like to seem to be in this state by saying that this state seems to possess redness – due to the apple seeming to appear red. You have thus provided one phenomenal property of the state of looking at a red apple, i.e. redness. In other words, any experiential state possesses a phenomenal property, if it makes sense to ask for a description of what it is like to seem to be in that state, while any meaningful description will single out a phenomenal property of that state. It is now possible to define the following condition for reduction:

> For any state α, α can be reduced to some other state β, if and only if β possesses all the properties of α. Conversely, if α possesses some property, for which no combination of β-properties suffice, then α is irreducible to β.

Returning to the above question, "Can the state of referent-resolution be reduced to the state of shifting one's attentional focus?", it is now possible, with the conceptual toolkit above, to sketch an answer. Again, I will begin by performing an anatomy of the two states. What is it like for you to seem to be in a state of shifting your attentional focus? If you are like me, in the sinewave scenario, you found yourself concentrating on the tones of the sinewave, extending your attentional focus on these tones, molding your focal point according to the dynamics provided by the original speech. This is not to say that you changed the sound, but rather, you molded your attention in order to shape the reception of the sound. In addition, you

[3] The emphasis on 'seeming to be' is meant to indicate that the experience need not be veridical; i.e. it does not matter whether the agent is under an illusion or hallucinating; all that matters here is that it merely seems to be so.

found yourself emphasizing those tones that seemed to match the original speech, and thus prolonging your attention to them. So, one may say that the state of shifting one's attentional focus possesses two properties: the molding property and the emphasizing property.

Let us now move to the state of referent-resolution. In our scenario, if you are like me, you found yourself visualizing the different objects the sounds refer to. One might think that the sounds refer to words, but, if you are like me, you did not find yourself visualizing (or imagining in any other way) any words. So for instance, for the sinewave, whose original speech is "The kettle boils quickly", I found myself, upon hearing the sinewave a second time, visualizing a boiling kettle. I will call this the "imagery property" of the state.

Comparing the state of referent-resolution with the state of shifting one's attentional focus, it becomes clear that shifting one's attentional focus is a poor candidate for reduction. It seems to me that there is no possible way to reduce the imagery property to some combination of the molding and emphasizing properties. Perhaps, molding and emphasizing are required to hear the word 'kettle', but hearing 'kettle' is not sufficient for explaining the visualization of an imagined kettle-percept. As it concerns the overarching discussion, the imagery property is, however, not sufficient for establishing sinewave speech as evidence of irreducible cognitive phenomenology, since image-based thinking can be reduced to sensory phenomenology. But the imagery property is not the only distinguishing characteristic of referent-resolution. While listening to the sinewave in the informed case, I seemed to imagine not only a boiling kettle, but also a context in which it is appropriate to utter the statement "The kettle boils quickly". When I reflect on what I had in mind, I realize that I was aware of a situation, in which two friends were meeting for tea, with the speaker of the utterance being the host, implicitly telling his guest by the statement "The kettle boils quickly" that the tea would have soon been ready. Now, you may have had a different situation in mind; perhaps, your situation was so transient that it did not provide an imaginary host; but regardless of how you thought the situation to be, I am convinced that an awareness of context is a necessary property of being in a state of referent-resolution and an awareness of context "feels" like something (more on that feeling later).

Notice that an awareness of context cannot be reduced to an awareness of sensory imagery, whether this imagery is imagined or otherwise. Suppose you watch two films: the first, a horror film; the second, a drama about modern medicine. Suppose further that the producers of both films purchased identical stock footage for a scene involving surgical equipment. So, the only difference between the scenes is the context in which each is situated; and awareness of this context cannot be reduced to an awareness of on-screen images. You might claim, however, that the context can be reduced to the previous scenes in the film. It seems to me that this claim may be unpacked in two ways: either atomically or holistically. Atomically, the context is reduced to the sum of all individual scenes. But this does not hold. Upon seeing the surgical equipment, one does not recall and visualize the previous scenes. This means, the awareness of context does not reduce to awareness of the previous scenes. Holistically, context is reduced to the individual scenes taken as a whole. But this does not seem to support the claim. I find it difficult to imagine what "the scenes taken as a whole" would look like. And yet, I *do* seem to have an awareness of the scenes taken as a whole, from which the context for the surgical equipment is determined. So, it seems that an awareness of context is not reducible to an awareness of sensory imagery, imagined or otherwise.

III.

Above I said that an awareness of context is a necessary property of being in a state of referent-resolution. In other words, without being aware of some context implied by a meaningful statement, one cannot be said to be in a state of having resolved the linguistic reference of that statement, and thus cannot be said to have understood the statement. This is a large claim, and the irreducibility of cognitive phenomenology as it pertains to speech perception hinges on it, so allow me to unpack it.

The claim can be unpacked as follows: given any linguistic statement, understanding that statement requires one to be aware of a possible context, in which it is meaningful to utter it. I will argue for this claim with three examples: indexicality, ambiguity and vagueness. The first example involves indexicals. A statement such as "I am here now" is only meaningful if one can resolve the referents of 'I', 'here' and 'now', which change depending on the context. Therefore, understanding the meaning of that statement requires being aware of context.

The second example dispenses with indexicals and instead focuses on semantic ambiguity. A statement such as "The letter is in the drawer" can only be meaningful if you seem to be aware of the correct referent of 'letter'. If you think that 'letter' refers to an alphabetic symbol, and not a missive, then this statement verges on meaninglessness. That said, perhaps you and I are both working at the cinema one summer. The latest blockbuster has come to our town, and we are charged with the task of updating the marquee. You go into the office and grab the box of marquee-letters, but notice that 'A' is missing. So you shout from behind the desk "Where is A?" and I respond "The letter is in the drawer!". In order to find the appropriate referent for 'letter' in the first scenario, one must draw an inference from the adverbial phrase 'in the drawer'. But, as the second scenario shows, the referent of 'letter' is not explicated by the adverbial phrase, but merely implied by it; thus awareness of context is needed in order to resolve semantic ambiguity.

The third example dispenses with indexicality and semantic ambiguity and focuses on vagueness. Consider the statement "The man threw the ball". At first glance, the referents of this statement do not seem to depend on the context. But, depending on the situation, one could be talking about either the man or the ball. Consider the following questions: "Who threw the ball?" and "What did the man throw?". For each of these questions, a different context is implied, and without at least one of them, the statement remains vague. I would therefore claim that the person who is unaware of a possible context has not understood the statement and has not resolved its referents.

IV.

Returning to our discussion of irreducible cognitive phenomenology, I will now discuss how referent-resolution fits into the larger picture. If it can be shown that the phenomenal property of the awareness of context involved in the state of referent-resolution cannot be reduced to any combination of the properties of a wholly sensory state, imagined or otherwise, then the state of referent-resolution is a possible candidate for irreducible cognitive phenomenology, and, furthermore, sinewave speech provides evidence for it.

In what follows, I will examine whether indexicality offers support for irreducible cognitive phenomenology. To begin, I will stipulate the state in question as the state in which you are aware of the referents of the indexicals 'I', 'here' and 'now' in the statement "I am here now". When one is in a state of referent-resolution of the above statement, what is one aware of? In order to address this question, I will draw upon the direct reference theory developed by Kaplan, who builds a conceptual framework for determining the referent of an indexical (Kaplan, 1989). Kaplan's framework is based on two notions: content and character. He holds that the content of a statement consists of those factors that determine the truth-value of the statement, and argues that the context of a statement containing an indexical needs to be, by way of the indexical's character, determined before the content can be determined. So, character + context = content.

Since the content is equated with the referent of the indexical, I will focus on resolving the content of the indexicals – for the sake of brevity, I will consider only the indexical 'I' in the above statement. I will call 'I''s content a "phenomenal property" because it seems to make sense to ask what it is like to be in a state of referent-resolution of the statement “I am here now” and you might reply meaningfully “It is like being aware of the content of ‘I’”. Understood thusly, awareness of ‘I’’s content becomes a phenomenal property of the experiential state of referent-resolution. So the question becomes, can this property be reduced to any combination of the properties of a wholly sensory state?

Consider hearing and seeing someone saying, “I am here now”. You hear the word ‘I’ being spoken and see the word being enunciated. According to Kaplan, one arrives at the content by way of the indexical’s character. The character of ‘I’ is a function that takes some context as argument and returns the agent of that context, while the agent is the content of ‘I’ (Kaplan, 1979). This means that being in a state that possesses awareness of ‘I’’s content consists of being aware of a context through which ‘I’ comes to have its content. So, a question arises about whether hearing and/or seeing the speaker seems to make you aware of the context required to render the content of ‘I’.

I have serious doubts that a wholly sensory state seems to make you aware of the context necessary for rendering the content of ‘I’. In order to determine whether the sensory state can do this, we should answer this question: what is required of the sensory state in order for it to render the content of ‘I’? It seems plausible to respond that two conditions need be met: 1) the sensory state seems to make you aware of the context, and 2) the content of the sensory state is sufficient for rendering the content of ‘I’. That said, I believe that conditions 1) and 2) cannot be met by the sensory state.

The first condition hinges upon the sensory state seeming to make you aware of the context. Arguably, context is the background information relative and relevant to a situation, where ‘background information’ means ‘other than the presently and immediately available information’. In order for the sensory state to meet the first condition, its content would need to include some background information, that is to say, the content of the sensory state would need to include some information other than the presently and immediately available moment; but the sensory state of looking and hearing someone saying “I am here now” cannot include anything other than the present moment. This means that the content of the sensory state alone cannot seem to make you aware of the context.

Now consider the second condition. You might respond by claiming that the speaker-percept is sufficient for rendering the content of ‘I’. But this is only possible if the speaker is the referent of ‘I’; and that is not necessarily the case. For example, the speaker could explicitly say, “He said, ‘I am here now’”, so that ‘I’ no longer refers to herself. Alternatively, ‘that he said it’ could be implied by the speaker. Either way, ‘I’ does not always refer to the agent speaking, so there is no guarantee that the speaker-percept is sufficient for rendering the content of ‘I’. So, the second condition is not met as well.

This means that the state of referent-resolution, due to its dependency on awareness of context, possesses properties that cannot be reduced to properties of wholly sensory states, imagined or otherwise. In turn, this means that, provided that referent-resolution possesses phenomenology, it would be a candidate for irreducible cognitive phenomenology, while sinewave speech provides evidence for it.

V.

It is time now to return to sinewave speech. Recall what it was like for you to hear the sinewave as speech. Ask yourself, “was there something it was like for me to resolve the referents of the strange sounds?”. If you are like me, along with a host of other phenomenal properties, you also had a feeling of understanding, or an experience of meaning, which you did not undergo the first time. I believe that this feeling of understanding is no different than an awareness of context, which I have argued is a candidate for irreducible cognitive phenomenology.[4]

In addition to the evidence from the arguments above, I would like to share some further evidence for irreducible cognitive phenomenology provided by the neurological research on sinewaves. Several studies provide reason to believe that phonetic perception, compared to acoustic perception, involves distinct neurological locations and processes (Lambertz *et al.*, 2004; Benson *et al.*, 2005; Möttönen *et al.*, 2005). In these studies, listeners engaged with sinewaves in much the same way as presented here: in the naive case, sinewaves were heard as beeps and whistles, while in the informed case, sinewaves were heard as speech. Brain imaging (fMRI) revealed increased activation of the superior temporal sulcus in informed cases only, providing evidence for the claim that phonetic perception involves a distinct neurological mechanism. If one assumes a correlation between brain activity and phenomenology, the independence of phonetic perception from acoustic perception bolsters the present case for irreducible cognitive phenomenology. Furthermore, according to Lambertz *et al.* (2004), sinewaves show that perceiving phonemes has an inhibitory influence on perceiving acoustics. If one assumes the correlation mentioned above, then the inhibiting effect on acoustic perception by phonetic perception lends support to the claim that cognitive phenomenology is decoupled from sensory phenomenology, which, in turn, lends support to the claim that the phenomenology of speech perception involves irreducible cognitive phenomenology.

VI. In conclusion, I would like to respond to some objections. Prinz (2012) provides four challenges to the existence of cognitive phenomenology. I will argue that the case presented here meets these challenges. Fish (2013) raises concerns regarding the phenomenological method. I will argue that the contrast provided by sinewaves addresses his concerns. Brogaard (2016) defends the view that meanings are perceived. I will argue that meanings are the result of cognitive, not sensory, phenomenology.

Prinz Prinz (2012) raises four challenges for cognitive phenomenology: *distinctiveness*, *isolability*, *inaccessibility*, and *inner speech*. Concerning *distinctiveness*, the challenge is to provide a contrast in which the difference between the two cases is solely cognitive. By keeping the sensory data fixed, sinewave contrasts seem to meet this challenge. Regarding *isolability*, Prinz believes that if cognitive phenomenology cannot be isolated from other episodes, then there is no phenomenology associated with cognition. The feeling of understanding, as the result of an awareness of context, can be isolated: I have no trouble isolating the moment that “it all clicked” for me. Thirdly, due to sub-personal factors involved in language comprehension, such as parsing syntax, Prinz thinks it is plausible to maintain that there is no phenomenology of cognition. The feeling of understanding, however, is readily *accessible* to introspection. Finally, Prinz argues that any putative cognitive phenomenology is reducible to sensory phenomenology of either mental imagery and/or *inner speech*; but as I argued above, due to its dependency on awareness of context, the feeling of understanding is neither reducible to sensory phenomenology, nor derived mental imagery. So, the case presented here meets Prinz’s challenges.

[4] If you find that the feeling of understanding still eludes you, I suggest that you read the analysis of understanding an ambiguous newspaper headline at the end of this paper; there you may find that the feeling of understanding is more palpable.

Fish

Fish (2013) raises the following objection to Siegel's method of phenomenal contrast, "… unless we can eliminate the concern that the phenomenological method reproduces antecedent assumptions, rather than delivers new evidence, such a claim should be treated with skepticism" (p. 53). Call this *Fish's Challenge*. The challenge is directed at what Fish calls the "phenomenological method", which includes the contrasts presented by Siegel and the contrast presented here. That said, I would like to point out a fundamental difference between the contrast discussed here and Siegel's; I believe this difference enables our contrast to meet Fish's Challenge.

Consider the contrast (presented by Siegel) concerning seeing trees as pine trees. In the naive case, you have little experience with the concept PINE TREE, and so you do not discern pines from other trees. In the informed case, on the other hand, you have acquired some skills regarding the concept and, therefore, you are able to discern pines from the rest. Siegel asks you to introspect and infer the best explanation for the difference. Fish's Challenge results from discrepancies between Siegel's and others' testimonies regarding what it is found upon introspection (see e.g. Lyons, 2005). However, Siegel's method of phenomenal contrast seems to depend on more components than the one presented here. To see this, compare the pine-tree contrast to the sinewave contrast. First, concerning the pine-tree contrast, you do not undergo the experience of seeing a pine tree as a pine tree: you imagine the experience. So when asked to introspect, you do not introspect the experience: you introspect an imagination of the experience. Regarding the method presented here, you actually undergo the experience. So, when you introspect, you introspect the experience, not an imagination of the experience.

Returning to Fish's Challenge, it seems to me that there are two ways to unpack it: one trivial, the other not. The trivial account can be expressed as follows: results produced by the phenomenological method should be met with skepticism, so long as this method is based on prior assumptions. I find this account trivial because a method lacking any prior assumptions would be implausible. The non-trivial account can be read in the following terms: the phenomenological method reproduces assumptions that may be avoided, and until they are, the results of the method should be met with skepticism. It seems to me that Siegel's method reproduces assumptions because, at least in part, it depends on imagining undergoing the experience. By avoiding the need to imagine, the method presented here, call it "phenomenological anatomy", arguably eliminates the assumptions involved in Siegel's method and so meets the non-trivial reading of Fish's Challenge.

Brogaard

In explaining the perceptual view of linguistic comprehension, Brogaard (2016) describes reading ambiguous newspaper headlines, "…our expectations at a higher level of processing automatically influence lower-level processing, quickly generating an appearance of the intended meaning" (pp. 12-13). What is exactly meant by "an appearance of the intended meaning" is not clear. Since Brogaard defends the perceptual view, a plausible reading may be that the perceptual appearance of the sentence changes. That said, when I read these headlines, I do not undergo any experience that could be described as a change in the appearance of the sentence. In what follows, I will perform an anatomy of the episode involved in undergoing an experience of understanding an ambiguous newspaper headline taken from Brogaard's paper. Before I begin, I suggest you spend a moment understanding the sentence yourself. Here is the sentence: "Eye drops off the shelf".

If you are like me, this sentence will seem meaningless to you at first; that is to say, you did not undergo a feeling of understanding the content of the sentence, and this is why you probably reread it. If you are like me, upon re-reading the sentence, you had a feeling of understanding the literal content, viz. '(the) eye drops off the shelf'. Furthermore, the feeling exhibited a particular quality, which can be described as a "lack of confidence", and so you read the sentence again. Upon third reading, ‘eye drops’ seemed to form a compound noun easily and was read almost effortlessly alongside ‘off the shelf’. At this moment, you underwent a feeling of understanding the content of the sentence as based upon the (pharmaceutical) context and the pun involving the word ‘drops’; this feeling also seemed to exhibit two (metacognitive) qualities, which can be described as “fluency” – it felt easy to understand – as well as “confidence” – you felt confident you had understood it – and thus you no longer needed to re-read the sentence.

Let us take stock of what this anatomy says about the appearance of meaning. The meaning of the sentence *did* appear, though it did not appear in perception; one might say it appeared in cognition. Similar to how the solution to a math problem appears as the result of deduction, the meaning of the sentence appeared as the result of an awareness of context. Thus, the appearance of meaning does not result from the sensory content of perception, aided by sub-personal, higher-level processing (as Brogaard suggests). Instead, the appearance of meaning results from the cognition involved in resolving linguistic referents in context. So, an appearance of meaning is perhaps better understood as an experience of meaning, whose upshot is the feeling of understanding, which is, as argued above, an example of irreducible cognitive phenomenology.

REFERENCES

Benson, R., Richardson, M., Whalen, D.H., & Lai, S. (2005). Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *Neuroimage*, 31, 342-353.

Brogaard, B. (2016). In defense of hearing meanings. *Synthese*, 1-17. DOI:10.1007/s11229-016-1178-x.

Chudnoff, E. (2015). *Cognitive Phenomenology*. New York: Routledge.

Fish, W. (2013). High-level properties and visual experience. *Philosophical Studies*, 162, 43-55.

Kaplan. D. (1979). On the Logic of Demonstratives. *Journal of Philosophical Logic*, 8 (1), 81-98.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry & H. Wettstein (Eds.), *Themes from Kaplan*. New York: Oxford University Press, 494-563.

Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2004). Neural correlates of switching from auditory to speech perception. *Neuroimage*, 24, 21-33.

Lyons, J. (2005). Perceptual belief and nonexperiential looks. In J. Hawthorne (Ed.), *Philosophical Perspectives, 19: Epistemology*. Oxford: Blackwell.

Möttönen, R., Calvert, G., Jääskeläinen, I., Matthews, P., Thesen, T., Tumainen, J., & Sams, M. (2005). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage*, 30, 563-569.

O’Callaghan, C. (2011). Against Hearing Meanings. *The Philosophical Quarterly*, 61 (245), 783-807.

Prinz, J. (2012). *The Conscious Brain*. New York: Oxford University Press.

Siegel, S. (2007). How Can We Discover the Contents of Experience?. *The Southern Journal of Philosophy*, 45, 127-142.

JOE HIGGINS
University of St. Andrews and University of Stirling (SASP)
jbeh@st-andrews.ac.uk, joseph.higgins@stir.ac.uk

EMBODIED MIND – ENSOCIALLED BODY: NAVIGATING BODILY AND SOCIAL PROCESSES WITHIN ACCOUNTS OF HUMAN COGNITIVE AGENCY[1]

abstract

There is a prevalent tension within recent cognitive scientific accounts of human selfhood in that either bodily processes or social processes are explanatorily favored at the expense of the other. This tension is elucidated by the body-social problem (Kyselo, 2014) and at its heart is ambiguity regarding the body's role within embodied cognitive science. Drawing on a range of phenomenological and empirical insights, I propose that we can avoid the problem by embracing the concept of an ensocialled body, in which all organic bodily processes are simultaneously social processes from the perspective of human cognitive agency.

keywords

selfhood, cognitive science, embodiment, ensocialment, Body-Social problem, enactivism

Phenomenology and Mind, n. 12 - 2017, pp. 228-237
DOI: 10.13128/Phe_Mi-21121
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

In broad terms, embodied cognitive science construes mind as a dynamic phenomenon that depends non-trivially – sometimes constitutively – on an agent's physical body and surrounding world. However, a persistent problem for this broadly construed paradigm has been a lack of clarity over the exact role of "body" for embodied cognitive agents. This lack of clarity has recently been highlighted by the *body-social problem* (Kyselo, 2014), which describes how, on the one hand, we have bodily oriented cognitive science that risks an account of human selfhood[2] that is individualistically confined by the physical boundary of the organic body, thereby downplaying the significance of social processes in the individuation of the self. On the other hand, we have socially oriented cognitive science, which risks prioritizing explanatory focus on constituting social relations to the extent that any notion of bodily individuality seems to be "lost" to supra-individual social organizations. Exactly how one can reconcile these dichotomizing perspectives on selfhood will be the target of this paper.
My claim, presented in section 3, will be that one route to resolving the body-social problem is to understand human selfhood as relying on the concept of an *ensocialled body*, in which all Self-constituting[3] organic bodily processes are simultaneously social processes. This conception implies that there is no ontological separation between "bodily" and "social" processes as far as human selfhood is concerned. Instead, the body is an inherently social phenomenon, such that *both* bodily and social processes are indispensable to the *constitution* of selfhood.[4] This claim will be supported by a range of phenomenological and empirical insights.

[1] I would like to thank Mike Wheeler for valuable comments regarding an earlier version of this paper. I would also like to note that some passages of sections 2 and 3 have been adapted with revision from "Biosocial Selfhood: overcoming the 'body-social problem' within the individuation of the human self" (Higgins, 2017).

[2] Throughout this paper, I will take it as granted that 'human cognitive agency' equates to 'human selfhood'. The two phrases will thus be used interchangeably.

[3] 'Self-constituting' refers to constitution of the human self, rather than referring to an entity's ability to form or compose itself in a uniquely reflexive manner.

[4] The characterisation of *constitution* to which I will adhere is that provided by De Jaegher *et al.* (2010) and subsequently followed by Kyselo (2014). On this view, for a given phenomenon X, "P is a *constitutive element* if P is part of the processes that produce X" (De Jaegher *et al.*, 2010, p. 433). This is contrasted with the notions that "F is a *contextual factor* if variations in F produce variations in X, [and] C is an *enabling condition* if the absence of C prevents X from occurring" (*ibid.*).

2. An Overview of the *Body-Social Problem*

In a recent paper, Miriam Kyselo (2014) describes the body-social problem as the "mutual tension" (p. 1) between *bodily oriented* approaches to selfhood and *socially oriented* approaches to selfhood. Whilst theories of selfhood that fall under the former approach risk designating the body as a medium of conceptual isolation – in a move that is comparable to orthodox cognitive science's conceptual isolation of neural machinery, which the "embodied turn" endeavours to overcome –, theories of selfhood that fall under the latter approach risk losing any notion of 'individuals' to the dynamics of various social organizations (in a manner that will be explained shortly) (*ibid.*). Human selfhood is thus primordially either an individualistic bodily phenomenon or an essentially social phenomenon:

> (i) For bodily oriented approaches, selfhood inheres in the biological body and is essentially independent from the sociocultural world. Such an *embodied self* may be importantly embedded in a scaffolding environment, but resists constitutive dependence on this environment in virtue of the body's status as an organismic body. As Kyselo (2014) explains, "there is nothing social about the organismic or the moving body *per se*" (p. 4); it is, fundamentally, a biological phenomenon. On such a view, the social world amounts to nothing more than a context for the embodied self. In other words, selfhood emerges from biological-organismic autonomy and, as such, is a permanent potential means of isolation from the (non-biologically autonomous) social world.
>
> (ii) For socially oriented approaches, selfhood inheres in social interactions, such that a cognitive agent is intersubjectively dependent. On this view, selfhood emerges through interactions with others or with socionormative organizations; the body becomes a mere context for selfhood (or, at best, an enabling condition). Rather than being fundamentally isolated through one's biological-organismic body, socially oriented approaches propose that the self is existentially "open" to assimilation into relational dynamics that are generated through interactions.

The broad scope of the body-social problem is that (i) and (ii) each respectively obscure important insights into the roles that social and bodily processes play in the individuation of the human self (Kyselo, 2014). If one endorses (i) then one risks paying "lip service" to the social world (*ivi*, p. 4), treating it as a mere causal or contextual contributor to the self, whereas if one endorses (ii) then one risks an analogous disservice to the organismic body. The problem is that once one has acknowledged the "beyond-the-brain" progression of cognitive science – so as to embrace the integral dependence of cognition on worldly features – then one should be committed to appreciating the important insights that *both* bodily- and socially-oriented perspectives have delivered.[5] A failure to do this may engender a philosophically hollow stance towards either bodily or social processes as regards the individuation of the self, or, worse, may render bodily and social processes "mutually exclusive" (Kyselo, 2014, p. 4).

As Kyselo (2014) explains, the objective should be to integrate bodily and social processes without conceptually isolating either set of processes, thereby rejecting tendencies to prejudicially cut the cognitive cake. Contrary to Kyselo (2014), however, I believe that only a constitutive fusion of "body" and "social", such that the body is a fundamentally "ensocialled" phenomenon (see section 3), will satisfactorily overcome the "body-social problem" (see Higgins, 2017 for an expanded discussion of this).

[5] See, for example, De Haan (2010), Steiner & Stewart (2009) and Vygotsky (1986) for views on the socially constituted self, and Gallagher (2005) and Zahavi (2005; 2014) for views on the primordially embodied self.

Before progressing to the solution of the *ensocialled body*, it is worth quickly considering a crystallized version of the body-social problem. This occurs within the framework of biological enactivism, which is a highly influential position within the field of embodied cognitive science; indeed, it is the position that provides the conceptual background to Varela, Thompson and Rosch's (1991) *The Embodied Mind*, which is one of the "bibles" of embodied cognitive science. For biological enactivism, a cognitive agent is one that is *autonomous*, *emergent*, *embodied* and *experiential*, so as to become a *sense-maker* (De Jaegher & Di Paolo, 2007). Each of these criteria is interdependent. So, according to enactive theory, embodiment is necessary for cognitive agents as the physical manifestation of experiential and emergent autonomy, in virtue of which an agent is able to make sense of the world; that is, the body is the animate locus of autonomy through which meaningful activity is created (*ibid*.). On the basis of this simplistic definition that I have provided, it is easy to see how the body may amount to a phenomenon of isolation, in that it is a unique locus of personal autonomy for each and every cognitive agent.

However, it is important to the theory of enactivism that agents can also make sense of the world in a *participatory* manner (*ibid*.). This is captured by the theory of *participatory sense-making,* which claims that when two or more individuals interact with one another their intentional activity can become dynamically coordinated in such a way that a new relational system *with its own autonomy* can emerge between them (*ibid*.; Luhmann, 2002). Such a relational system has its own autonomy in virtue of the mutual regulation between the coordinated behavior of the individuals, which generates and sustains the interaction, and the reciprocal influence of the interactive dynamics on the individuals' behavior. A specific domain of relationality is thus manifest, through which the involved individuals can "sense-make" (i.e. cognize) in a "participatory" manner (De Jaegher & Di Paolo, 2007). The issue is that once we have an autonomous "participatory" organization alongside "individual" autonomy, the body-social problem rears its ugly head. As Kyselo (2014) explains, on one interpretation, (a) individuals are "lost" within the interaction because their "intrinsic purpose" – which is bodily manifest – seems to be directed at the generation and maintenance of relational dynamics, such that the individuals are *heteronomous* with respect to their cognitive activity. These interacting agents are individuated as "constituents" within a social process that has its own autonomous ("group") identity. We thus run the risk of losing grip on a meaningful sense of individuality for each of the interacting agents. On an alternative interpretation, (b) each of the interacting agents "is individuated from others *qua* being *embodied*" (*ivi*, p. 7), such that the individuals' organic bodies manifest a conceptual boundary of delineation. Yet differentiating each individual from the other and from their jointly created interactive dynamics in virtue of their presence as bodily beings condemns the body as "a locus of isolation, not a means of connection and engagement" (*ibid*.). This interpretation thus runs the risk of losing grip on the meaningful influence of the social world.

The question is how should we fairly adjudicate between these seemingly disparate bodily and social autonomies when it comes to the individuation of selfhood?

3. The *Ensocialled* *Body*

One solution to the body-social problem is, I believe, to characterize the body as an *ensocialled* phenomenon. I will initially elucidate this solution through consideration of the pervasive constitution of cognitive agents by social "norms",[6] before clarifying the body's essential role in accomplishing the lived and living performance of these social norms.

[6] Here, and throughout this paper, I am using 'norms' (and 'normativity') in a very broad sense, to capture those principles by which human behavior (and cognition) is deemed appropriate, equating loosely to *ways of being*. In this sense, norms are generally implicit and always socially permeated, such that they simply encapsulate *what one does*, as much as what one *ought* to do.

Firstly, by 'ensocialled' I mean that a bodied agent is always and irrevocably social. Unlike the terms 'socialized' or 'enculturated', which suggest the assimilation or transformation of an agent within a specific socio-cultural domain, 'ensocialled' is intended to convey a fundamental feature of one's existence. That is, an ensocialled agent does not grow *into* the social world, nor can she shed her social nature by living as a hermit; instead, an ensocialled agent is a constitutively social being. This means that, foundationally, the body is a social phenomenon, such that the ongoing generation of one's selfhood is constituted by the active experience of the socially permeated body.

In order to better grasp this idea, consider that the fictional character of Robert Neville (from Matheson's *I Am Legend*, 1954) may be considered "de-socialized", as he is completely isolated from other humans, but is still ensocialled (and always will be) because he continues to live (and think) in accordance with recognizable social norms. That is, his thoughts and behavior cannot *but* be socio-normatively permeated. Think, for instance, of any number of everyday acts: how one carries oneself whilst walking, how loudly one talks, how far one stands from another person during conversation, or how one dresses – all of these are executed through implicit conformance to the intersubjective norms of *what one does*. The myriad occurrences of these norms amounts to a normative "lifeworld" through which human agents inevitably live (cf. Ikäheimo, 2009), such that one conforms to socio-normative standards regardless of the presence of others.[7] Indeed, it is not just overt behavior that is suggestive of such norms; there is also evidence that one's linguistic descriptions will vary in accordance with one's societal conformance (Athanasopoulos *et al*., 2014), as will one's neural responses to perceiving objects (Goh and Park, 2009), to evaluating threats (Park & Kitayama, 2012), and to judging the performances of in- and out-group team members (Molenberghs *et al*., 2012). Further to this line of thought, there is also considerable empirical support for the view that gender, which is commonly taken to be a social construction, produces notable divergences in cognition and behavior; for example, genders adopt differing cognitive strategies when faced with tasks requiring creative or generative responses (Abraham *et al*., 2013; Abraham, 2015), as well as undergoing the activation of differing neural regions during affective experiences (Moriguchi *et al*., 2013), and showing differing sensitivity to physical pain (Wiesenfeld-Hallin, 2015).[8] The enaction of socio-normative conformances thus runs "deeper" than just overt bodily activity, to a point where the experientially lived nature of one's bodily agency – neurally, affectively, linguistically and behaviorally – is implicitly bound to social practices and miens. This holds even if one is an isolated hermit: the very foundations of one's agency remain suffused with social ways of life.[9] One therefore simply cannot avoid the pervasiveness of these social norms as constituents of one's cognitive agency.

[7] "Private" behavior may seem to be an exception here, in that it is often claimed that one may behave differently when alone and behind closed doors. But the very fact that being alone is occasion to act differently is itself telling of social norms of when and where to behave in certain ways. As this section of the paper goes on to explain, both iconoclastic and conformist activities modulate the same normative structures in virtue of what becomes implicitly sedimented as the common disposition for *what one does* in a given situation.

[8] With the issue of gender, physiological differences in sex will have some role to play, but I believe that gender divergence is primarily generated through the *normativity* that pervades social conceptions of female and male roles and capacities.

[9] Of course, over time, a hermit's cognitive and behavioral enactions may stray from those of regularly interacting social groups (to which the hermit previously belonged). But the socio-normative behavior of an ensocialled agent is not merely a socially conferred property; it is an *(en)active* engagement with socially established and maintained domains of normativity, which, in the absence of interactions with others, will persist in virtue of deeply sedimented conformances from the hermit's early life. If the hermit were to re-engage with social peers, s/he would thus bodily present sufficient traces of implicitly recognizable normativity such that s/he could be acquiesced to as harboring a shared normative background.

Indeed, full extraction from such social norms amounts to death or a comatose state in which one can no longer meaningfully enact the social world. This is not to say that one's selfhood does not undergo (potentially significant) changes throughout a lifetime, but that these changes can only ever occur within an existential domain that retains certain socionormative connections to a specific (embodied and ensocialled) lifeworld.

Crucially, the social domains of normativity to which one conforms are not a static phenomenological bedrock, but are modulated via our expressive bodily activity. Bodied engagements – either through observation or direct interaction – are the currency by which social norms are generated, maintained and transformed within and across dyads, cliques, groups, institutions and cultures, with these social norms then determining the very nature of bodied engagements. In terms of interaction dynamics, the idea is that bodies are permeated by social norms and bodily activity thereby discloses the socionormative lifeworld, with generated and modulated social norms then feeding back and canalizing individuals' range of potential bodily actions which will, in turn, generate and modulate further norms which will then instantiate further feedback (and so on). Social norms are thereby reinforced or modified by aggregations of collective *embodied* behaviors. In this way, our bodies are not sites of isolation; rather, they are that through which our ensocialled agency is lived – the animate locus of sociality that characterizes our existence. It is our bodily belonging to – and ongoing modulation of – social norms that canalizes our agential existence, from both behavioral and cognitive perspectives.

In order to clarify the manner in which our bodies escape characterization as isolating phenomena, we can refer back to the exposition of the body-social problem within the enactive theory of participatory sense-making. Recall that the issue is whether cognitive agents should be individuated in accordance with the autonomous social organizations that are generated during interactions, or with the organismic autonomy of their individual bodies (Kyselo, 2014). The response that the concept of ensocialment gives us is that the body is never a purely organic phenomenon as far as cognitive agency is concerned, in that it is always directed towards enacting some social norm that is a pervasive constituent of our agential existence. The idea is that just as two individuals can play *participatory* roles in generating a dyadic relational domain that they communally enact through their expressive embodiment, so, too, can society-wide collectives of individuals be seen as playing *participatory* roles in generating society-wide relational domains of normativity that are enacted across a society through the expressive embodiment of everyday behaviors. The bodily activity of any given individual is therefore normatively laden so as to modulate the myriad socio-normative domains that we participate in. In this way, to move one's organic body is thus not to merely perturb the physical world; it is also a perturbation of the socio-normative world that we enact.

In order to further develop this concept of an ensocialled body that is essential to our cognitive agency, one could claim that our bodies are, in an important sense, "linguistic" entities, in that we are bodily sensitive to the social world in such a way that intersubjective activity is habitually rendered intelligible (Cuffari *et al*., 2014). In virtue of being constituted by pervasive social norms, our bodies are never merely organic phenomena from the perspective of cognitive agency; instead, they are always socially expressive and engaged in the ongoing modulation of the socio-normative "lifeworld" that we all enact. Our bodies thus have a "linguistic" essence, assimilating and dispersing norms through our animate conduct. As Cuffari *et al*. (2014) put it: "world-engagement is an integrated whole of embodied interpreting [...] embedded in horizons of social normativity" (p. 1115). However, it is important that this view of "linguistic" bodies does not downplay the fundamentally *ensocialled* nature of the body that I am espousing. It is not the case that embodiment and ensocialment are separate spheres of agency that are merely "integrated"; rather, embodiment and ensocialment are, from the perspective of human cognitive agency, ontologically dependent on one another. Without the socio-normative constitution of embodied agency and the bodily manifestation of social norms, one simply does not exist as a bona fide human self.

What's more, this ontological dependence of embodiment and ensocialment is evident from the earliest moments of life through to the final moments. There is a plethora of empirical data which suggests that, from birth, humans are potentiated to recognize and interact with other humans, with our bodies being the means by which such social accomplishments are achieved. Within minutes of birth, for instance, neonates display meaningful responsiveness to human interaction, such as imitating facial gestures (Anisfield, 1996; Meltzoff & Moore, 1977) and certain hand movements (Nagy *et al.*, 2005). Such fledgling responsiveness combines with studies showing the preferential attunement of neonates towards face-like configurations (Johnson *et al.*, 1991; Valenza *et al.*, 1996), maternal odor (Macfarlane, 1975) and human speech (Vouloumanos *et al.*, 2010). Preferences of this ilk support the view that there is an inherent human disposition towards parsing the world into human and non-human entities (Meltzoff & Brooks, 2001).[10] On the basis of these nascent abilities to recognize and interact with others, newborns gradually begin to accrue further social capacities such as gaze-following at approximately 9 months of age (Senju *et al.*, 2008), joint attention at 9-14 months (Phillips *et al.*, 1992) and comprehension of goal-directed behavior at 18 months (Meltzoff & Brooks, 2001). The significance of such empirical evidence is to bolster the claim that human bodily activity is inherently socially imbued from the first moments of life, with more complex behavior developing from the foundation of this bodily immersion in the social world.

Naturally, one may still debate the exact point at which a foetus or neonate becomes a bona fide human self. For the ensocialled account that I am putting forward, what matters is that the earliest tentative signs of intelligible activity in neonates are socially predicated. Neonatal activity seems to be inherently disposed towards social interaction and this nascent bodily behavior concurrently generates and modulates the normative lifeworld that a neonate will enact. Even if one wishes to claim that selfhood emerges with basic bodily self-awareness (Gallagher, 2005, pp. 72-85), the ensocialled view would posit that such awareness is simultaneously a kind of social (self-)understanding of one's own presence (and modulatory capacity) within the normative "lifeworld" of human cognitive agency.

Thus, even as neonates, our bodies are ensocialled in that *any* bodily activity has a normative component that is an expressive modulation of possibilities for action for others. Such normativity first takes hold through the nascent capacity to recognize, imitate and interact with others, so that the very bonds of social normativity in which our existence is rooted are bodily manifest at a foundational level and then continuously modulated in bodily expressivity and interactions throughout life.

To close this section, it will help to refer the notion of the ensocialled body back to the body-social problem. The claim is that once the body is rightly conceived of as *ensocialled* there is no longer any tension between the individuating credentials of bodily and social processes because the body is fundamentally a social entity as far as human cognitive agency is concerned. To be a human self is to be an embodied modulator of the socio-normative lifeworld that we each enact.

[10] The Macfarlane (1975) study, amongst others (e.g. DeCasper & Fifer, 1980; Pascalis *et al.*, 1994), suggests that this inherent disposition is perhaps even more refined, such that the social world is parsed on a mother/non-mother basis. This remains supportive of my claim regarding the ontological entwinement of embodiment and ensocialment, with a neonate being bodily attuned to social phenomena through its mother perhaps even prior to birth (see also Marx & Nagy, 2015).

4. Back to Phenomenology?

One may worry that the account of an ensocialled body is merely a reformulation of the established notion of a *lived body* (Husserl, 1931/1960; Merleau-Ponty, 1945/1962). But the account that I have provided differs in important ways from this stalwart concept of phenomenology.

Firstly, the ensocialled body is relevant to all stages of human life, whereas many giants of phenomenology neglect the developmental aspect of our existence (e.g. Husserl, Heidegger, and Sartre). Secondly, the concept of an ensocialled body is not merely addressing one's lived subjectivity in the sense that we live *through* our bodies; rather, it is focusing on one's ensocialled nature such that the body is, inherently, the site of one's lived subjectivity and, simultaneously, the site of intersubjective expression. It is thus not an account of lived bodily subjectivity that is dressed with a distinct intersubjective dimension, but an account in which lived bodily subjectivity emerges within the intersubjective relations that it enacts. In slightly different terminology, the lived ensocialled body is permanently open to modulation by social norms, and will itself transform social norms as a modulating constituent, such that an organismic agent and societal worlds are two sides of the same coin. Thirdly, for many phenomenological and cognitive scientific accounts of the human self, the body has non-trivial status, but is still secondary to transcendental consciousness or socially constructed identity. For these accounts, the organic body is a sensorimotor constituent of the self, essential only insofar as it physically grounds and ratifies one's dynamic existence in the world. The ensocialled body, however, is more than just the sensorimotor aspect of the wider system that is the self; it is the incarnation of our social identity, both a bearer and emissary of bodily-social enacted worlds through which we individuate ourselves. In other words, the ensocialled body is not a mere "medium" for some primordial sense of self, nor is it an isolating site of lived subjectivity – it is the hub of animate agency in which socio-normative selfhood is manifest.

5. Conclusion

In this paper, I have shown how the human body should be viewed as a resolutely *ensocialled* phenomenon as far as the cognitive agency of selfhood is concerned. Whilst there are still various aspects of this account to be further developed, the significance of the claim is that accepting the mind as embodied is only a tentative step towards a full appreciation of the mind – and any self within which the mind inheres – as a genuinely enworlded phenomenon. The body (including the brain) that constitutes the mind is itself constituted by the social norms that permeate our everyday existence. The worlds of social mores, cultural dictions, postural expressions and emotional reserves that are generated and modulated through our social relations are all constitutive contributors to our cognitive praxis, which is physically manifest by our ensocialled bodies.

REFERENCES

Abraham, A. (2015). Gender and creativity: an overview of psychological and neuroscientific literature. *Brain Imaging and Behavior*, 10(2), 609-618.

Abraham, A., Thybusch, K., Pieritz, K. & Hermann, C. (2013). Gender differences in creative thinking: behavioral and fMRI findings. *Brain Imaging and Behavior*, 8(1), 39-51.

Anisfield, M. (1996). Only tongue protrusion modelling is matched by neonates. *Developmental Review*, 16(2), 149-161.

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N. & Thierry, G. (2014). Two Languages, Two Minds: Flexible Cognitive Processing Driven by Language of Operation. *Psychological Science*, 26(4), 518-526.

Cuffari, E., Di Paolo, E. & De Jaegher, H. (2014), From participatory sense-making to language: there and back again. *Phenomenology and the Cognitive Sciences*, 14(4), 1089-1125.

DeCasper, A. & Fifer, W. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174-1176.

DeHaan, S. (2010). Comment: the minimal self is a social self. In T. Fuchs, H. Sattel & P. Henningson (Eds.), *The Embodied Self*. Stuttgart: Schattauer, 12-17.

De Jaegher, H. & Di Paolo, E. (2007), Participatory sense-making: an enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6, 485-507.

De Jaegher, H., Di Paolo, E. & Gallagher, S. (2010), Can Social Interaction Constitute Social Cognition?. *Trends in Cognitive Sciences,* 14(10), 441-447.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Gallagher, S. & Meltzoff, A. (1996). The earliest sense of self and others: Merleau-Ponty and recent developmental studies. *Philosophical Psychology*, 9, 213-236.

Goh, J. and Park, D. (2009). Culture sculpts the perceptual brain. *Progress in Brain Research*, 178, 95-111.

Higgins, J. (2017). Biosocial Selfhood: overcoming the 'body-social problem' in the individuation of the human self. *Phenomenology and the Cognitive Sciences*, DOI: 10.1007/s11097-017-9514-2.

Husserl, E. (1931/1960), *Cartesian Meditations*, (Engl. Transl. by D Cairns). Dordrecht: Kluwer.

Ikäheimo, H. (2009). A vital human need - recognition as inclusion in personhood. *European Journal of Political theory*, 8, 31-45.

Johnson, M., Dziurawiec, S., Ellis, H. & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2), 1-19.

Kyselo, M. (2014). The Body Social: An Enactive Approach to the Self. *Frontiers in Psychology*, 5(986), DOI: 10.3389/fpsyg.2014.00986.

Luhmann, N. (2002/2012). *Introduction to Systems Theory* (Engl. Transl. by P. Gilgen), Cambridge: Polity Press.

Macfarlane, A. (1975). Olfaction in the development of social preferences in the human neonate. *Ciba Foundation Symposium,* 33, 103-117.

Marx, V. & Nagy, E. (2015). Fetal Behavioural Responses to Maternal Voice and Touch. *PLoS One*, 10(6), DOI: 10.1371/journal.pone.0129118, cited 2017.

Matheson, R. (1954). *I Am Legend*. New York: Gold Medal Books.

Meltzoff, A. & Brooks, R. (2001). 'Like me' as a building block for understanding other minds: Bodily acts, attention, and intention. In B. Malle, L. Moses & D. Baldwin (Eds.), *Intentions and Intentionality: Foundations of Social Cognition*. Cambridge: MIT Press, 171-191.

Meltzoff, A. & Moore, M. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, 198, 75-78.

Merleau-Ponty, M. (1945/2012). *Phenomenology of Perception*, (Engl. Transl. by D. Landes). London: Routledge.

Molenburghs, P., Halasz, V., Mattingley, J., Vanman, E. & Cunnington, R. (2012). Seeing is believing: Neural mechanisms of action-perception are biased by team membership. *Human Brain Mapping*, 34(9), 2055-2068.

Moriguchi, Y., Touroutoglou, A., Dickerson, B., & Feldman Barrett, L. (2013). Sex differences in the neural correlates of affective experience. *Social, Cognitive and Affective Neuroscience*, 9(5), 591-600.

Nagy, E., Kompagne, H., Orvos, H., Pal, A., Molnar, P., Jansky, I., Loveland, C. & Bardos, G. (2005). Index finger movement imitation by human neonates: motivation, learning, and left-hand preference. *Pediatric Research,* 58(4), 749-753.

Park, J. & Kitayama, S. (2010). Interdependent selves show face-induced facilitation of error processing: cultural neuroscience of self-threat. *Social Cognitive and Affective Neuroscience*, 9(2), 201-208.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C. & Fabre-Grenet, M. (1994). Mothers face recognition by neonates: a replication and extension. *Infant Behavior and Development*, 18(1), 79-85.

Phillips, W., Baron-Cohen, S. & Rutter, M. (1992). The role of eye contact in goal detection: Evidence from normal infants and children with autism or mental handicap. *Development and Psychopathology*, 4(3), 375-383.

Senju, A., Csibra, G. & Johnson, M. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, 108(2), 303-319.

Steiner, P. and Stewart, J. (2009). From autonomy to heteronomy (and back): The enaction of social life. *Phenomenology and the Cognitive Sciences*, 8(4), 527-550.

Valenza, E., Simion, F., Macchi Cassia, V. & Umilta, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 892-903.

Varela, F., Thompson, E. & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.

Vouloumanos, A., Hauser, M., Werker, J. & Martin, A. (2010). The Tuning of Human Neonates' Preference for Speech. *Child Development*, 81(2), 517-527.

Vygotsky, L. (1986). *Thought and Language*. Cambridge: MIT Press.

Wiesenfeld-Hallin, Z. (2015). Sex differences in pain perception. *Gender Medicine*, 2(3), 137-145.

Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-person Perspective*. Cambridge: MIT Press.

Zahavi, D. (2014). *Self & Other: exploring subjectivity, empathy, and shame*. Oxford: Oxford University Press.

HUGO DE BRITO MACHADO SEGUNDO
Federal University of Ceará (UFC - Brazil)
hugo.segundo@ufc.br

RAQUEL CAVALCANTI RAMOS MACHADO
Federal University of Ceará (UFC - Brazil)
raquelramosmachado@gmail.com

BIOLOGY, JUSTICE AND HUME'S GUILLOTINE

abstract

Biology and Neuroscience are addressing issues related to moral sentiments, but this does not mean that Philosophy has lost its importance in the debate. Paradoxically, the discovery that moral sentiments have evolutionary origins does not overcome the problem of "Hume's Guillotine". There are human characteristics which can be explained by natural selection and that are nonetheless culturally reproved. In order to choose or select which "natural" characteristics are to be promoted and which are to be discouraged, it is necessary to use a criterion that is not given by nature, although human capacities to discuss these criteria have been naturally shaped.

keywords

moral sentiments, evolutionary origins, moral philosophy

Phenomenology and Mind, n. 12 - 2017, pp. 238-246
DOI: 10.13128/Phe_Mi-21122
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness to him, though he derives nothing from it except the pleasure of seeing it (Smith, 1790, p. 4).

1. Introduction

In the last decades, some branches of scientific knowledge have turned their attention to the origins and foundations of moral sentiments (*e.g.,* Ruse, 1986; Waal, 1996; Changeux *et. al.*, 2005; Waal *et. al.*, 2014), a topic that for many centuries was addressed almost exclusively by philosophers. The conclusions of these scientific approaches are surprising. To a great extent, in the end, they converge with what was already considered to be known about the topic, but the fact that these new studies are grounded in empirical experimentation, not in philosophical reflection, seems to provide more solid clues about the origins of such sentiments.

One question then arises: is philosophy still necessary to address this topic? As scientific knowledge regarding human beings and their brain advances, topics such as ethics, aesthetics and epistemology would have more solid or objective answers than those given by philosophy, whose relevance would be dwindling.

This text intends to readdress this question, especially dealing with "Hume's Guillotine" and its possible overcoming from such scientific studies. Its aim is to contribute in determining the role of moral philosophy in present days, addressing the following question: if science explains that *values* have biological origins, is it possible to say that the "is-ought" distinction is now losing its importance, because values (*ought*) derive directly from our biological constitution (*is*)? In this case, is philosophy – that in the past was undoubtedly more capable than science to deal with values – still relevant in this field?

2. Game Theory, Biology, Neuroscience and The Research About Moral and Ethics

Game theory deals with the mathematics of interactions. The study of game theory shows that there are different kinds of interactions, and that there are many variables or factors interfering in these interactions.

If two or more "players" – human beings, wolves, bacteria or even computer programs – are seeking for some result (related to food, money, sex partners, points, clients, or any other scarce resource) in an environment that allows one to gain without others having necessarily to suffer losses in the same proportion, cooperative strategies will naturally arise. That's what game theorists call *non zero-sum game*. But, of course, if most players adopt cooperative behavior, it may be interesting for one to take advantage of others by refraining from cooperating. This also gives rise to mechanisms that are able to protect cooperative individuals (and their group) from such "free riders", so *cooperation* is often followed by *retaliation* when a free rider is identified (Axelrod, 2010; Joyce, 2006).

The struggle for survival is a "non zero-sum game". It is not necessary for one living being to kill all the others to stay alive. On the contrary, many times cooperation is a good strategy, frequently observed among living beings. That's why cooperative behavior has been naturally selected, including mechanisms to identify and deal with free riders. Nature, indeed, has many examples of cooperation, and even of altruistic behavior (Darwin, 1871; Waal, 1996), since altruism may be seen as one of nature's ways of implementing cooperation (Joyce, 2006, p. 17). This may be the reason why social animals, like apes, dolphins, whales and wolves, have more developed and complex brains, which permit high level cooperation. And, for the same reason, moral sentiments were also naturally selected in these animals, especially because they act as a mechanism to prevent and punish free rider behavior (Greene, 2013; Waal, 1996).

In other words, game theory, evolutionary biology and the study of other animals' behavior, all these approaches work like Susan Haack's crossword puzzle (Haack, 1993): different studies of the same reality confirm each other's hypothesis. It also happens with neuroscience, which proposes that human brain is endowed with structures that make us experience unpleasant sensations when observing behaviors or situations harmful to our organism or to the social group we belong to. Disgust is common when someone is in front of blood, wounds, feces, vomit, or other substance that may be harmful if ingested, but also if one sees, *e.g.*, an innocent child being tortured. Languages have even the same words to designate both situations (Kelly, 2011; Joyce, 2006). Not surprisingly, persons with brain damages in some specific areas loose the capability of making moral judgments (Damasio, 2005; Joyce, 2006).

In the same vein, Antonio Damasio points out that values are linked, in their origin, to the *homeostatic equilibrium* necessary for the preservation of life (Damasio, 2005, p. 47). This equilibrium is related to temperature, food satiety, hydration, the normal functioning of the digestive system, and so on, situations invariably linked to feelings of pleasure or pain/discomfort. Thus, of course, situations are seen as something "good" or something "bad", and the mechanism of natural selection, as pointed out earlier, has extended these sensations also to other situations, especially to those related to the homeostasis of the social group, regarding social animals like humans, apes, dolphins etc., giving rise to moral sentiments (Greene, 2013). The same can be said of values such as 'beautiful' and 'ugly'. Not only altruistic behavior and moral sentiments would therefore have a natural origin, but values in general, including those related to aesthetics (Ramachandran, 2011).

3. Hume's Guillotine and The Paradoxical Origin of Morals

One could argue whether a philosophical approach to such questions would still be relevant, given the fact that human values and moral sentiments seem to be biological in their origin. After all, one could defend the necessity of studying such questions from a scientific perspective only, supposedly capable of endowing them with greater objectivity and certainty. However, even if moral values and moral sentiments have biological origins, as they probably do, this does not in itself provide a reason for obeying or promoting them. In other words, the fact that the "sense of justice" is natural from the biological point of view is not sufficient – alone – to justify the "obligatoriness" of the resulting moral duties.

The study of other animals' behavior suggests that they also have, in some way, the awareness of the difference between *is* and *ought*. Primates know that there are social rules they must follow, but they nonetheless eventually disregard them, and are aware of this. Frans de Waal relates, for example, the case of chimpanzees that have sexual intercourse with female partners of the alpha male of their group, expressing concern that the alpha might discover the infringement they committed (Waal *et al.*, 2014). Even so, in human societies, the difference between *is* and *ought* can be placed in a much clearer and striking way. Humans have more complex and developed brains, which permit, to an extent incomparably greater, to imagine different realities and scenarios, futures and possibilities, in order to create a deeper distinction between actual realities (is) and possible ones (ought), in order to make *moral judgements* possible (Joyce, 2006).

On the other hand, the study of biology, neuroscience, and the behavior of other animals, reveals that there are also natural foundations for *e.g.*, preconceptions, xenophobia and racism (Kelly, 2011; Greene, 2013), but, of course, this is not a reason to defend or promote these negative emotions. Feelings, emotions or sentiments associated with empathy and solidarity are usually manifested when the individual relates to those who are considered by them as members of the same group (seen as "equal"); however, in relation to individuals considered as members of other groups (seen as "different") opposite feelings usually emerge (Greene, 2013). If all these emotions or sentiments have, in some way, natural explanation, why some should be promoted, while others should be fought or avoided? In addition, sexual intercourse is naturally pleasurable for reproductive purposes, but nowadays, perhaps, the overwhelming majority of sexual relations are undertaken without this purpose. On the contrary, couples massively use contraceptive methods. This shows that the "natural" purpose of a behavior or sensation can be ignored, and, when acknowledged, it can be culturally modified.

David Hume's warning, according to which it is not possible, from an "is" statement, to extract an "ought" judgment (Hume, 1978) seems to remain current. This does not mean, of course, that one cannot make value judgments considering facts. Indeed, ought judgements cannot be based *only* on "is" statements. Or, in other words, "experience teaches us, to be sure, that something is constituted thus and so, but not that it could not be otherwise" (Kant, 1998, p. 137).

There is, then, a paradox. The human capacity to formulate value judgments has its origins explained by biological facts (an "is"), but the binding of such judgments cannot be justified or grounded only in these facts. This shows that there is still enough space for philosophical considerations on such questions, and philosophy should not be seen as an adversary of science, and vice versa. It is pointless to discuss if science has replaced philosophy, or to argue which of them is "more important". On the contrary, they are different and complementary ways of approaching the same realities, and both should dialogue with each other (Garson, 2015).

Better knowledge of facts, for example, allows the formulation of more appropriate ought judgments, reinforcing them. For example, the judgement according to which a person should not smoke is based on the metaphysical assumption that smokers' and other people's lives should be preserved (an "ought"). Even without any change specifically in this "ought" statement, the judgement could be reinforced if one verifies that tobacco is even more harmful to health than science used to consider (a fact, or an "is"). Or, by the same reason, the judgement would shift to its opposite if science surprisingly discovered that smoking is, in fact, not unhealthy. In other words, although ought judgments cannot be based on facts *alone*, they are made considering facts, so the best knowledge of these is undoubtedly important to such valuations.

It is also possible to raise a provocative question, analogous to what Plato proposes, through the character Socrates, when he inquires, in *Euthyphro*: "is the pious (τὸ ὅσιον) loved by the gods because it is pious, or is it pious because it is loved by the gods?" (10a.). In analogous terms, one might propose: do certain behaviors seem fair to us due to our naturally selected ability to consider them as such, or was the ability to regard such behaviors as fair selected because they are fair, and fairness favors survival? To put it another way: did natural selection and game theory create the notion of "fair behavior", or do they merely provide a means for humans to discover fairness? Science is not able to answer this question alone, without any philosophical help.

4. The Importance of Philosophy in Contemporary Debate

Although it contributes greatly to the clarification of the biological origins of moral sentiments, the scientific approach, as we can see, is incapable of solving a series of questions that remain. This is, basically, due to "Hume's Guillotine", which is not overcome just because our capacity of making moral judgements has evolutionary origins, or because some other animals are also capable of making the distinction between is and ought.

In this context, it is up to philosophical speculation to investigate, for example, the grounds on which feelings of empathy and solidarity must be nurtured, while those of aggression, racism and prejudice must be suppressed. Clarification on the origins of such sentiments in the evolutionary sphere undoubtedly helps this speculation, but they are not to be confused with it.

As we have said, xenophobia, racism and prejudice regarding "different" people may also have biological origins, since, in a very distant past, strangers indeed were often a threat. Our remote ancestors who had curiosity or sympathy for strangers, "outsiders" to their group, may not have lived long enough to leave offspring with the same genes. The contact between distinct human groups, in turn, was much rarer. Therefore, as Greene (2013) explains, our intuitive or even instinctive sentiments lead us to cooperate with people we see as equals, part of our same group (us), but to see as enemies or adversaries people considered as "different" (them).

No longer making a purely factual analysis – although starting from it – one could ask whether such premises, which led to the natural emerging of racist or xenophobic sentiments, are often present nowadays. And the answer is no, so these sentiments are no longer justified, even from a merely biological point of view. Comparatively, it is known that humans have a special preference for sugar and fat, because in a very remote past, in which our digestive system and our food preferences were shaped, such nutrients were decisive for survival. In an environment where food was not always available, and in which much energy was expended to obtain nutriment, the individual satiated with as many high-calorie foods as possible would have a much greater chance of survival, a reality that is no longer present in contemporary societies.

This means that in today's world the ease of obtaining high-calorie foods and the sedentary lifestyle provided by automobiles, elevators and related devices have turned obesity and diabetes into major problems, leading people to restrain their natural appetite for such nutrients.

As for xenophobia, racism and other kinds of prejudice towards different people, it could be said something similar. Also, in the contemporary world, human groups are no longer separated and isolated, and they are not dangerous to each other because of this. This also allows exploring ways of circumventing or departing those natural tendencies of hostility towards people seen as members of "another group".

In the contemporary world, in fact, people are highly interconnected, and their individuality is defined by such a varied range of characteristics that it is impossible to identify them as belonging only to a single social group. One can see another as "different" because of the color of their skin, but as "equal" if considering their religious or ideological beliefs, or their sport preferences. As Amartya Sen (2006) points out, the same person can be

> without any contradiction, an American citizen, of Caribbean origin, with African ancestry, a Christian, a liberal, a woman, a vegetarian, a long-distance runner, a historian, a schoolteacher, a novelist, a feminist, a heterosexual, a believer in gay and lesbian rights, a theater lover, an environmental activist, a tennis fan, a jazz musician, and someone who is deeply committed to the view that there are intelligent beings in outer space with whom it is urgent to talk (preferably in English). Each of these collectives, to all of which this person simultaneously belongs, gives her a particular identity (p. xii).

These are the innumerable "groups" to which this person belongs in contemporary world, and whenever they are developing aggressive feelings against others due to being part of a diverse group, they can remember that, considering different aspect of their individuality, they can be regarded as members of the same group. This is a way to deal with instincts, sentiments, emotions or feelings considered negative, once their causes are known, in order to neutralize their effects. That is another example of the richness provided by the dialogue between science and philosophy on this topic.

As Lukes (2008, p. 252) points out, there is an innate or biological morality, and another perfected by society, and in this social or cultural improvement philosophy has the important role of directing how and in which terms improvements should occur. In similar terms, it is possible to say that natural selection (and perhaps also cultural and sexual selections) gave to human beings the capacity of making moral judgments, and a common and universal core to the content of a few of them, leaving open, however, a wide range of possibilities for determining the content of the others, according to the environment and cultural variances. In R. Joyce's words (2006), no one "would deny that cultural learning plays a central role in determining *the content* of the moral judgments that an individual ends up making; the claim is that there is a specialized innate mechanism (or series of mechanisms) designed to enable this type of learning" (p. 137).

It is also possible to inquire, as pointed out – and this paper is not intended to answer this question, but only to introduce it –, if natural selection created moral sentiments from nothing, or if it merely gave to some animals, and more specifically to humans, mechanisms to allow them to access or to know this supra-sensitive reality that would exist anyway.

To clarify the argument, it may be useful to remember Karl Popper's theory of three worlds (Popper, 1999). According to this theory, reality would be divided into three distinct and related worlds. "World 1" would be that composed of physical particles, or, in other words, by matter. This is the case of a stone, the sun shining, a river running. "World 2", on the other side, would be composed by the result of the brain processes in someone's head, considering not the brain, as a physical organ (part of "World 1"), but the mind, as the result or the effect of brain functioning. A vase of flowers on a table is part of "World 1". The image of this vase, formed in the mind of the one who observes it, integrates "World 2". Finally, there is "World 3", integrated by thoughts and ideas, once detached from the mind of those who formulated them. Hamlet, for example, was one day only part of "World 2" of Shakespeare's consciousness. After being written and divulged, it became part of "World 3", and today, even if this or that book (made of paper and ink, part of "World 1") is destroyed, and even after the death of Shakespeare, such work continues to exist in "World 3", so much that you, dear reader, know who Hamlet is.

There are "World 3" realities which, if all humans died, would disappear as well. This is the case of languages, for example. But in relation to other parts of "World 3", it may be possible to cogitate about their complete autonomy. Even if all people disappeared from the face of the earth, prime numbers will continue to exist, and to be divisible only by one and by themselves, just as the sum of the square of the legs of a straight triangle will remain equal to the square of its hypotenuse.

In this vein, it is possible to ask whether morality cannot be equated with mathematics regarding this specific topic. Natural selection has given human beings, and perhaps some other animals (although to a lesser extent), mechanisms and structures which make them capable of knowing or accessing these parts of "World 3", while not being part of them. In the same way, natural selection gave human beings a brain capable of abstractions, making it possible to know prime numbers and geometrical theorems, whilst not creating or instituting such realities. This is defended by Dworkin (2011), who believes not only in the objectivity of values, but in the fact that they integrate, like mathematics, a diverse – not physical – level of reality, remaining, in this level, *real*. According to Dworkin (2011),

> Hume's principle, properly understood, supports not skepticism about moral truth but rather the independence of morality as a separate department of knowledge with its own standards of inquiry and justification. It requires us to reject the Enlightenment's epistemological code for the moral domain (p. 17).

Regarding to the topic of this paper, Dworkin (2011) also writes that

> Neo-Darwinian theories about the development of moral beliefs and institutions, for instance, are external but no way skeptical. There is no inconsistency in holding the following set of opinions: (1) that a wired-in condemnation of murder had survival value in the ancestral savannahs, (2) that this fact figures in the best explanation why moral condemnation of murder is so widespread across history and cultures, and (3) that it is objectively true that murder is morally wrong. The first two of these claims are anthropological and the third is moral; there can be no conflict in combining the moral with the anthropology or any other biological or social science (p. 35).

There are those who affirm, however, that morality is different from mathematics, because the latter is objective, and the first is not (Joyce, 2006). In the same order of ideas, Ruse (1986) claims that "our morality is a function of our actual human nature and that it cannot be divorced from the contingencies of our evolution. Morality, as we know it, cannot have the necessity or objectivity sought by the Kantian and Rawlsian" (p. 110). And there are even those who claim that not only morality, but even mathematical realities do not exist "in themselves", but only as creations of the human mind, due to natural selection (Dehaene, 2005, p. 145). This last statement can be questioned, nevertheless, at least as far as mathematics is concerned, with the consideration that relations between numbers, and between geometric forms, are independent of the human mind. They are even independent of the existence of an observer, and they are true also in machinery, for example. The fact that reality fits mathematics, which can be argued from the observation of facts, does not mean that human brains created mathematical entities like prime numbers. Indeed, it is also possible to affirm that humans understand or access supra-sensitive mathematical realities from the observation of empirical reality, but the former exist independently of the latter.[1] The same, perhaps, can also be said regarding Justice and morality, and even if we do not go into this subject here, this discussion is sufficient to ground the claim according to which philosophy still has a wide field on which it can contribute to the analysis of many questions related to the topic.

Moreover, it cannot be said that only things that exist independently of all human beings are "objective". One must differentiate, as John Searle does (2004), ontological objectivity, and epistemic objectivity. In the first meaning, we identify as "objective" entities that exist independently of subjects able to observe them. This is the case of stars, rocks, rivers, lions etc. In the second meaning, we can say that something is "objective" because of the possibility of making statements about it, which are independent of the personal preferences of those who

[1] To think that entities as numbers or equations only exist "inside" the brain seems, paradoxically, to incur in a revisited version of idealism. That is to say, in order to dismiss philosophy for a more objective and scientific analysis of reality, one returns to the philosophical conception according to which reality is entirely constituted by the human mind itself.

make them. That is the case of the claim 'Velázquez is a Spanish painter'. Spain, as a national state, is an *institutional reality*, created as a human convention. If all human beings disappeared, 'Spain' would no longer exist, but this does not remove the objectivity of the statement. In this meaning, objective statements can be made about literature, law, money, games, and an infinity of other institutional realities. The same could be said about the objectivity of morality, even if considered as a human creation (or an "illusion") based on evolutionarily foundations. Again, this shows that Philosophy – not just biology – has a broad spectrum of research in the field, especially because it is possible to discuss about realities from "World 3" independently from the correspondence of these realities in "World 1". Therefore, one can discuss the notion of 'progress' in ethics, as a criterion to "salvage the notion of 'objectively better than' that occurs in these claims and counterclaims" about the topic (Kitcher, 2011, p. 210). And even if one accepts that ethical values are indeed emotional and subjective, it is also possible to debate the conclusions that could be drawn from this premise, as does Kelsen, who bases the very necessity of democracy on the axiological relativism in which he believed (Kelsen, 2013).

In a way or another, scientific findings about morality and values have the great merit of showing that Justice has no necessary relation to religion, nor it is a pure and abstract construction of reason, or a mere product of culture, since it is a consequence of moral sentiments that are prior to all of them.

5. Concluding Remarks

The scientific study of values, moral sentiments, and the sense of justice, in the fields of biology, neuroscience and game theory, should not be seen as an opponent of philosophical speculation, which would have been overcome by it. On the contrary, such visions complement each other, providing mutual contributions for a better understanding of the same realities.

It is still not possible to base value judgments on factual or descriptive statements. But it is indispensable to know the facts, their origin and the reason why they emerged, in order to judge them better. That is why the identification of the motives behind moral sentiments, as well as the biological origin of prejudice and xenophobia, helps in the justification – which is philosophical – of the motives why some of them must be unfolded and enlarged, and others suppressed and combated, and provides clues on how to avoid or minimize human tendencies that, while natural, are nevertheless seen as undesirable.

REFERENCES

Changeux, J.P., Damasio, A.R., Singer, W., & Christen, Y. (Eds.) (2005). *Neurobiology of Human Values*. Heidelberg: Springer.

Damasio, A. (2005). The neurobiological grounding of human values. In J.P. Changeux, A.R. Damasio, W. Singer, Y. Christen (Eds.), *Neurobiology of Human Values*. Heidelberg: Springer, 47-56.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. New York: Princeton University Press.

Dehaene, S. (2005). How a primate brain come to know some mathematical truths. In J.P. Changeux, A.R. Damasio, W. Singer, Y. Christen (Eds.), *Neurobiology of Human Values*. Heidelberg: Springer, 143-155.

Dworkin, R. (2011). *Justice for Hedgehogs*. Cambridge: Harvard University Press.

Garson, J. (2015). *The biological mind. A philosophical introduction*. New York: Routledge.

Greene, J. (2013). *Moral tribes*. New York: Penguin Press.

Haack, S. (1993). *Evidence and Inquiry: towards reconstruction in epistemology*. Cambridge: Blackwell.

Hume, D. (1978). *Treatise of human nature.* London: Oxford.
Joyce, R (2006). *The evolution of morality*. Cambridge: MIT Press.
Kant, I. (1998). *Critique of pure reason* (Eng. Transl. by P. Guyer and A.W. Wood). Cambridge: Cambridge University Press.
Kelly, D. (2011). *Yuck! The Nature and Moral Significance of Disgust*. Cambridge: MIT Press.
Kelsen, H. (2013). *The essence and value of democracy* (Eng. Transl. by B. Graf). Maryland: Rowman & Littlefield Publishers.
Kitcher, P. (2011). *The Ethical Project.* Cambridge: Harvard University Press.
Lukes, S. (2008). *Moral relativism.* New York: Picador.
Popper, K. (1999). *All life is problem solving* (Eng. Transl. by P. Camiller). London: Routledge.
Ramachandran, V.S. (2011). *The tell tale brain: a neuroscientist's quest for what makes us human*. New York: W.W. Norton & Company.
Ruse, M. (1986). Evolutionary ethics: a phoenix arisen. *Zygon*, 21(1), 95-112.
Searle, J. (2004). *Freedom and Neurobiology: Reflections on Free Will, Language, and Political Power* (3rd ed.). New York: Columbia University Press.
Sen, A. (2006). *Identity and violence.* New York: W.W. Norton & Company.
Smith, A. (1790). *The theory of moral sentiments.* London: A. Millar.
Waal, F. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge: Harvard University Press.
de Waal, F., Churchland, P., Smith, P., Pievani, T., Parmigiani, S. (Eds.). (2014). *Evolved morality. The biology and philosophy of human conscience*. Boston: Brill.
Zimmerman, A. (2010). *Moral Epistemology*. New York: Routledge.

RYAN ADAMS
Franciscan University of Steubenville
radams003@student.franciscan.edu

ON SOLIDARITY: GRAMSCI'S OBJECTIVITY AS A CORRECTIVE TO BUBER'S I-IT

abstract

I and Thou sets out a dichotomy of human interactions between the merely objective I-It and the intense intersubjective relationship of the I-Thou, creating a problem of how one is to differentiate the I-It relations that are healthy and natural, and those that are limiting and detrimental. As a corrective to this ambiguity, I posit the principle of solidarity as a relation which retains the personhood of the Other yet still confines it to what Buber calls the I-It "relation". To do this I will discuss similar attitudes such as sympathy and camaraderie using them to draw out the meaning of solidarity in contradistinction to them, and show how solidarity functions as Gramsci's Objectivity.

keywords

Martin Buber, Antonio Gramsci, intersubjectivity, objectivity, solidarity

Phenomenology and Mind, n. 12 - 2017, pp. 248-255
DOI: 10.13128/Phe_Mi-21123
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

ON SOLIDARITY: GRAMSCI'S OBJECTIVITY AS A CORRECTIVE TO BUBER'S I-IT

Rather the house of man about which he is concerned now stands between houses, between neighboring houses, between the houses of his neighbors (Buber, 2002, pp. 94-95).

1. Introduction

Eric Mohr points out in his work "Mixing Fire and Water: A Critical Phenomenology" (2016) that the social role of phenomenology will be a key feature to the future of phenomenological research. Mohr seeks to accomplish this by bringing together the social analysis of Critical Theory, with the phenomenological method of Max Scheler, into a "critical phenomenology". As the field of critical phenomenology grows[1] it will be necessary to consider more in depth the notions of community and solidarity from within the perspective of phenomenology. Key to a social understanding of phenomenology is the personalist phenomenological tradition, which Mohr obviously appreciates as evidenced by his reliance on Scheler. However, if one is to understand the particularly social aspects of phenomenology, it is important to return to their dialogical roots. As such, this paper will seek to explore the value as well as the limitations of Buber's account of community and social interaction. In this way, this paper is also a work of Critical Phenomenology, working to bring into dialogue the Dialogical analysis of Buber with the social theory of Gramsci, and to use the interaction of their thought as a way of investigating the idea of solidarity and community.

In *I and Thou*, Martin Buber sets out a dichotomy of human interactions between the merely objective view of the I-It, which calculates and measures, and the intense intersubjective relationship of the I-Thou. This distinction creates a problem of how one is to differentiate the various sorts of I-It "relations". Some I-It relations are clearly healthy and natural, while others are limiting and detrimental to human interaction. As a corrective measure to the ambiguity of Buber's work, I posit, over the course of this paper, the principle of solidarity as a non-I-Thou relation which retains the dignity and personhood of the Other in a way that still confines it to what Buber calls the I-It "relation". To do this, I will discuss similar attitudes such as sympathy and camaraderie, and use them to draw out the meaning of solidarity in contradistinction to them. In drawing out solidarity, I will also address the meaning of community, with particular attention paid to Buber's understanding of community.

Buber's distinction between the living dynamic existence of community and merely useful organization of collective, which he derives from the It/Thou distinction, provides the basis

[1] Evidenced by the inauguration of the new *Journal of Critical Phenomenology* this year.

for a strong critique of ideological thought (Buber, 2002, p. 35). In this way, Buber's dialogical thought is always a careful and helpful resource in avoiding a kind of social or personal dogmatism which hinders discourse and continued thought. His rejection of over-systematic thought allows for his analysis to dig into the fabric of individual experiences and arrive at conclusions as they are found, rather than speculated about. In this way, Buber's work serves as a means to avoid ever being too comfortable with one's thought.

The interpersonal nature of the relationship of solidarity seems to throw a wrench into Buber's program, since for Buber it is clear that the I-It is not properly a "relation"; instead, the I-It is fundamentally a material interaction, rather than an interpersonal one. I will view Buber through the lens of the rejection of materialism by Antonio Gramsci, as well as view Gramsci through the lens of Buber's discussion of community and his rejection of Marxism. In Gramsci's revised view of the relationship between man and the material, which requires a community with others to develop the ways of relating to the world, there is now room for solidarity in even the material interactions of mankind. Through all of this, I will attempt to address the ambiguity of the I-It and posit solidarity as a corrective.

2. The Ambiguity of the I-It

I and Thou (Buber, 1970) creates a fruitful lens for analyzing personal interaction with the world as a whole and with fellow-persons in particular. Buber starts by saying that the "basic word I-It can never be spoken with one's whole being" (1970, p. 54). To begin in this way is central to the claim of the book that the I-It is not entirely human. Were this dichotomy correct, then it would be obvious from the language Buber uses to describe the two that the I-It is sub-human, or at the very least sub-personal, in the fact that the I-It does not allow one to speak with one's "whole being". The language Buber uses to continue his discussion of the I-It would seem to back up this idea. He says that in the I-It "man goes over the surface of things and experiences them" (1970, p. 55).

Thus taken, the I-Thou is far superior. It was even contended by Buber's contemporary Rosenstock-Hussey (1958) that his dichotomy of It/Thou made his work a kind of Gnosticism and sought to bring the It into the I-Thou relation itself to avoid this problem (Theunissen, 1984, p. 265). Yet, Buber states that "I-It does not come from evil" (1970, p. 95). In this, Buber makes it clear that the distinction between the I-It and the I-Thou is not necessarily a moral one. Still, if one considers all the morally deficient ways of relating to another person, it's clear that they fit within the I-It framework. Certainly, one cannot have a relationship with another person wherein they treat the Other as a mere means, or an object of pleasure, within the relationship of the I-Thou, with all its unspeakable reverence for the centrality of the personhood of the Other. So, while the I-It is not necessarily an immoral mode of relating to the Other, certainly all morally deficient modes of relating to the Other are *necessarily* I-It "relations".[2] This creates a baffling ambiguity in Buber's work. There is a difference between the I-It relations which are morally objectionable, and those that are "natural" in Buber's account. Yet, his work seems to leave this question completely unaddressed.

If the Ego is the only possible "I" of the I-It, then it seems that Buber has made a very particular and strong condemnation of the I-It mode of being. How then does one maintain that the I-It does not arise from evil? It seems unreasonable to think that such a morally questionable, if not outright immoral, way of relating to the Other could possibly belong to a "basic word" which is not in some way also morally bankrupt. All the same, Buber still contends quite vehemently that there is nothing immoral about the I-It, and that it is a basic

[2] I hesitate to use the term 'relation' regarding I-It, because it's clear that for Buber only the I-Thou is a "relation". However, for lack of a better term 'relation' will have to be used here.

and natural way of relating to the world around us. He even goes so far as to say that "without It a human being cannot live" (1970, p. 85). Leaving the moral question unanswered, Buber seems content with holding fast to the notion that there's nothing wrong with the I-It.

There is another problem which arises from Buber's description of the I-It/I-Thou make-up of human experiences of the Other. Not only must there be some specific part of the I-It which is not immoral (since Buber is careful to make clear that it is not), there must also be a specific part of the spectrum of I-It that forms the basis for I-Thou relationships. While the distinct ontological primacy of the I-It cannot be established by Buber's work, it would seem natural enough to assume that there is a certain temporal primacy to the I-It, as that would be the most common initial reaction to Other persons. Even if one does not hold to the temporal primacy of the I-It, thus making the I-Thou entirely primary, there still has to be a sort of I-It which allows for the oscillation back into the I-Thou. Buber offers nothing by way of examination into what it is in the I-It which might allow the I-Thou to be achieved. It seems perfectly reasonable to contend that for Buber there is a sort of I-It which precludes the I-Thou and stands as an impediment to the development of that real relationship; however, there must be a kind of I-It which is not an affront to the dignity of the person, from the vantage-point of which one may have the I-Thou.

For this problem, there can be only one solution: that there are distinctions within the I-It relationship which Buber has, for whatever reason, ignored altogether. While I will not set out here to discover all the distinct variants of the I-It, I will attempt to understand one possible variant. This variant, I would argue, is the basis for the growth of the I-Thou, or at the very least, should one deny any possible primacy of the I-It, the ability to oscillate successfully between the two. In this role, I would suggest the specific relationship of Solidarity as the most fitting candidate.

3. Buber and Community

In Buber's work, the notion of community is described in very simple terms, as he claims that "a people is community to the extent that it is communally disposed" (1967, p. 67). At first glance this does very little to elucidate the meaning of the term in Buber's usage, but it provides a good outline for Buber's thought on community, that a community is formed whenever the members of the group in question begins to think, actively, of themselves as a community, to reflect from the depths on that which grounds it (Buber, 1967, p. 211). This is what separates the ideas of collective and community in Buber's work: a collective is simply an organization ordered toward a goal, and a community is an organism "struggling for its own reality as a community" (Buber, 2002, p. 35).

Understood in this way, it is easy to see why Buber bases his understanding of community in the idea of "mutuality" (Theunissen, 1984, pp. 274-278). Community requires the intentional participation of all members. Mutuality seems to make community a dialogical venture, in the way that Buber claims it to be. Thus taken, the community is an I-Thou relationship, while the collective is an I-It. This need not mean that the collective (working with others towards a goal) is bad, but that it is insufficient, just as the I-It is insufficient in the interpersonal sphere generally. The I-Thou requirement for community means that there either must be an I-Thou interaction between each of the members, which would limit communities down to incredibly small groups, or else it must be an I-Thou relationship between the individual and the community itself. The communal "Thou" aspect in Buber's thought is part of what leads Susser to claim that Buber is best understood within the Volkish Tradition (Susser, 1977, pp. 75-95). While it may be true that there is a particular and strong type of community which relates to itself as an I-Thou, there must be a kind of base-level of interaction which forms the groundwork for this type of relationship to the community. It is this type of community interaction that I want to call solidarity.

4. Solidarity

Solidarity must be a kind of I-It attitude. This seems counterintuitive, since solidarity is always solidarity-with, yet it would be inappropriate to claim that it belongs, in Buber's language, to the I-Thou. Certainly, solidarity requires a recognition of value which seems out of place within the I-It as Buber explains it, but its universality and the fact that it requires no truly "direct" contact with the Other makes it an unfit candidate for the category of the I-Thou. As such, it must be an I-It relation, yet its particular moral and interpersonal content makes it a much more interesting case for the notion of the I-It than the ones outlined in Buber's work. In order to really understand solidarity's relationship with the I-Thou, and its place in the I-It, it will be necessary to give some attention to the notion of "solidarity".

David Heyd's description of solidarity is that it is a "local, partial and reflective emotion" (2015, pp. 55-64). Throughout his analysis, he is careful to make a sharp distinction between solidarity, justice, and sympathy. All too often these concepts are confused. It is a fair confusion, since solidarity is often the way in which one works towards justice, and it is reasonable to feel a sort of sympathy for those who are oppressed by injustice. All the same, they must be kept definitively separate as ideas. Solidarity cannot be said to be a sort of sympathy, since it does not require a "feeling" of the same emotional/mental state as the one with whom you feel solidarity. One expresses solidarity most impressively when one is not the victim of any oppression, but is still willing to act in solidarity with one who is being oppressed. However, there is still a sense in which solidarity arises out of a recognition of the unity of persons, such that in a certain respect one expresses solidarity because one knows that the struggles of the Other affect them. This is the basis for the old union mantra "an injury to one is an injury to all".

Despite the profound insights of Heyd, there is still a problem in his description of solidarity. Heyd contends that solidarity is an emotion. This seems to be the prevailing view among contemporary research on the topic. Certainly, this has a particular reasonability, in that one may without a doubt say that they "feel solidarity", but I would suggest that this is not a true description of solidarity. Instead, it seems that the feeling described as "solidarity" is rather the emotion of camaraderie. Solidarity is rather the willingness to sacrifice for the perceived community. In this regard, Karol Wojtyla's work *The Acting Person* provides an exquisite description of solidarity in much the same terms, saying "'Solidarity' means a constant readiness to accept and to realize one's share in the community because of one's membership within that particular community" (1979, p. 285).

Another problem in Heyd's description of solidarity is that of its apparent partiality. It is true that there is something class-based about solidarity, but that does not mean that it is necessarily "partial", and especially not necessarily antagonistic. The community based nature of solidarity, the sense of responsibility, lends itself to the notion of antagonism within any social or economic system which is based on antagonism. Thus, solidarity in oppressed communities becomes clearly antagonistic. Yet, an interesting feature of solidarity is that it need not be antagonistic.

Heyd does not seem to understand the difference between particular expressions of solidarity, strikes, boycotts, and so on, and solidarity itself. Solidarity, if it is to mean anything more than an emotional state of experiencing camaraderie, must be present before the expression of solidarity. A worker does not go on strike in a sudden fit of solidarity towards his fellow workers, but because the commitment of solidarity is already present between them. This however, would seem to create something of a problem in my thesis that solidarity is a form of the I-It "basic word". The basic words, in Buber's usage, are pre-ethical relations to the world. There is something more ontological than just an ethical commitment to the notion of solidarity, in that its basis is community. To explain the way in which the community is the ground-work of the I-It, I will need to address the works of Antonio Gramsci.

5. Gramsci, Objectivity, and Buber

While it may seem that the more abstract or metaphysical notion of materialism is not pertinent to the deeply personal and intersubjective concept of the "basic words" in Buber, it is worth noting that for Buber, material is the basis of the I-It. Buber's contention that the I-It is of the material world would seem to disarm any attempt to claim solidarity, a necessarily intersubjective relation, is a form of the I-It "relation". However, that is only within the particularly reductionistic account of material reality as understood by the doctrinaire materialists. As such, it is worthwhile to consider the ways in which one may approach the notion of material being, to understand the ways in which one may rightfully approach the I-It.

Gramsci's view of the material dialectics within Marx's writings is peculiar (Kearney, 1993, pp. 171-173). For Gramsci, this conception of the interaction between economics and society is not only too simplistic, it cannot even rightly be called a "dialectic" in any meaningful sense. Thus, Gramsci rejects this notion of materialism, claiming it to be little more than fatalism, or even a bizarre theism wherein the material world is a sort of pantheistic "god", whose laws are followed absolutely, and which reduces human decision and community to nothing more than deterministic laws of physics playing out in a specific way (1971, p. 665).

In rejecting this doctrinaire materialist view, Gramsci is faced with a problem: he must maintain the material dialectic of Marx. He wishes to keep the two parts of the dialectic, the base and the superstructure, suspended in relation, and not allow either one of them to totally overtake the other. Were Gramsci to insist on idealism in the strict sense, that would invalidate the dialectic. Instead of this simple assertion, Gramsci makes a much more complicated and nuanced claim. Culture is, for Gramsci, the grounding of the dialectical relation between the base and the superstructure in the Marxian material dialectic. This may seem strange since in common Marxian analysis the culture is distinctly part of the superstructure, but for Gramsci, the culture is not merely a product of the economic structures, but is instead a functional part of the creation of those economic structures and the philosophical and social entities which form the "superstructure" of the dialectic (1971, pp. 376-379).

The exact features of this interplay come from a complicated sociological analysis, and would not be entirely appropriate to this paper, however it suffices for now to say that Gramsci makes the addition of a heavy emphasis on culture to the Marxist view of politics (1971, pp. 765-766). Not that this is absent from Marx or from the Marxist tradition prior to Gramsci, but that in his work, it takes on the additional role of being the objective foundation of community. Taken as such, Gramsci's view is that the material, objective basis of a community is fundamentally an intersubjective reality. Thus, the solidarity, the recognition of oneself as a part of a community and the willingness to sacrifice for that community is itself also an intersubjective, but material, objective reality.

Buber was certainly not a Marxist; he rejects Marx outright saying that he does "not believe in Marx's 'gestation' of the new form of society" (2002, p. 92). It is, in fact, the non-personal element of Marx's view which he rejects more clearly. The notion that the world is simply functioning apart from human activity, he rejects and labels as *Apocalyptic* to which he opposes his own *Prophetic* view, that the new form of society must be brought about by human action, and not in spite of it (Silberstein, 1990, pp. 188-189). He is unabashed in claiming himself as a "Utopian Socialist", a term of scorn in Marxist circles (Buber, 1958, p. 10). However, unlike the bulk of the "utopian socialists" whom Marx criticizes who were called such for their apparent refusal to interact with the given conditions of the world, Buber is clear that "we must be quite unromantic, and, living wholly in the present, out of the recalcitrant material of our own day in history, fashion a true community" (Buber, 1958, p. 15).

This would seem to make the interaction between Buber and Gramsci's thought somewhat more outlandish, but this need not be the case. Buber's "utopian" and "volkish" vision of socialist revolution may be understood as a socialism which appreciates the central importance of social institutions and cultural cohesion as the force for organization (Susser, 1977, pp. 88-90). Thus, in his rejection of Marxism, he is upholding the social, rather than the political, as the central force for organization (Buber, 1958, p. 82). While Gramsci is still a Marxist, his rejection of doctrinaire materialism puts him in a very similar position to that of Buber. Instead of simply focusing on political action and armed conflict, Gramsci's interpersonal view of the material dialectic makes his analysis similarly focused on the social center and on the institutions of a community as the "Prophetic" center of social change. Considered in Gramsci's way, the objective, or material, is a dialectic between the rote material of the world, in connection to a large, complex series of intersubjective relations. These relations are intersubjective in that they are produced by no single individual on their own, but only through the interaction with the ideas of the society as a whole. In this way, Gramsci's view of the dialectic requires a sense of solidarity in the way that I have explained it order to maintain its structural integrity. Solidarity, in Wojtyla's language, is the recognition of the community, and the willingness to do what is necessary for the good of the community. Culture requires this internalized notion of community as well. Thus, for there to be any "objectivity" in Gramsci's usage, there must be a conception of oneself as belonging to a community. Here, there is a distinct case of an I-It "relation" which is both entirely objective and open to the personhood of other human beings. More than simply open to it, the personhood of the Other is entirely necessary for this view. Without it, there would be no way to formulate a notion of community, in which the specific content of an entity in Gramsci's analysis can express itself.

6. Conclusion

It can thus be seen that solidarity is an acceptable corrective to the ambiguity of the I-It in Buber's work. Solidarity shows itself as a separate experience than that of emotion or of sympathy, but rather as a recognition of the existence of a community, of one's place within it, and the readiness to sacrifice for that community. However, this highly personal idea of solidarity seemed to have clashed with the contention of Buber that the I-It is primarily material. Through Gramsci, it can be seen that even the most mundane and material of objects is related to any human being through a complicated system of intersubjective relations which make up the "community", or what he calls "objectivity". Even though Buber and Gramsci do not ultimately agree on the subject of Marx, they do both still retain the core idea that it is the social and communal element which makes up the change in society. This agreement means that for both Buber and Gramsci, the material, objective basis of the community is in the common social features. Solidarity has been distinguished from similar emotions such as camaraderie, as well as from the I-Thou in Buber. Unlike the critique of Rosenstock-Hussey, this view retains the distinction and stability of the I-It/I-Thou relations as Buber explains them. Thus taken, solidarity makes itself distinct as an I-It relation which is open to, and even in some ways reliant upon, the deep interpersonal relation of the I-Thou. Solidarity, distinct from the obvious negative Egoism that the I-It can take the form of, stands as a morally beneficial way of "relating" to the Other in a way which does not require the same overwhelming singularity of the I-Thou.

REFERENCES

Buber, M. (1956). *The Writings of Martin Buber*,W. Herberg (Ed.). New York: Meridian Books.
Buber, M. (1958). *Paths in Utopia* (Engl. Transl. by R.F.C. Hull). Boston, MA: Beacon Press.
Buber, M. (1967). *A Believing Humanism* (Engl. Transl. by M. Friedman). New York: Simon and Schuster.
Buber, M. (1970). *I and Thou* (Engl. Transl. by W. Kaufmann). New York: Touchstone.

Buber, M. (2002). *Between Man and Man* (Engl. Transl. by R.G. Smith). London: Routledge.
Gramsci, A. (1971). *Selections from the Prison Notebooks*, Q. Hoare & G. Nowell Smith (Eds.). London: Lawrence & Wishart.
Heyd, D. (2015). Solidarity: A Local, Partial and Reflective Emotion, *Diametros*, 43, 55-64.
Kearny, R. (1993). *Modern Movements in European Philosophy*, (2nd ed.). Manchester: Manchester University Press.
Mohr, E. (2016). Mixing Fire and Water: A Critical Phenomenology, in J.E. Hackett & J.A. Simmons (Eds.), *Phenomenology for the Twenty-First Century*. London: Palgrave Macmillan, 97-116.
Rosenstock-Hussey, E. (1958). *Das Geheimnis der Universitat Aufsatze und Reden aus den Jahren 1950 bis 1957*, G. Muller (Ed.). Stuttgart: W. Kohlhammer-Verlag.
Theunissen, M. (1984). *The Other: Studies in the Social Ontology of Husserl, Heidegger, Sartre, and Buber* (Engl. Transl. by C. Macann). Cambridge, MA: MIT Press.
Susser, B. (1977). Ideological Multivalence: Martin Buber and the German Volkish Tradition. *Politicl Theory*, 5(1), 75-96.
Wojtyla, K. (1979). *The Acting Person* (Engl. Transl. by A. Potocki). Dordrecht: D. Reidel.

CORRADO CLAVERINI
Vita-Salute San Raffaele University
c.claverini@studenti.unisr.it

THE ITALIAN "DIFFERENCE". PHILOSOPHY BETWEEN OLD AND NEW TENDENCIES IN CONTEMPORARY ITALY

abstract

Back in vogue today is the tendency of Italian philosophy toward reflection on itself that has always characterized an important part of our historiographical tradition. The present essay firstly analyzes the various interpretative positions in respect to the legitimacy, the risks, and the benefits of such a discourse, which intends to distinguish the different traditions of thought by resorting to a criterion of territorial or national kind. Secondly, the essay examines diverse paradigms that identify – in "precursory genius"; in ethical and civil vocation; and in "living thought" – the distinctive hallmark of the Italian philosophical tradition from the Renaissance to today.

keywords

Italian philosophy, Roberto Esposito, living thought, Eugenio Garin, civil engagement

Phenomenology and Mind, n. 12 - 2017, pp. 256-262
DOI: 10.13128/Phe_Mi-21124
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Where Is Italian Philosophy Heading?

Thirty years ago, many answers were given to the question at the center of a book edited by Jader Jacobelli: "where – if anywhere at all – is Italian philosophy going?" (Jacobelli, 1986, p. VI, my translation). That was in 1986 and many things have changed since then: the Berlin wall has fallen, and we have to deal with an ever more globalized world in which there are those who have proclaimed the "end of history" (Fukuyama, 1992), others a "clash of civilizations" (Huntington, 1996); those who speak of an "age of sad passions" (Benasayag & Schmit, 2003), or who ask: "what happened to the future?" (Augé, 2008). In this scenario, we should definitely hazard some new answers to Jacobelli's question, although it would not be strange for some to contest the legitimacy of the question itself. It is necessary not so much to ask where – and indeed if – Italian philosophy "is going" but rather "does it make sense to speak of an Italian philosophy at all?". In effect, since especially the Second World War, Jacobelli's question has been central in numerous publications on the current state of Italian philosophy and the character of contemporary philosophy in Italy. However, in recent years, although there are still books asking searching questions on the state of health of, and the most fruitful areas in, Italian philosophical research, there has been, *rightly*, more caution and attention paid to the preliminary question that should always be kept in mind when referring to philosophy in terms of nationality: is it legitimate to talk about an Italian philosophy? Does it make sense? Or, when it comes to philosophy, should you avoid making distinctions on a national basis?[1]

2. Roberto Esposito and the *Italian Theory*

These questions are at the center of a book that must be credited with having revived the discussion on the issue of nationality of philosophy: *Pensiero vivente* (Esposito, 2010). Esposito's book about the origin and relevance of an Italian philosophy responds positively to the question of the legitimacy of a discourse on that philosophy, where the adjective 'Italian' does not refer to the state or the nation, but to the Italian *territory*. In fact, according to Esposito, on the one hand, Italy has not taken part in the constitution process of modern nation states that affected early modern Europe (in particular France, England and Spain) and, second, neither the Italian people nor intellectuals have ever had a national consciousness. The numerous patriotic appeals of authors such as Dante, Petrarch and Machiavelli up to Foscolo, Manzoni, Mazzini and Gioberti have a purely rhetorical and literary character (*ivi*, p. 19). In this light, we might add the fact that Italian intellectuals have been in the main cosmopolitans,

[1] I have also addressed related issues elsewhere (Claverini, 2016).

as shown from the beginnings of Italian philosophy (in particular, during Scholasticism and Renaissance humanism). Therefore, when Esposito uses the adjective 'Italian' in reference to the philosophical culture produced in Italy, he means something different from both the state and the nation, that is, "a set of environmental, linguistic, 'tonal' characteristics connoting a specific mode that is unmistakable when compared to other styles of thought" (*ivi*, p. 12). This set of characteristics is what Esposito calls "territory", a geophilosophical concept that does not so much refer to "a specific geographical area" (*ibidem*), but emphasizes the movement of "deterritorialization" and "reterritorialization" that has often characterized philosophy (not only Italian). Twentieth-century European philosophy is a clear demonstration of this movement: see, for example, the "deterritorialization" of German philosophy at the time of Nazism and its "reterritorialization" in the United States. However, if philosophers such as Adorno, Horkheimer and Marcuse were forced to emigrate for political reasons, since the sixties there has been another movement of "deterritorialization" (spontaneous, this time). In fact, since 1966, the year of a famous conference organized by the John Hopkins University – *The languages of criticism and the science of man* –, many French philosophers and intellectuals have been called to teach or to participate in conferences in the United States. Similarly, in recent years, again in the United States, Esposito has detected a growing interest for certain Italian philosophies and he substantiates this by quoting three recent anthologies written in English: *Recording Metaphysics. The New Italian Philosophy* (Borradori, 1988), *Radical Thought in Italy. A Potential Politics* (Hardt & Virno, 1996) and *The Italian Difference between Nihilism and Biopolitics* (Chiesa & Toscano, 2009). In particular, Esposito focuses on the anthology by Chiesa and Toscano, in which the "Italian difference" is found in the categories of nihilism and biopolitics. Although the first was born in Germany and the second one in France, it should be noted that the contemporary Italian thought has often reinterpreted the German and French philosophies in an original way, focusing its reflections on the category of secularization (Vattimo and Marramao), on political theology (Tronti and Cacciari) and on the already mentioned biopolitics (Negri, Agamben and Esposito). Therefore, according to Esposito, the Italian, French and German philosophies of the twentieth century have had similar outcomes, namely their common "American destiny", to emphasize which we may speak respectively of *Italian Thought* (or *Italian Theory*),[2] *French Theory*[3] and *German Philosophy*.[4]

3. Is There a Specific Italian Philosophical Tradition?

Having clarified the way in which Esposito uses the notion of "territory", rather than that of the nation or state, to refer to the philosophical culture produced in Italy, it remains to be explained in what sense it is legitimate to hold a discourse on this kind of philosophy. Does it make sense resorting to a territorial criterion in order to distinguish between the various traditions of thought? Assuming that it is sensible and legitimate, is it not also risky? And, finally, as specifically regards Italian philosophy, where would it start and what would be the specific character of this tradition of thought?

Regarding the first question, it is necessary immediately to emphasize that Esposito was not the first to defend the legitimacy of a discourse of this kind. The issue of the nationality of philosophy was born with Bertrando Spaventa (see Spaventa, 1862) and developed by Giovanni Gentile (see Gentile, 1904-1915 and Gentile, 1918). However, this issue was also addressed outside idealist philosophy, specifically by Eugenio Garin and his school (particularly Michele

[2] On *Italian Thought* see, other than Esposito (2010), Gentili (2012), Gentili & Stimilli (2015), Maltese & Mariscalco (2016).

[3] On *French Theory* see, in particular, Cusset (2003).

[4] On the distinction and definition of *German Philosophy*, *French Theory* and *Italian Thought* see, respectively, the second, third and fourth chapters of Esposito (2016).

Ciliberto, 2012). In fact, Garin insists on the admissibility of a specific Italian philosophy in the *Introduzione* to his *Storia della filosofia italiana*. According to Garin, when doing philosophy, you cannot ever fail to keep in mind "its essential connection with a specific period of time" (Garin, 1947, p. liii). In other words, philosophy is historically determined, that is to say, it has "a precise connection with definite historical situations, with conditions and limits actually determined or determinable" (*ibidem*). In short – continues Garin – "if ideas are not, and indeed they are not, born by parthenogenesis, and the philosophical discourse is always, using a Platonic expression, 'an illegitimate discourse', the historical reality of philosophizing will always assume an implicit relation to specific situations, within space-time dimensions" (*ibidem*). On this point we could hardly wish for a clearer message from the author of *Filosofia come sapere storico* (Garin, 1959).

4. Historicism and Chauvinism

In stressing the fact that philosophy is always located within specific dimensions of space and time, we must also make a number of clarifications to avoid unfortunate misunderstandings. One thinker who highlights the risks of a discourse that distinguishes the various traditions of thought by a territorial criterion is Alain Badiou (2012), who admits that the term 'French philosophy' might appear contradictory (either philosophy is universal, or does not exist), chauvinistic, imperialist and anti-American.

The alleged contradiction of terms such as 'Italian philosophy' or 'French philosophy' must not be insisted on any more than necessary. Far from being contradictory, these expressions show the undeniable link existing between philosophy and history, a connection that does not affect in any way the universal validity of philosophy. In other words, the particular genesis of an idea does not compromise in any way its universal value. Admitting that philosophy is historically determined does not mean being historicist or reducing ideas to their history. There is an endless dialectic between universality and particularity, philosophy and history, internationality and nationality.

As such discourses do not fall into the danger of historicism, so they do not necessarily invoke chauvinism. National sentiment and cosmopolitanism can live together, as demonstrated, for example, by Giuseppe Mazzini. In other words, we can talk about nation without thereby being nationalists. We must not confuse the healthy national sentiment (or patriotism) with nationalism. The language of patriotism is linked to "the common liberty of a people" (Viroli, 1995, p. 1) and not the supremacy of the people over all others. Patriotism implies solidarity of an oppressed people with everyone else in the same situation, as it allows "to recognize a foreigner as a fellow in the common struggle for liberty" (*ivi*, p. 144). On the contrary, the language of nationalism is used "to defend or reinforce the cultural, linguistic, and ethnic oneness and homogeneity of a people" (*ivi*, p. 1). The purpose of nationalism is to impose the domination of one people over the other, while the purpose of patriotism is to extend freedom to all peoples.

5. National Culture and Globalization

Therefore, accepting neither historicism nor chauvinism, imperialism nor ethnocentrism; there is also another point to emphasize: namely, the fact that to insist on national philosophical traditions also means resisting the abstract conception of a universality as a cancellation of all particular differences. If you have to guard well from the perversion of healthy national spirit into nationalism, it is similarly necessary to stem the process of globalization in its most extreme dynamic, in favor of a genuine internationalism. One of the most obvious aspects of globalization, namely the reduction of multiple cultures to a single "world-culture", is the continued decline in linguistic diversity. According to the twentieth edition of *Ethnologue: languages of the world* (2017), out of a total of 7,099 known languages in the world, many are at risk of extinction: 1,547 (or 22%) are threatened or shifting (levels

6b and 7 of the EGIDS – the *Expanded Graded Intergenerational Disruption Scale*), while 920 (or 13%) are moribund, nearly extinct or dormant (levels 8a, 8b and 9). Finally, the number of extinct languages (level 10) from 1950 is 360.[5] This means – concludes the Ethnologue – a rate of loss amounting to 6 languages per year. Thus, for example, in 1992, the Ubykh language was declared extinct following the death of Tevfik Esenç, the only one who was using it. Similarly, in 2008, the death of Marie Smith Jones and her sister Sophie Borodkin meant the disappearance of the Eyak language in Alaska. Just as the risk of extinction of many animal and plant species is a threat to nature, so the linguistic diversity reduction process causes incalculable damage to culture. Another treatise would be needed, in this regard, to explore the serious linguistic and stylistic impoverishment that goes hand in hand with the process just described – the constant decrease in the number of languages spoken in the world.

6. The Italian "Difference"

Following on from the preceding arguments presented, it is clear that addressing the issue of the nationality of philosophy is only one of the pieces that make up a discourse of a more general order in which culture in a broad sense is invested. Distinguishing different national philosophical traditions is not only legitimate and sensible, as has been shown, but it is necessary and vital in today's globalized world. This need manifests itself in numerous publications on the subject, addressed not only by Esposito and by the *Italian Theory*, but also by a number of scholars that, in Garin's wake, reflect on Italian philosophy (prominent among whom is Ciliberto, 2012). But if on the question of the beginning of the Italian philosophical tradition there is substantial agreement among scholars, we cannot say the same with regard to the particular characteristics of this tradition of thought. The Middle Ages is the period of gestation of a specifically Italian philosophy, whose real beginning should be placed in Renaissance humanism. On this point, the idealists Spaventa and Gentile agree, as do Garin and his school. Likewise, *Pensiero vivente* begins its genealogical analysis of Italian philosophy in the chapter *La vertigine dell'Umanesimo*. However, interpretations disagree on identifying the specific characters of the Italian philosophical tradition: is there a common thread that binds the different Italian philosophers from Renaissance humanism to the contemporary world? Are there privileged themes? What are the categories of thought and philosophical attitudes historically popular in Italy?

In answering these questions, we can look, for example, at Spaventa who states that the Italian philosophical genius is distinguished by being a "precursory genius" since Telesio foreruns the reflections of Bacon and Locke, Campanella precedes Descartes in the conceptualization of the *cogito*, Bruno's pantheism anticipates that of Spinoza and, finally, Vico begins the "Copernican Revolution" completed by Kant and thinks historically long before German idealism. Modern philosophy, born in Italy and developed abroad, sublates (in the sense of Hegel's *aufheben*) with the thought of Galuppi, Rosmini and Gioberti. The Spaventian circle made up of forerunners and sublations (*Aufhebung*) is taken up by Gentile, while it is abandoned anti-idealistically by Garin. The latter, precisely in reference to the particular characteristics of the Italian philosophical tradition, writes that: "instead of the great systematic constructions, a science of the human being and of its activities, a secular and earthly philosophy [...] was preferred" (Garin, 1947, p. lviii). In other words, the Italian philosophy was essentially "philology in Vichian sense as the science of human communication; [...] politics and morality as the urgency of the problem of the State and of the Church-State" (*ibidem*) and "religion understood especially as the need for clarification of the earthly function of the Church" (*ibidem*). To use Remo Bodei's words, the Italian philosophical tradition has always preferred "impure reason"

[5] https://www.ethnologue.com/endangered-languages (accessed June 22nd, 2017).

(Bodei, 1998, p. 75) to pure reason. Later, not only the already mentioned Ciliberto (2012), but many are those who, in Garin’s wake, have stressed particularly the ethical and civil vocation of Italian philosophy. According to Carlo Augusto Viano, in the Italian philosophical tradition, “civil engagement has always prevailed over conceptual accumulation” (Viano, 1982, p. 55, my translation). Similarly, Mario Perniola (1984) indicates civil activism as one of the four main features of Italian thought together with philology, eclecticism and militancy. For his part, the aforementioned Bodei saw in Italian philosophy “a constant civil vocation” (Bodei, 1998, p. 74). Recently, the same interpretative thesis was supported by Martirano & Cacciatore (2008). The latter, in particular, reviewing *Pensiero vivente* by Esposito, has highlighted how “the constant pursuit of the relationship between history and philosophy and its ethical and civil dimension” (Cacciatore, 2012, p. 141, my translation) constitutes the very essence of the Italian philosophical tradition in a manner surely greater than the category of life. In fact, according to Esposito, contrary to Garin and to all scholars mentioned up to now, life would be the privileged object of investigation of Italian thought.

From “precursory genius” to “living thought”, up to “impure reason” and ethical and civil vocation, there are many interpretive paradigms. In their difference, and greater or lesser plausibility, the self-reflection of Italian philosophy has always played an important part of our historiographical tradition. The motto “know thyself” addressed to the essence of Italian philosophy has been programmatic since the unification of Italy up to today. The persistence of the question on the existence of a specific Italian philosophical tradition perhaps says a lot more about this tradition than do the various responses provided by Spaventa up to those by the *Italian theory*. This question, among other things, has never been only “who are we?” but has always implied another query: “who do we want to be?”. If our past can provide some clues, then the undeniable ethical and civil vocation of our philosophy as well as its interest in concrete life must be a warning: our essence should not be forgotten, but reaffirmed as an endless task. In conclusion, the past of Italian philosophy can provide useful guidance on what should be the future of philosophy, not just abstract theory, but its actual practice; not only theoretical inspiration, but ethical and civil vocation.

REFERENCES

Augé, M. (2008). *Où est passé l’avenir?*. Paris: Editions Panama.

Badiou, A. (2012). *L’aventure de la philosophie française*. Paris: La fabrique.

Benasayag, M. & Schmit, G. (2003). *Les passions tristes. Souffrance psychique et crise sociale*. Paris: Éditions La Découverte.

Bodei, R. (1998/2006). *Il noi diviso. Ethos e idee dell’Italia repubblicana*. (Engl. Transl. by J. Parzen & A. Thomas). *We, the Divided: Ethos, Politics and Culture in Post-War Italy, 1943-2006*. New York: Agincourt Press.

Borradori, G. (Ed.) (1988). *Recording Metaphysics. The New Italian Philosophy*. Evanston: Northwestern University Press.

Cacciatore, G. (2012). ‘Pensiero vivente’ e pensiero storico. Un paradigma possibile per ripensare la tradizione filosofica italiana. *Iride*, 65, XXV, 135-142.

Cacciatore, G. & Martirano, M. (Eds.) (2008). *Momenti della filosofia civile italiana*. Reggio Calabria: La Città del Sole.

Chiesa, L. & Toscano, A. (Eds.) (2009). *The Italian Difference between Nihilism and Biopolitics*. Melbourne: re.press.

Ciliberto, M. (Ed.) (2012). *Il contributo italiano alla storia del pensiero. Ottava appendice*. Rome: Istituto dell’Enciclopedia Italiana.

Claverini, C. (2016). La filosofia italiana come problema. Da Bertrando Spaventa all’Italian Theory. *Giornale Critico di Storia delle Idee*, 15/16, 179-188.

Cusset, F. (2003). *French Theory: Foucault, Derrida, Deleuze, & Cie et les mutations de la vie intellectuelle aux États-Unis*. Paris: Éditions La Découverte.

Esposito, R. (2010/2012). *Pensiero vivente. Origine e attualità della filosofia italiana*. (Engl. Transl. by Z. Hanafi). *Living Thought: The Origins and Actuality of Italian Philosophy*. Stanford: Stanford University Press.

Esposito, R. (2016). *Da fuori. Una filosofia per l'Europa*. Turin: Einaudi.

Fukuyama, F. (1992). *The End of History and the Last Man.* New York: The Free Press.

Garin, E. (1947/2008). *Storia della filosofia italiana*. (Engl. Transl. by G. Pinton). *History of Italian Philosophy*. Amsterdam: Rodopi.

Garin, E. (1959/1990). *La filosofia come sapere storico*. Rome-Bari: Laterza.

Gentile, G. (1904-1915/1969). *Storia della filosofia italiana fino a Lorenzo Valla*. In G. Gentile, *Storia della filosofia italiana*. Florence: Sansoni.

Gentile, G. (1918/1963). *Il carattere storico della filosofia italiana*. In G. Gentile, *I problemi della Scolastica e il pensiero italiano*, III edizione riveduta. Florence: Sansoni, 209-236.

Gentili, D. (2012). *Italian Theory. Dall'operaismo alla biopolitica*. Bologna: Il Mulino.

Gentili, D. & Stimilli, E. (2015). *Differenze italiane. Politica e filosofia: mappe e sconfinamenti*. Rome: Derive Approdi.

Hardt, M. & Virno, P. (Eds.) (1996), *Radical Thought in Italy. A Potential Politics*. Minneapolis: University of Minnesota Press.

Huntington, S.P. (1996). *The Clash of Civilizations and the Remaking of World Order*. New York: Simon & Schuster.

Jacobelli, J. (1986). *Dove va la filosofia italiana?.* Rome-Bari: Laterza.

Maltese, P. & Mariscalco, D. (2016). *Vita, politica, rappresentazione. A partire dall'Italian Theory*. Verona: Ombre Corte.

Perniola, M. (1984). The Difference of the Italian Philosophical Culture. *Graduate Faculty Philosophy Journal*, 10(1), 103-116.

Spaventa, B. (1862/1908). *La filosofia italiana nelle sue relazioni con la filosofia europea*. Bari: Laterza.

Viano, C.A. (1982). *Il carattere della filosofia italiana contemporanea*. In N. Bobbio *et al*., *La cultura filosofica italiana dal 1945 al 1980*. Naples: Guida, 9-56.

Viroli, M. (1995/1995). *Per amore della patria. Patriottismo e nazionalismo nella storia*. Bari: Laterza. Engl. Ed. *For Love of Country: An Essay on Patriotism and Nationalism*. Oxford: Oxford University Press.

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

Graphic Design

Dondina e associati (print version)

Graphic Layout

Direweb (on line version)

50,00 €