1506
UNIVERSITÀ DEGLI STUDI DI URBINO CARLO BO

DiSPeA
DIPARTIMENTO DI SCIENZE PURE E APPLICATE

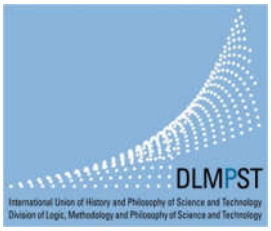
Dottorato di Ricerca in Scienze di Base e Applicazioni

SILFS
società italiana di logica e filosofia delle scienze

REGIONE AUTÒNOMA DE SARDIGNA
REGIONE AUTONOMA DELLA SARDEGNA

UNIVERSITÀ DI CAGLIARI

DIPARTIMENTO DI PEDAGOGIA, PSICOLOGIA, FILOSOFIA






# BOOK OF ABSTRACTS

## 4th SILFS POSTGRADUATE CONFERENCE ON LOGIC AND PHILOSOPHY OF SCIENCE

## UNIVERSITY OF URBINO, URBINO (ITALY)

## 3rd-7th of June 2019

# PROGRAM

## 3rd of JUNE 2019, Monday

09:30 Registration
Venue: Palazzo Albani, Univ. of Urbino,
via Timoteo Viti 10, Urbino, 60129 Italy

10:00 Welcome Coffee

10:30 Welcome Address by **Vilberto Stocchi** (Chancellor of the Univ. of Urbino), **Gino Tarozzi** (Univ. of Urbino) and **Roberto Giuntini** (Univ. of Cagliari, President of SILFS)

11:00 **Alessandro Guaitolini** (Roma Tre Orchestra)
*J. S. Bach, Suite No. 2 in D minor, BWV 1008*

11:30 – 12:30 **Invited Speaker:** **Gerhard Heinzmann** (Université de Lorraine) [Pag. 8]
*Formalizations of constructiveness: predicative and operative mathematics from Weyl to Feferman, and developments of Hilbert's meta-mathematics.*
Chair: Gino Tarozzi (University of Urbino)

## FIRST SESSION: Philosophy of Mind and Cognitive Science [Pag. 13]

*Chair*: **Francesco Bianchini** (University of Bologna)

15:00 **Vanja Subotić** (University of Belgrade) [Pag. 13]
*Can Connectionism Save Usage-based Theories? A reappraisal of the connectionism vs. symbolism debate*

15:40 **Marco Facchin** (IUSS, Pavia) [Pag. 16]
*Can "Basic Minds" ever meet content? A problem for Radical Enactivism*

16:20 Coffee & Tea

*Chair*: **Luisa Damiano** (University of Messina)

16:40 **Stefano Calboli** (University of Urbino), **Vincenzo Fano** (University of Urbino) and **Roberto Macrelli** (University of Urbino) [Pag. 19]
*The Moral Decoy Effect. Asymmetric Dominance Effect in Morality and Its Political Implications*

17:20 **Robert Chis-Ciure** (Uni. of Bucharest) and **Francesco Ellia** (Uni. of Bologna) [Pag. 20]
*Facing up to the Hard Problem as an Integrated Information Theorist*

18:00 – 19:00 **Invited Speaker:** **Stefania Centrone** (Technische Universität Berlin) [Pag. 8]
*Husserl and Weyl*
Chair: Vincenzo Fano (University of Urbino)

## 4th of JUNE 2019, Tuesday

## SECOND SESSION: General Philosophy of Science [Pag. 26]

*Chair*: **Mario Alai** (University of Urbino)

09:30 **Eugenio Petrovich** (University of Siena) [Pag. 26]
*Bridging the Gap between General Philosophy of Science and Scientometrics: Towards an Epistemological Theory of Citations*

10:10 **Alejandra Casas Munoz** (University of Bristol) [Pag. 27]
*An Inferential Conception of Scientific Explanation*

10:50 Coffee & Tea

*Chair*: **Marco Giunti** (University of Cagliari)

11:10 **Alberto Corti** (University of Urbino) [Pag. 30]
*Scientific Realism without Reality? What remains when metaphysics is left out*

## THIRD SESSION: Classical and Non-Classical Logics [Pag. 36]

*Chair*: **Sara Negri** (University of Helsinki)

15:00 **Stefano Bonzio** (Marche Polytechnic University), **Tommaso Flaminio** (Artificial Intelligence Research Institute, IIIA — Spanish National Research Council, CSIC) and **Paolo Galeazzi** (University of Copenhagen) [Pag. 36]
*Sure-wins under coherence*

15:40 **Michele Pra Baldi** (University of Cagliari) [Pag. 40]
*The lattice of logics of variable inclusion*

16:20 Coffee & Tea

*Chair*: **Roberto Giuntini** (University of Cagliari)

16:40 **Sara Negri** (University of Helsinki) and **Edi Pavlovic** (University of Helsinki) [Pag. 40]
*DSTIT modalities through a sequent calculus*

17:20 **Davide Dalla Rosa** (University of Padua) [Pag. 41]
*In which sense is Kant's categorical syllogistic non-classical?*

18:00 – 19:00 **Invited Speaker:** **Sara Negri** (University of Helsinki) [Pag. 8]
*Unveiling the constructive core of classical theories: A contribution to 90 years of Glivenko's theorem*
Chair: Roberto Giuntini (University of Cagliari)

20:00 ***SOCIAL DINNER***

## 5th of JUNE 2019, Wednesday

## FOURTH SESSION: Philosophy and Foundations of Physics [Pag. 44]

*Chair*: **Laura Felline** (University of Rome III)

09:30 **Frida Trotter** (University of Lausanne) [Pag. 44]
*The [un]observability of the entangled state*

10:10 **Ivan Chajda** (University of Olomouc), **Davide Fazio** (University of Cagliari) and **Antonio Ledda** (University of Cagliari) [Pag. 48]
*The Generalized Orthomodularity Property: Configurations, Pastings and Completions*

10:50 Coffee & Tea

*Chair*: **Giuseppe Sergioli** (University of Cagliari)

11:10 **Silvia Bianchi** (IUSS, Pavia) [Pag. 49]
*Introducing Thin Objects in Mathematical Structuralism: Ontological Dependence and Grounding for a Weak Approach*

11:50 **Andrea Oldofredi** (University of Lausanne) [Pag. 55]
*An Internal Realist Interpretation of the Primitive Ontology Programme*

## FIFTH SESSION: Philosophy of Social Sciences [Pag. 61]

*Chair*: **Pierluigi Barrotta** (University of Pisa)

15:00 **Giulia Miotti** (Sapienza University of Rome) [Pag. 61]
*Imperfect Knowledge and Non-Equilibrium in Finance: The Efficientist Approach In The Light Of Fallibilism*

15:40 **Stefano Vaselli** (University of Turin) [Pag. 61]
*Is Methodological Individualism Without Ontological Individualism Possible*

16:20 Coffee & Tea

*Chair*: **Eleonora Montuschi** (Ca' Foscari University of Venice)

16:40 – 17:40 ***Winner of SILFS Philosophy of Science Prize***
**Remco Heesen** (University of Western Australia) [Pag. 9]
*The Necessity of Commensuration Bias in Grant Peer Review*

## 6th of JUNE 2019, Thursday

## SIXTH SESSION: Philosophy of Biology and Health Sciences [Pag. 66]

*Chair*: **Gilberto Corbellini** (Sapienza University of Rome, CNR-DSU)

09:30 **Federico Boem** (University of Milan), **Stefano Bonzio** (Marche Polytechnic University) and **Barbara Osimani** (Marche Polytechnic University; LMU Münich) [Pag. 66]
*The Cochrane case: an epistemic analysis on decision-making and trust in science in the age of information*

10:10 **Silvano Zipoli Caiani** (University of Florence), **Federico Boem** (University of Milan) and **Gabriele Ferretti** (University of Florence) [Pag. 68]
*Out of our Skull, within our Skin: The Gut Microbiota and the Extended Mind Thesis*

10:50 Coffee & Tea

*Chair*: **Cristina Amoretti** (University of Genoa)

11:10 **Chiara Beneduce** (University Campus Bio-Medico of Rome) [Pag. 71]
*"complexio". A systemic approach to organism's dynamics*

11:50 ***General assembly of SILFS***

## SEVENTH SESSION: Foundations of Logic and Mathematics [Pag. 74]

*Chair*: **Emiliano Ippoliti** (Sapienza Univ. of Rome)

15:00 **Ludovica Conti** (University of Pavia) [Pag. 74]
*Russell's Paradox ways out*

15:40 **Claudio Ternullo** (Kurt Gödel Research Center for Mathematical Logic, University of Vienna) and **Luca Zanetti** (IUSS, Pavia) [Pag. 74]
*From Bolzano to Frege: A Cantorian Path*

16:20 Coffee & Tea

*Chair*: **Laura Crosilla** (Univ. of Birmingham)

16:40 **Matteo Zicchetti** (University of Bristol) [Pag. 75]
*Truth-theories, Cognitive Projects and Trustworthiness*

17:20 **Michele Lubrano** (University of Turin) [Pag. 84]
*Difference-making and explanation in mathematics*

18:00 – 19:30 **Invited Speaker:** **Heinrich Wansing** (Ruhr-Universität Bochum) [Pag. 9]
*Connexive Heyting-Brouwer logic*
Chair: Francesco Bianchini (University of Bologna)

## 7th of JUNE 2018, Friday

### EIGHTH SESSION: Foundations of Computing and Artificial Intelligence [Pag. 88]

*Chair*: **Guglielmo Tamburrini** (University of Naples Federico II)

09:30 **Mirko Tagliaferri** (University of Urbino) [Pag. 88]
*How to Build a Formal Notion of Trust*

10:10 **Sandro Sozzo** (University of Leicester) [Pag. 89]
*Entanglement and Quantum Structures in Concept Combinations*

10:50 Coffee & Tea

*Chair*: **Edoardo Datteri** (University of Milano-Bicocca)

11:10 **Silvia Crafa** (University of Padua) and **Lucrezia Pelizzon** (University of Cagliari) [Pag. 89]
*Epistemological questions for a philosophical education in artificial intelligence*

11:50 ***Best Presentations Awards***
Chair: Roberto Giuntini (University of Cagliari) and Gino Tarozzi (University of Urbino)

*Closing*

**Further information** can be found at www.silfs.it or by writing to
Pierluigi Graziani pierluigi.graziani@uniurb.it

**Scientific Committee:**
The Council of SILFS, Italian Society for Logic and the Philosophy of Science (www.silfs.it).

**Organizing Committee:**
Mario Alai (University of Urbino)
Francesco Bianchini (University of Bologna)
Vincenzo Fano (University of Urbino)
Marco Giunti (University of Cagliari)
Roberto Giuntini (University of Cagliari)
Pierluigi Graziani (University of Urbino)
Giuseppe Sergioli (University of Cagliari)
Gino Tarozzi (University of Urbino)

## INVITED SPEAKERS

## ABSTRACTS

**Gerhard Heinzmann** (Université de Lorraine)
*Formalizations of constructiveness: predicative and operative mathematics from Weyl to Feferman, and developments of Hilbert's meta-mathematics.*

Formalizations of constructiveness: predicative and operative mathematics from Weyl to Feferman, and developments of Hilbert's meta-mathematics.
In my lecture, I propose to clarify the notion of constructiveness in mathematics by two approaches that dominate the discussion until today: one pursues the idea of avoiding a circularity in the formation of mathematical concepts and is oriented to the concept of predicativity introduced in 1906/1907 by Poincaré and Russell, specified in 1918 by Weyl and then, after the Second World War, by Lorenzen and Hao Wang, and finally formalized by Solomon Feferman using a "constructive" ordinal number.
The other attempt to formalize constructiveness emanates from Hilbert's program and from the necessity of transgressing the original idea of finite meta-mathematics by attempting a new demarcation line between evident reasoning and suspicious reasoning through the definition of constructive ordinals.

**Stefania Centrone** (Technische Universität Berlin)
*Husserl and Weyl*

The paper focuses on the influence Husserl's phenomenology might have had on the ideas of Hermann Weyl and contrast their views on specific issues, in particular those concerning the nature of mathematical knowledge, the ontological status of mathematical objects, their conception of logic and symbolization. It bases on an analytical and internal reading of two important works by Weyl, *Das Kontinuum* (1918) and *Raum – Zeit – Materie* (1919), in which Husserl is explicitly mentioned, as well as those parts of Husserl's work, Weyl explicitly refers to. Some biographical details are recalled now and again to better follow Weyl's formalistic beginnings as a student of Hilbert, his confronting with foundational questions, his intuitionistic turn, his partial comeback to formalistic positions as well as Husserl's influence on his ideas.

**Sara Negri** (University of Helsinki)
**Unveiling the constructive core of classical theories: A contribution to 90 years of Glivenko's theorem**

Glivenko's well known result of 1929 established that a negated propositional formula provable in classical logic is even provable intuitionistically. Similar later transfers from classical to intuitionistic provability therefore fall under the nomenclature of Glivenko-style results: these are results about classes of formulas for which classical provability yields intuitionistic provability. The interest in isolating such classes lies in the fact that it may be easier to prove theorems by the use of classical rather than intuitionistic logic. Further, since a proof in intuitionistic logic can be associated to a lambda term and thus obtain a computational meaning, such results have more recently been gathered together under
the conceptual umbrella "computational content of classical theories."
They also belong to a more general shift of perspective in foundations: rather than developing constructive mathematics separately, as in Brouwer's program, one studies which parts of classical mathematics can be directly translated into constructive terms.
We shall survey how Glivenko-style results can be easily obtained by the choice of suitable
sequent calculi for classical and intuitionistic logic, by the conversion of axioms into inference rules, and by the procedure of geometrization of first order logic.

**Remco Heesen** (University of Western Australia)
*The Necessity of Commensuration Bias in Grant Peer Review*

Peer reviewers at many funding agencies and scientific journals are asked to score submissions both on individual criteria and overall. The overall scores should be some kind of aggregate of the criteria scores. Carole Lee identifies this as a potential locus for bias to enter the peer review process, which she calls commensuration bias. Here I view the aggregation of scores through the lens of social choice theory. I argue that in many situations, especially when reviewing grant proposals, it is impossible to avoid commensuration bias.

**Heinrich Wansing** (Ruhr-Universität Bochum)
*Connexive Heyting-Brouwer logic*

Systems of *connexive logic* and the *bi-intuitionistic* logic BiInt that is also known as *Heyting-Brouwer logic* have been carefully studied since the 1960s and 1970s with various philosophical and mathematical motivations, see [2, 13, 14, 28] and [5, 10, 19, 20, 21]. The characteristic principles of connexive logic are usually traced back to Aristotle and Boethius, and the co-implication of BiInt can be traced back to Skolem [23].

A distinctive feature of connexive logics is that they validate the so-called
Aristotle's theses: $\sim(\alpha \to \sim\alpha)$ and $\sim(\sim\alpha \to \alpha)$, and
Boethius' theses: $(\alpha \to \beta) \to \sim(\alpha \to \sim\beta)$ and $(\alpha \to \sim\beta) \to \sim(\alpha \to \beta)$ .

An intuitionistic (or constructive) connexive modal logic, CK, which is a constructive connexive analogue of the smallest normal modal logic K, was introduced in [25] by extending a certain basic constructive connexive logic, C, which is a connexive variant of *Nelson's paraconsistent logic* [1, 7, 8, 15, 16]. A *classical connexive modal logic* called CS4, which is based on the positive normal modal logic S4, was introduced in [6] as a Gentzen-type sequent calculus. The Kripke-completeness and cut-elimination theorems for CS4 were shown, and CS4 was shown to be embeddable into positive S4 and to be decidable. Moreover, it was shown in [6] that the basic constructive connexive logic C can be faithfully embedded into CS4 and into a subsystem of CS4 lacking syntactic duality between necessity and possibility.

*Heyting-Brower logic*, which is an extension of both *dual-intuitionistic logic*, DualInt, and intuitionistic logic, Int, was introduced by Rauszer [19, 20, 21], who proved algebraic and Kripke completeness theorems for BiInt. As was shown by Uustalu in 2003, cf. [17], the original Gentzen-type sequent calculus by Rauszer [19] does not enjoy cut-elimination, and various kinds of sequent systems for BiInt have been presented in the literature, including cut-free display sequent calculi in [5, 26], see also [17] and [18] for a comparison between sequent calculi for BiInt. Moreover, BiInt is known to be a logic that has a faithful embedding into the future-past tense logic KtT4 [11], and a modal logic based on BiInt was studied by Lukowski in [12].

Dual-intuitionistic logics are logics which have a Gentzen-type sequent calculus in which sequents have the restriction that the antecedent contains at most one formula [3, 4, 24]. This restriction of being singular in the antecedent is syntactically dual to that in Gentzen's sequent calculus LJ for intuitionistic logic, which is singular in the consequent. Historically speaking, the logics in the set of logics containing Czermak's *dual-intuitionistic calculus* [3], Goodman's *logic of contradiction* or *anti-intuitionistic logic* [4], and Urbas's extensions of Czermak's and Goodman's logics [24] were collectively referred to by Urbas as dual-intuitionistic logics. The dual-intuitionistic logic referred to as DualInt in [5, 27] is the implication-free fragment of BiInt. An interpretation of DualInt as the *logic of scientific research* was presented by Shramko in [22].

In this talk, which is based on [9], the two approaches are combined and the *bi-intuitionistic connexive logic* (or *connexive Heyting-Brower logic*), BCL, is introduced as a Gentzen-type sequent calculus. The logic BCL may be seen as an extension of the connexive logic C from [25] by the co-implication of BiInt, using a connexive understanding of negated co-implication. Another understanding of co-implication is developed in [27, 29, 31], and a natural deduction proof system and formulas-as-types notion of construction for a bi-connexive logic 2C that assumes this understanding of co-implication is presented in [30].

In this talk, in a first step, the logic BCL is introduced as a Gentzen-type sequent calculus, and a dual-valuation-style Kripke semantics for BCL is defined. BCL is constructed on the basis of Takeuti's cut-free Gentzen-type sequent calculus LJ' for Int. Gentzen-type sequent calculi ICL, DCL, BL, IL and DL for

*intuitionistic connexive logic, dual-intuitionistic connexive logic, bi-intuitionistic logic, intuitionistic logic* and *dual-intuitionistic logic*, respectively, are defined as subsystems of BCL.

In a second step, some theorems for syntactically and semantically embedding BCL into BL are proved, and using these theorems, the completeness theorem with respect to the Kripke semantics for BCL is shown as a central result. The cut-elimination theorems for ICL and DCL are shown using some restricted versions of the syntactical embedding theorem of BCL int BL. The cut-elimination theorem does *not* hold for BCL and BL.

Next, some theorems for syntactically embedding ICL into DCL and viceversa are shown. These theorems reveal that ICL and DCL are syntactically dual to each other in a certain sense. Thus, it is shown in these theorems that BCL is constructed based on a duality principle of the characteristic subsystems. Finally, if time permits, sound and complete tableau calculi will be presented for BCL and its subsystems ICL, DCL, BL, IL and DL using triply-signed formulas.


[1] A. Almukdad and D. Nelson, "Constructible falsity and inexact predicates", *Journal of Symbolic Logic* 49 (1984), 231-233.

[2] R. Angell, "A propositional logics with subjunctive conditionals", *Journal of Symbolic Logic* 27 (1962), 327-343.

[3] J. Czermak, "A remark on Gentzen's calculus of sequents", *Notre Dame Journal of Formal Logic* 18 (1977), 471-474.

[4] N.D. Goodman, "The logic of contradiction", *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 27 (1981), 119-126.

[5] R. Goré, "Dual intuitionistic logic revisited", *Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods* (TABLEAUX 2000), Lecture Notes in Artificial Intelligence 1847, Berlin, Springer, 2000, 252-267.

[6] N. Kamide and H. Wansing, "Connexive modal logic based on positive S4", in: *Logic without Frontiers: Festschrift for Walter Alexandre Carnielli on the occasion of his 60th Birthday*. (Jean-Yves Bézian and Marcelo Esteban Coniglio, eds.), College Publications, London, 2011, 389-410.

[7] N. Kamide and H. Wansing, "Proof theory of Nelson's paraconsistent logic: A uniform perspective", *Theoretical Computer Science* 415 (2012), 1-38.

[8] N. Kamide and H. Wansing, *Proof theory of N4-Related Paraconsistent Logics*, Studies in Logic 54, London, College Publications, 2015.

[9] N. Kamide and H. Wansing, "Completeness of connexive Heyting-Brouwer logic", 2015, submitted to *IFCoLog Journal of Logic and their Applications* 3 (2016), 441-466.

[10] E.G.K. López-Escobar, "On intuitionistic sentential connectives L", *Revista Columbiana de Matemáticas* 19 (1985), 117-130.

[11] P. Lukowski, "Modal interpretation of Heyting-Brouwer logic", *Bulletin of the Section of Logic* 25 (1996), 80-83.

[12] P. Lukowski, "A deductive-reductive form of logic: Intuitionistic S4 modalities", *Logic and Logical Philosophy* 10 (2002), 79-91.

[13] S. McCall, "Connexive implication", *Journal of Symbolic Logic* 31 (1966), 415-433.

[14] S. McCall, "A history of connexivity", in D.M. Gabbay et al. (eds.), *Handbook of the History of Logic. Volume 11. Logic: A History of its Central Concepts*, Amsterdam, Elsevier, 415-449.

[15] D. Nelson, "Constructible falsity", *Journal of Symbolic Logic* 14 (1949), 16-26.

[16] S.P. Odintsov, *Constructive Negations and Paraconsistency*, Springer, Dordrecht, 2008.

[17] L. Pinto and T. Uustalu, "Relating sequent calculi for bi-intuitionistic propositional logic", in: S. van Bakel, S. Berardi and U. Berger (eds.), *Proceedings Third International Workshop on Classical Logic and Computation*, Electronic Proceedings in Theoretical Computer Science, Vol. 47, 2010, 57-72.

[18] L. Postniece, *Proof Theory and Proof Search of Bi-Intuitionistic and Tense Logic*, Ph.D. thesis, The Australian National University, Camberra, 2010.

[19] C. Rauszer, "A formalization of the propositional calculus of H-B logic", *Studia Logica* 33 (1974), 23-34.

[20] C. Rauszer, "Applications of Kripke models to Heyting-Brouwer logic", *Studia Logica* 36 (1977), 61-71.

[21] C. Rauszer, "An algebraic and Kripke-style approach to a certain extension of intuitionistic logic", *Dissertationes Mathematicae*, Polish Scientific Publishers, 1980, 1-67.

[22] Y. Shrambo, "Dual intuitionistic logic and a variety of negations: The logic of scientific research", *Studia Logica* 80 (2005), 347-367.
[23] T. Skolem, "Untersuchungen über die axiome des Klassenkalküls und über Produktations und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen", in: *Skrifter utgit av Videnskabsselskapet i Kristiania*, Vol. 3, 1919; reprinted in *T. Skolem, Selected Works in Logic*, J.E. Fenstad (ed.), Universitetforlaget, Oslo, 1970, 67-101.
[24] I. Urbas, "Dual-Intuitionistic logic", *Notre Dame Journal of Formal Logic* 37 (1996), 440-451.
[25] H. Wansing, "Connexive modal logic", in: R. Schmidt et al. (eds.), *Advances in Modal Logic*, Vol. 5, London, College Publications, 2005, 367-383.
[26] H. Wansing, "Constructive negation, implication, and co-implication", *Journal of Applied Non-Classical Logics* 18 (2008), 341-364.
[27] H. Wansing, "Falsification, natural deduction and bi-intuitionistic logic", *Journal of Logic and Computation* 26 (2016), 425-450.
[28] H. Wansing, "Connexive Logic", The Stanford Encyclopedia of Philosophy (fall 2014 Edition), E. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2014/entries/logic-connexive/
[29] H. Wansing, "On split negation, strong negation, information, falsification, and verification", in: K. Bimbó (ed.), J. Michael Dunn on Information Based Logics, Springer, Dordrecht, 2016, 161-189.
[30] H. Wansing, "Natural deduction for bi-connexive logic and a two-sorted typed λ-calculus", 2015, submitted to IFCoLog Journal of Logic and their Applications 3 (2016), 413-439.
[31] H. Wansing, "A more general general proof theory", Journal of Applied Logic 25 (2017), 23-46.

## FIRST SESSION: Philosophy of Mind and Cognitive Science

**Vanja Subotić** (University of Belgrade)
*Can Connectionism Save Usage-based Theories? A reappraisal of the connectionism vs. symbolism debate*

As Rogers & McClelland (2014) rightly point out, cognitive science may be defined as the effort to answer three questions:
(1) What processes support the complex behavior of intelligent systems;
(2) What kinds of representations do such processes operate on;
(3) What is the core basis of such processes and representations, i.e. are they innate or learnable through experience?
Theoretical commitments made by choosing answers to these three questions have serious philosophical and methodological consequences about assumptions concerning the nature of human cognition. Throughout the history of cognitive science, it is possible to distinguish between two paradigms, each taking its side in the older, full-blooded philosophical debate between rationalists and empiricists, and thus making theoretical commitments in accordance with chosen sides. Namely, from the late 1950s through the 1970s and 1980s, cognitive science had been shaped by the mind-as-computer metaphor, thereby drawing an analogy between the brain as hardware and the mind as software. Symbolic paradigm provided the following answers to abovementioned questions:
(1) Cognitive processes are like digital computer programs because they resemble ordered lists of explicit or implicit rules; and they are modular and sequential, which means that each process follows domain-specific rules and that each process waits for its predecessor to finish in order to compute the appropriate output;
(2) Representations are discrete and symbolic. They have combinatorial syntax and semantics, which means that structurally molecular representations have syntactic constituents that are themselves either structurally molecular or atomic and that the semantic content of a molecular representation is a function of the semantic contents of its syntactic constituents;
(3) Knowledge is largely innate since the number of possible ordered lists of rules is virtually unbounded, so the initial constraints must be prespecified rather than learned. Symbolic models of language processing were heavily influenced by Noam Chomsky (e.g. 1965), who had made sharp distinction between linguistic competence and linguistic performance, thereby arguing that certain sets of structural, linguistic rules are innate which allows speakers to acquire native languages quickly and rather accurately; i.e. that each speaker is endowed with "universal grammar". In such models, performance is an imperfect reflection of abstract encoded competence constituted by Chomsky's speculations about the "universal grammar" (cf. Plaut 2000).
However, in the 1980s and especially in 1990s, the connectionist paradigm has become prominent in cognitive science, mostly because of its neural plausibility. Connectionism provided the following innovative answers to abovementioned three questions:
(1) Cognitive processes are like analogue computer programs because the aim is to find the most highly associated output corresponding to an arbitrary input within the connectionist network. Weights among connections between input units and output units are in fact adjusted until the statistical properties of input units are recapitulated among the environmental events. This detection of statistical patterns is produced by hidden units that are abstract and that are not directly connected to the environment as input and output units are;
(2) Representations are parallelly or neurologically distributed within a neural network and over microfeatures or lower level units. By giving a complete, formal and precise account of microlevel, or subsymbolic level – where states of units' activation correspond to patterns of statistical and neural activity – it is possible to simultaneously obtain approximately true generalizations at
macrolevel, or symbolic level;
(3) Knowledge is largely learnable from experience concerning various environmental factors and events. A plethora of learning procedures is available in connectionist research: backpropagation or error correction, Hebbian learning, etc.
Starting with formative work by Feldman & Ballard (1982), as well as McClelland & Rumelhart (1986), connectionism has been severely criticized by leading traditional cognitive scientists and linguists such as Fodor & Pylyshyn (1988), Pinker & Prince (1988), or Marcus (1998). Generally, these authors concluded that connectionism cannot be a viable alternative to symbolism, even though connectionist models can be

useful for cognitive research and accepted in the mainstream cognitive science as long as they are mere implementations of symbolic architectures with biological flavor. On the other hand, cognitive scientists such as Smolensky (1987, 1988) and Clark (1990) defended connectionism by emphasizing the revolutionary character of this approach to cognitive modelling. Yet, Smolensky (1999) and Clark (1993) nonconcurred on the status of connectionism when construed as a hypothesis about cognitive architecture.

In other words, Smolensky endorsed the middle way and claimed that eliminative connectionism, which aims to put symbolism and nativism ad acta, represents impractical delusion: connectionism and Chomskyan tendencies in linguistics which are embedded in symbolism are not incompatible at all, rather they should be both regarded as valuable strategies in language research. Specifically, connectionism, in that case, should be regarded as a commitment to a certain way of modelling computational mechanisms, and Chomskyan tendencies, especially nativism, should be regarded as key theoretical commitments to a certain way of explaining the empirical generalizations provided by connectionist models.

Smolensky's remarks were written during the particularly heated controversy which surrounded language research in 1999 (cf. Christiansen & Chatter 1999). Namely, models of language processing were an unsurmountable obstacle for connectionist researchers, because it seemed that every model had to include at least some "hand-wired" rule into the neural network. Even the first ambitious model PARSNIP (Hanson & Kegl 1987) which learned to parse tagged sentences, and to assign grammatical structures to novel sentences after training on corpora, proved to be prone to prespecified instructions. It seemed that connectionism failed to provide us with means to reject linguistic nativism along with several other Chomskyan ideas. Nevertheless, by using dynamical systems theory and novel recurrent neural networks, Elman (1991) and Tabor & Tanenhaus (1999) made promising results. Such connectionist models of language processing were quite different from symbolic ones: instead of focusing on abstract competence, the aim was to model performance of actual language users, i.e. to articulate the computational principles that account for linguistic usage (Plaut 2000). Elman et al. (1996) used connectionist models of language processing for assessing the disparate answers on the question (3). The authors claimed that they are using connectionist models in order to show how domain-specific representations can emerge from domain-general architectures [such as a connectionist architecture] and learning algorithms and how these can ultimately result in a process of modularization as the end product of development rather than its starting point (1996: 115). Therefore, the main goal was to make connectionism a viable alternative to symbolism by making the need for innate symbolic architecture in language research a non-fundamental one for explaining the linguistic behavior.

Marcus (1998) saw this goal as a failed attempt to provide the first computational account of Piagetian constructivism, viz. the view that learning leads the child to develop new types of representations of abstract categories such as noun or verb. Moreover, he claimed that Elman et al.
(1996) provide us with:

(i) only a strawman argument against nativism;

(ii)models that are replete with innately defined modules and innately defined representations;

(iii) deeply flawed methodology, since the models are unable to generalize in the ways that humans do to items that include properties that did not appear in the training set.

A fortiori, connectionist models do not provide any support for Piagetian constructivism, or valid objection to Chomskyan nativism, but the fact that the authors were not able to deliver connectionist models that are consistent with their goal does not mean that the endeavour of constructivism is doomed (cf. Marcus 1998).

My aim in this talk is two-fold: I will argue in favor of eliminative connectionism, since I believe that by putting aside nativism connectionism can emerge as a viable alternative to symbolism in both methodological and architectural terms; and at the same time I will examine whether connectionism can provide support for usage-based theory à la Tomasello (2003) rather than for Piagetian constructivism. Contra Marcus, I will draw heavily on brand new connectionist models which are implementing convolutional neural networks (Karpathy & Fei Fei 2015, Karpathy & Fei Fei & Johnson 2016). In these models convolutional neural networks (CNN) are combined with bi-directional and multi-modal recurrent neural networks (RNN) in such a way that CNN is used for image classification and object detection, bi-directional RNN for determining the sequence of the words in sentences of corpora, and multi-modal RNN for generating the novel descriptions of image regions by using inferred alignments of two modalities. Such a boosted connectionist architecture may well be indicative of how children acquire knowledge of the objects of reference which surround them. Therefore, in my opinion, there is no need to presuppose innate linguistic competence in this case, construed as a domain-specific ability of human children, but rather what is innate

can be a mechanism which is construed as some sort of a domain-general ability, used for a plethora of higher-cognitive processes. This means that what we need is, in fact, a theory of language acquisition and processing which must specify a priori only a single set of general learning processes with which to learn everything about a language, including these correspondences between visual input and verbal output. Thus, contra Marcus and Smolensky, I will be stating that the methodological progress of connectionist models gives us good reasons to reject linguistic nativism in favor of more flexible theoretical commitment, such as usage-based theory (UBT) in linguistics. According to the proponents of UBT, children come to the process of language acquisition, at around one year of age, equipped with two sets of cognitive skills, both evolved for other, more general functions before linguistic communication emerged in the human species: intention-reading and pattern-finding. Intention-reading is what children must do in order to discern the intentions of adults when they use linguistic conventions to achieve social ends, and thereby to learn these conventions from them culturally. Pattern-finding is what children must do to go productively beyond the individual utterances they hear adults using around them to create abstract linguistic constructions. As a summary term for such general mechanisms such as categorization, analogy and categorial-based induction, pattern-finding is the central cognitive construct. A pioneer of UBT, Michael Tomasello has remarked that even though connectionism is way more in accordance with UBT than symbolism, nevertheless it has its own limitations, viz. ignorance of communicative intentions and visual pieces of information, as well as work with rather small units such as words and morphemes. Yet, models that have been proposed by Karpathy & Fei Fei seem to be going in the direction of providing the computational framework for at least one essential part of the UBT, i.e. pattern-finding. In conclusion, drawing on Steedman (1999) I will suggest the way in which connectionist research program should progress in order to account for the other part of UBT, i.e. intention-reading. Namely, it is likely that such a research program would proceed by first conceptualizing primary bodily actions and sensations, then coordinating perception and primary actions like reaching, then conceptualizing identity, permanence and location of objects. Later stages would have to include the conceptualization of more complex events including intrinsic actions of objects themselves (such as falling), translations and events involving multiple participants, intermediate participants including tools, and goals.


Clark, A. 1990. “Connectionism, Competence, and Explanation”. British Journal for the Philosophy of Science 41: 195-222.

_____. 1993. Associative Engines: Connectionism, Concepts, and Representational Change. Cambridge, MA: The MIT Press.

Christiansen, M. H. & Chater, N. 1999. “Connectionist Natural Language Processing: The State of Art”. Cognitive Science 23 (4): 417-437.

Chomsky, N. 1965. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

Elman, J. L. 1991. “Distributed Representations, Simple Recurrent Networks, and Grammatical Structure”. Machine Learning 7: 195-225.

_____. et al. 1996. Rethinking Innateness: A Connectionist Perspective on Development. Cambridge, MA: MIT Press.

Feldman, J. L. & Ballard, D. H. 1982. “Connectionist Models and their Properties”. Cognitive Science 6 (12): 205-254.

Fodor, J. & Pylyshyn, Z. 1988. “Connectionism and Cognitive Architecture”. Cognition 28: 3-71. Hanson, S. J. & Kegl, J. 1987. “PARSNIP: A Connectionist Network that Learns Natural Language

Grammar from Exposure to Natural Language Sentences”. In: Proceedings of the 8th Annual Meeting of the Cognitive Science Society. Hillsdale, NJ: Erlbaum: 106-119.

Karpathy, A. & Fei Fei, L. 2015. “Deep Visual Semantic Alignments for Generating Image Descriptions”. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition: 3128-3137.

_____. & Johnson, J. 2016. “Visualizing and Understanding Recurrent Networks”. Proceedings of ICLR 2016: 1-12.

Marcus, G. F. 1998. “Can Connectionism Save Constructivism?”. Cognition 66: 153-182.

Marr, D. 1982. Vision. San Francisco, CA: W. H. Freeman.

McClelland, J. L. & Rumelhart, D. E. (eds). 1986. Parallel Distributed Processing Vol. 1. Cambridge, MA: MIT Press.

Mikolov, T. et al. 2010. “Recurrent Neural Network Based Language Model”. Proceedings of INTERSPEECH 2010: 1045-1048. 63
Mikolov, T. & Zweig, G. 2012. “Context Dependent Recurrent Neural Network Language Model”. Proceedings of SLT Workshop 2012: 234-239.
Pinker, S. & Prince, S. 1988. “On Language and Connectionism”. Cognition 28: 73-193.
Plaut, D. 2000. “Connectionist Modeling of Language: Examples and Implications”. In: Banich, M. T. & Mack, M. (eds.). Mind, Brain, and Language: Multidsciplinary Perspectives. Mahweh, NJ: Erlbaum: 1-14.
Rogers, T. T. & McClelland, J. 2014. “Parallel Distributed Processing at 25: Further Explorations in the Microstructure of Cognition”. Cognitive Science 38: 1024-1077.
Seidenberg, M. S. & MacDonald, M. C. 1999. “A Probabilistic Constraint Approach to Language Acquisition and Processing”. Cognitive Science 23 (4): 569-588.
Smolensky, P. 1987. “The Constituent Structure of Connectionist Mental States: A Reply to Fodor & Pylyshyn”. Southern Journal of Philosophy XXVI: 137-162.
_____. 1988. “On the Proper Treatment of Connectionism”. Behavioral and Brain Sciences 11: 1-73.
_____. 1999. “Grammar-based Connectionist Approaches to Language”. Cognitive Science 23 (4): 589-613.
Steedman, M. 1999. “Connectionist Sentence Processing in Perspective”. Cognitive Science 23 (4): 615-634.
Tabor, W. & Tanenhaus, M. K. 1999. “Dynamical Models of Sentence Processing”. Cognitive Science 23 (4): 491-515.
Tomasello, M. 2003. Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge, MA: Harvard University Press

**Marco Facchin** (IUSS, Pavia)
*Can "Basic Minds" ever meet content? A problem for Radical Enactivism*

Enactivism argues that cognitive processes are best understood in terms of dynamical agent-environment couplings, rather than inner computations (Varela, Thompson & Rosch 1991, Thompson 2007, Noë 2009, Di Paolo & Thompson 2014, Gallagher 2015). For this reason, enactivism argues that cognitive agents need not internally represent any kind of content. Rather, content is enacted in a lived and embodied process of sense making (Thompson 2011), within which the world is encountered as significant by the agent.

Radical Enactivism, just as regular enactivism, is committed to the antirepresentational claim about the explanation of cognitive processes. However, it adds that minimal cognitive processes performed by basic minds should be understood as contentless (Hutto & Myin 2013); while non-basic ones are just content involving, as opposed to content-based (Hutto & Myin 2017). To account for the emergence of mental content and its involvement in non-basic cognitive processes, Radical Enactivism resorts to Natural Origins of Content (i. e. NOC, see Hutto & Satne 2015): a research program aimed to spell out the origin and functions of content in naturalistic and broadly speaking developmental terms.

In §1, I will introduce NOC and highlight its important merits. Theoretically, NOC's background is Haugeland's (1990) discussion of intentionality, re-interpreted in radical enactivist terms. The thesis that intentional states just are content-involving states is denied, due to its incompatibility with Radical Enactivism. The three competing strategies Haugeland individuated (Neo-Cartesianism, Neo-Behaviorism and Neo-Pragmatism) are then re-interpreted as three distinct, yet mutually supporting, steps to provide a rational account for both contentless and content involving cognition. The proposed account is, broadly speaking, a history of content emergence from basic cognitive interactions, that also describes how content is involved, and manipulated, in non-basic cognitive processes.

The first step NOC takes is that of providing an account of Ur-Intentionality: the minimal form of contentless intentionality that relates basic minds to their environment. The account is achieved by stripping Millikan's teleosemantics (Millikan 1984) of any commitment to content-bearing mediators. This provides a theoretically robust mechanism to account for intentionality in a naturalistic, normative, and yet contentless, fashion.

Once Neo-Cartesian strategies have provided the account of Ur-Intentionality and basic cognition in the first step, Neo-Behaviorist strategies can be used to account for content involving cognition, or so NOC argues. In its second step, Neo-Behavioristic conceptual resources are used to provide a space where to put content in: namely, the linguistic space of accounting for something in terms of reasons. This space is constituted by our ascriptive practices (like Dennett's Intentional Stance, see Dennett 1987) that spell out what accounting for something in terms of reasons amounts to. At the same time ascriptive practices provide a paradigmatic case to clarify how content is involved in non-basic cognition: by being publicly manipulated in a content-sensitive space.

Lastly, Neo-Pragmatist conceptual resources are put in the task of reconstructing a (broadly speaking) developmental history of cognition, highlighting how content has emerged from per se contentless cognitive processes. This means to show how our content involving cognitive niche gets structured by our dynamical interactions. The idea NOC defends in its third step is that such a space has emerged by means of Ur-Intentional mechanisms for social learning and social coordination which allow for the emergence of the local, cultural norms. These norm, in turn, structure our linguistic, content sensitive, cognitive niche.

NOC has at least two important merits that need to be highlighted here: (1) it provides a unique way to answer, at least in principle, the representation hungry problem objection (Clark & Toribio 1994), thus showing enactivism to be explanatory complete. Secondly (2) it allows to integrate contributions from disciplinary fields (such as cognitive archaeology and cognitive anthropology) whose contributions to the sciences of the mind have often been downplayed; thus promising a more nuanced comprehension of the human mind. Both results depend on NOC's acceptance of a historically oriented style of explanation for cognitive processes; and both concur in defining (Radical) Enactivism as a radical explanatory alternative to classical (and non-classical) cognitivism.

As praiseworthy as NOC is, however, I shall, in §2, show two important shortcomings it suffers from. On the one hand (A) the Ur-Intentional account of cognition boils down to a form of behaviorism, for it only accounts for "Perception-Action routines forged […] through a long history of selection of consequences" (Hutto & Myin 2017, p. 115), which just are behavioral responses learned by reinforcement. On the other hand (B) ascriptive practices, NOC's paradigmatic case of content involving cognition, seem insufficient to single-handedly account for all cases of content involving cognition (Hutto & Myin 2017, §7.5, §8.5). NOC, however, does not seem to allow for any kind of expansion (or generalization) of its account.

To see why NOC suffers from (A), consider Garcia Cases (Garcia et al. 1955). They show that some minimal cognitive phenomena, such as a rat learning to avoid poisoned food, are possible even in absence of a long history of selection of sensorimotor routines. Furthermore studies on Observational Learning (Bandura 1965) have shown that at least some sensorimotor routines are actually acquired regardless what consequences they are shown to bring about. Hence some minimal sensorimotor coordination can be acquired without any kind of selection of consequences. This shows that the Ur-Intentional account of basic cognition, at least in the form proposed by NOC, is explanatory incomplete: there are real, and well attested, basic cognitive phenomena it cannot account for.

As for (B), ascriptive practices can indeed provide a paradigm of what content-involving cognition amounts to. But it is clear that such a paradigm falls short in accounting for all the possible cases of content involving cognition, such as judging the two lines of the Müller-Lyer illusion as being equal in length despite their appearance. To expand NOC's initial insight, recent development of Radical Enactivism have revolved around two broader accounts of non-basic cognition. The first one generalizes the model of content involvement suggested by ascriptive practices, proposing a "generalized judicative stance" (Hutto & Myin 2017 § 7.5): an attitude allowing for the manipulation of content in any inferential process. However, this account is vacuous, for such an attitude should be explained by "an organism's interactional history" and thus in Ur-Intentional terms, that exclude, according to Radical Enactivism, any significant involvement of content. The other account (Hutto & Myin 2017 § 8.5) revolves around the adoption of the framework of Material Engagement Theory (Malafouris 2013), according to which content is manipulated through the manipulation of specific content bearing external structures. Such an account may work, but unfortunately Material Engagement Theory is conceptually incompatible with Radical Enactivism. This is due to the fact

that the Material Engagement Theory is a variant of the Manipulation Thesis (Menary 2007) - asserting that some cognitive processes are possible only when a cognitive agent is coupled with a suitable cognitive artifact – that Radical Enactivism explicitly rejects (Hutto & Myin 2013 ch. 7). Despite the surface-level similarity the two theses have, in fact, they show major differences on fundamental matters such as the information-theoretic framework, Darwinian psychological continuity, and the metaphysical commitment to the extended mind thesis.

In §3, however, I will show that the Skilled Intentionality Framework can provide NOC with just the right set of conceptual resources to overcome its difficulties. The Skilled Intentionality Framework is a radical embodied cognitive science (Chemero 2009)-inspired enactive account of cognition, that is best thought as alternative to the Ur-Intentionality framework (Kiverstein & Rietveld 2015).
Being a framework committed to spell out the consequences the Free Energy Principle (Firston 2009, Allen & Firston 2018) has for the philosophy of mind, the Skilled Intentionality Framework comes with an in-build way to get a look into the black box (Bruineberg & Rietveld 2014). This is per se sufficient to avoid behaviorism, and its flawed commitment to explain cognition in terms of only sensorimotor routines acquired in virtue of a long history of selection of consequences. This is also due to the commitment of the Skilled Intentionality Framework to an affordance based kind of psychological explanations (Rietveld et al. 2018). Since affordances are not stimuli, the dynamical coupling an agent undertakes with them cannot be learned just by means of reinforcements.

Furthermore, this commitment to affordances has two important upshots: on the one hand, due to the fact that affordances have in an important sense a perspective quality, the Skilled Intentionality Framework vindicates the original enactivist claim (Bruner 1990 Ch. 1, Varela Thompson & Rosch 1991 Ch. 1) that subjectivity is ineliminable from our psychological theorizing. On the other hand, a commitment to the affordance-based talk makes the framework well poised to expand NOC's account of non-basic cognition (Kiverstein & Riteveld 2018). This is because a commitment to affordances allows for the integration of Material Engagement Theory conceptual resources, since affordance talk is per se committed to Darwinian psychological continuity and the information theoretic approach (Bruineberg et al. 2018); and it is at least metaphysically compatible with the extended mind thesis. Moreover, canonical (socially determined) affordances (Costall 2012) may provide a way to bind even closer Material Engagement Theory and NOC's insistence on the role of social practices in the emergence of content.

On the surface, the relationship between NOC and the Skilled Intentionality Framework seems as idyllic as it can possibly be, for the latter provides the former just the right kind of conceptual resources to overcome its own shortcomings. However, I shall argue in §4 that a Skilled Intentionality-inspired NOC is incompatible with Radical Enactivism. This is due to the fact that exertions of skills, even in basic-mind level cases, do involve some kind of content, since they do respect the relevant constraints by which contentful vehicles are identified (Rowlands 2006, Gallagher 2017 ch. 5). Thus a Skilled Intentionality-inspired NOC is committed precisely to the thesis Radical Enactivism denies: that minimal cognition is contentful. This puts Radical Enactivism in a dire scenario, for it can either (a) abandon NOC, and succumb to the representation hungry problem objection, or (b) accept NOC, ceasing to be radical in any significant sense.

I will conclude arguing that (b) is by far the best option. Accepting that minimal cognition is contentful is not, per se, sufficient to push us into adopting any form of representationalism. This is due to the fact that representationalism and cognitivism require a special kind of content bearing vehicles, located inside the agent and bearing a deeply reconstructive, mirror-of-nature-like, kind of content (Clark 1997 pp. 21-23, Clark 2015); a characterization that skilled engagements are obviously ill-suited to fit. Furthermore, as predictive models of the brain gain in popularity, conceptual radicalism seems to be the position at the losing end of the reflections in philosophy of mind and cognitive science. For predictive models are gaining attention not by being yet another radical break within the mind/brain sciences, but instead by providing a cohesive unitary framework into which many different insights, taken from the whole spectrum of competing research programs of post-classical cognitive science, can be integrated in an unitary, conceptually and empirically well-grounded, framework (Clark 2016).

Micah ALLEN, Karl FIRSTON, 2018 – From Cognitivism to Autopoiesis: Towards a Computational Framework for the Embodied Mind, in Synthese, 195.6.
Albert BANDURA, 1965 – Influence of models' reinforcement contingencies on the acquisition of imitative responses, in Journal of Personality and Social Psychology, 1.6.
Jelle BRUINEBERG, Erick RIETVELD, 2014 – Self-Organization, Free Energy Minimization and the Optimal Grip on a Field of Affordances, in Frontiers in Human Neurosciences, 8.
Jelle BRUINEBERG, et al. – General echological information supports engagements with affordances for "higher" cognition, in Synthese, https://doi.org/10.1007/s11229-018-1716-9
Jerome BRUNER, 1990 – Acts of Meaning, Harvard University Press.
Anthony CHEMERO, 2009 – Radical Embodied Cognitive Science, The MIT Press.
Alan COSTALL, 2012 – Canonical Affordances in Context, in Avant 3.2.
Andy CLARK, 1997 – Being There: Putting Brain, Body and World together again, The MIT Press.
Andy CLARK, 2015 – Predicting Peace: The End Of the Representation Wars, in T. Metzinger, J. Windt (eds.), Open Mind, Frankfurt am Main, The MIND group.
Andy CLARK, 2016 – Surfing Uncertainty, Oxford University Press.
Andy CLARK, Josefa TORIBIO, 1994 – Doing without representing? In Synthese, 101.3.
Daniel DENNETT, 1987 – The Intentional Stance, The MIT Press.
Ezequel DI PAOLO, Evan THOMPSON, 2014 – The Enactive Approach, in L. Shapiro (ed.) - The Routledge Handbook of Embodied cognition, Routledge.
Karl FIRSTON, 2009 – The Free-energy Principle: a rough guide to the brain?, in Trends in Cognitive Sciences, 13.7.
John GARCIA et al. 1955 – Conditioned Aversion to Saccharine Resulting from Exposure to Gamma Radiation, in Science, 122.
Shaun GALLAGHER, 2015 – Do we (or our brains) actively represent or enactively engage with the world? In A. Engel, K. Friston, D. Kragic (eds.) The Pragmatic Turn, The MIT Press.
Shaun GALLAGHER, 2017 – Enactivist Interventions, Oxford University Press.
John HAUGELAND, 1990 – The Intentionality All-Stars, in J. Haugeland (1998), in Having Thought, Harvard University Press.
Daniel HUTTO, Erick MYIN, 2013 – Radicalizing Enactivism, The MIT press.
Daniel HUTTO, Erick MYIN, 2017 – Evolving Enactivism, The MIT Press.
Daniel HUTTO, Glenda SATNE, 2015 – The natural origins of content, in Philosophia, 43,3.
Julian KIVERSTEIN, Erick RIETVELD, 2015 – The primacy of skilled intentionality: on Hutto & Satne's natural origins of content, in Philosophia, 43,3.
Julian KIVERSTEIN, Erick RIETVELD, 2018 – Reconceiving Representation-Hungry Cognition: an Ecological Enactive Proposal, in Adaptive Behavior, 1059712318772778
Lambros MALAFOURIS, 2013 – How Things Shape the Mind, The MIT Press.
Richard MENARY, 2007 – Cognitive Integration, mind and cognition unbound, Palagrave McMillan
Ruth G. MILLIKAN, 1984 – Language, Thought and other Biological Categories, The MIT Press.
Alva NOË, 2009 – Out of Our Heads, Hill & Wong.
Erick RIETVELD, et al. 2018 – Ecological-enactive Cognition as Engaging with a Field of Relevant Affordances: The Skilled Intentionality Framework (SIF), in A. Newen, L. De Bruin & S. Gallagher (eds.), The Oxford Handbook of 4E Cognition, Oxford University Press.
Mark ROWLANDS, 2006 – Body Language, The MIT press.
Evan THOMPSON, 2007 – Mind in Life, Harvard University Press.
Evan THOMPSON, 2011 – Living Ways of Sense-making, in Philosophy Today, 55.
Francisco VARELA, Evan THOMPSON, Eleanor ROSCH, 1991 – The Embodied Mind, The MIT Press.

**Stefano Calboli** (University of Urbino), **Vincenzo Fano** (University of Urbino), **Roberto Macrelli** (University of Urbino)
*The Moral Decoy Effect. Asymmetric Dominance Effect in Morality and Its Political Implications*

One of the heuristics and biases program's line of research is devoted to investigate framing and context effects. Behavioral economics is the most important applications of the heuristics and biases program hence these effects have been investigated to a very lesser extend outside the realm of economic choices. The

present research is meant to cast light on the role of one type of context effect, found in the consumers' choice area, in shaping the human beings' moral judgments, namely the asymmetric dominance effect.
Our investigation of this effect in moral domain exploits questionnaires based on a classical scenario that involve a conflict between moral requirements, the well-known footbridge problem.
The experiment employed a between-subjects design, assigning participants randomly to one of two conditions. Participants assigned to the control group were asked to judge the moral acceptability of pushing and do not pushing the workman on the footbridge in the classic footbridge problem. Otherwise, participants assigned to the treatment group receive a footbridge scenario which contains an added decoy option. The decoy option consists in the act of pushing a workman wearing a light backpack, doing that would allow to avoid the deaths of merely three of the five workmen.
Our main hypothesis is that the presence of the decoy increases the moral acceptability of acting. To confirm our hypothesis would be a nontrivial result for the reason that adding the decoy should be a neutral variation from a moral viewpoint. Unlike the vast majority of the researches on the role of cognitive biases in shaping moral decisions the present research has not been thought-out to support or criticize a specific descriptive moral theory. Instead, we highlight the importance of the kind of findings we have obtained for a public policy issue. Indeed, our results enrich and extend the debate on the legitimacy of exploiting cognitive biases to influence the citizens' choices. Such issues have been extensively addressed concerning the economic choices within the debate around libertarian paternalism and nudges, but the same cannot be said for the use of nudges to address moral choices.

**Robert Chis-Ciure** (University of Bucharest), **Francesco Ellia** (University of Bologna)
*Facing up to the Hard Problem as an Integrated Information Theorist*

## 1. Introduction

Consciousness seems particularly hard to fit into our scientific worldview when we consider its subjective aspect. Neurobiological theories that account for consciousness starting from its physical substrate seem unable to explain the problem posed by experience. Why certain neural processes are accompanied by certain experiential features, while others are not? Why seeing red gives that sensation, while pain a different one? This is the Hard Problem of consciousness and any theory that attempts to explain this phenomenal feature of reality needs to address it. In this contribution we discuss how the Hard Problem affects the Integrated Information Theory (IIT), which is currently regarded as one of the most prominent neurobiological theories of consciousness. We first introduce IIT, then we present our own Layered View of the Hard Problem. We show that, if our analysis of the Hard Problem is correct, then the integrated information theorist has to reject the Hard Problem in its current formulation.

## 2. A Brief on Integrated Information Theory

In its various forms (Tononi 2004, Balduzzi & Tononi 2008, Oizumi et al. 2014, Tononi et al. 2016), Integrated Information Theory presents itself as a theory of consciousness that can quantify the degree and qualify the type of subjective experience that a system exhibits. "Understanding consciousness requires not only empirical studies of its neural correlates, but also a principled theoretical approach that can provide explanatory, inferential, and predictive power" (Oizumi et al. 2014). Such an approach is necessary since the neural and behavioral correlates of consciousness can be insufficient or misleading (e.g. locked-in syndrome), especially for system designs progressively divergent from that of neurotypical adult human cerebral cortex. As the scale of alien-ness increases, the difficulty in determining the presence and character of consciousness in such systems increases exponentially (from brain-damaged patients to babies, animals, and machines) (Tononi & Koch 2015).
IIT suggests a different approach to the problem, as figuring out how the brain can produce consciousness could not only be hard, but almost impossible (Tononi & Koch 2015). Rather than having a bottom-up approach, namely from neurons to consciousness, IIT adopts a top-down approach from phenomenology to the mechanism of consciousness. As such, the theory starts with five axioms that characterize the essential properties of every experience. IIT defines axioms as "self-evident truths" (Oizumi et al. 2014). These axioms are: intrinsic existence, composition, information, integration and exclusion. According to them, consciousness intrinsically exists, i.e. from the point of the system that has it, not that of an external observer; it is structured by many combined features (composition); it is informative, in the sense that every

experience differs from others (information), and integrated, in the sense that is strongly irreducible to non-interdependent components (integration). Finally, consciousness is exclusive, in the sense that any experience excludes all possible others (exclusion).
To each axiom corresponds a postulate. Postulates are defined as "assumptions about the physical world and specifically about the physical substrate of consciousness" and are "inferred" from axioms (Oizumi et al. 2014). The postulate of intrinsic existence states that, within a system, a complex of mechanisms in a state responsible for a given experience must exist intrinsically: it must have a cause-effect power upon itself. The postulate of composition states that, given an integrated system, we should be able to individualize subsystems that have a cause-effect power upon the whole system. The postulate of information states that the system must have a repertoire of activation states of the subsystems that compose the system, which defines the cause-effect power of such states and differentiates it from other possible states. The postulate of integration states that the cause-effect structure specified by the system must be irreducible to that specified by independent subsystems. The postulate of exclusion states that the cause-effect structure specified by the system is maximally irreducible intrinsically, namely that the cause-effect structure specified by the system must be specified over a single set of elements and spatio-temporal grain (Oizumi et al. 2014, Tononi 2015, Bucci & Grasso 2017).
Alongside these five axioms and postulates, IIT posits an ontological identity between the phenomenological properties of experience and causal properties of a physical system:

'The maximally irreducible conceptual structure (MICS) generated by a complex of elements is identical to its experience. The constellation of concepts of the MICS completely specifies the quality of the experience (its quale "sensu lato" (in the broad sense of the term)). Its irreducibility ΦMax specifies its quantity. The maximally irreducible cause-effect repertoire (MICE) of each concept within a MICS specifies what the concept is about (what it contributes to the quality of the experience, i.e. its quale "sensu stricto" (in the narrow sense of the term)), while its value of irreducibility φMax specifies how much the concept is present in the experience.' (Oizumi et al. 2014).

Therefore, IIT accounts for consciousness by positing a fundamental identity between an experience and a conceptual structure that is maximally irreducible intrinsically (MICS) (Tononi 2015). An important point is that this identity is not between consciousness and physical substrate. Instead of equating experience with some physical processes, IIT identifies it with the conceptual structure that is specified by a system's complex of elements in a state. The physical configuration and dynamics of elements in a complex specify a conceptual structure that is an experience: the properties of the experience map directly unto properties of the relevant conceptual structure. The presence, quantity and quality of an experience is given by the conceptual structure and the inherent relations between the concepts comprising it. "An experience is a 'form' in cause-effect space" (Tononi 2015). In other words, consciousness is how an integrated system exerts cause-effect power upon itself (or intrinsically), independent from an extrinsic observer.

## 3. The Layered View of the Hard Problem

The joint endeavor of philosophy and cognitive sciences to explain this most intimate and yet elusive phenomenon of consciousness has been permeated by a methodological distinction between easy problems and the Hard Problem (HP) of consciousness (Chalmers 1995/2010). This distinction can be prima facie understood as a difference in the explanations needed to account for their respective explananda. On one hand, the easy problems are vulnerable to explanations in terms of structural configuration and functioning in physical systems, the kind of explanations obtained via the methods of natural sciences, including life sciences (e.g. cognitive neuroscience). This type of explanation is called reductive, since it implies an account of the explanandum in other terms (e.g. terms about physical processes). On the other hand, the HP of consciousness is resistant to such methods, requiring instead a nonreductive explanation, where consciousness itself is taken as fundamental, i.e. not explainable in more basic terms.
But what are the easy problems and the HP? According to Chalmers (1995/2010, 1996, 2004/2010, 2018), the easy problems are those of explaining various mental functions, like attention, perceptual integration, conscious access, reportability, memory, and others. In contrast, the HP is that of explaining what it is like to be us (Nagel 1974), i.e. phenomenal conscious (or conscious/subjective experience). For any system endowed with consciousness, there is something it is like to be that system; accordingly, for any conscious

mental state, there is something it is like to be in that state. This phenomenal character constitutes the explanandum of the HP.
However, this is incomplete. What justifies the distinction between the explanations required? Is there a reason why phenomenal consciousness cannot be explained in terms of physical processes? This cannot just be accepted dogmatically without justification. As Chalmers (1994/2010) puts it, it is a conceptual fact that mechanistic explanations of physical sciences are insufficient to explain experience but are adequate to account for the easy problems. To avoid pushing the dogmatic stance a level up (or down), a further justification is needed. The justification for this conceptual fact is given by another conceptual fact: the conceptual coherency of a scenario where, given any physical process, it could be instantiated in the absence of experience. In principle, we could conceive of any physical process that is put forward as the basis of consciousness as being instantiated without any phenomenal aspect at all. This conceptual coherency fact justifies the claim that the HP is impregnable to methods employed to explain the easy problems.
If the preceding analysis is correct, then HP has a layered structure. We propose the Layered View of the Hard Problem, according to which there is a phenomenal layer and a further conceptual layer that together constitute HP. The phenomenal layer is captured by the Nagelian fact that there is something it is like to be us. Call this the Monolayered Hard Problem (MHP).

(MHP) There is something it is like to be us and we need to explain this fact.

MHP makes experience an explanandum in its own right. Furthermore, this fact requires no justification, since it is something we are directly acquainted to. In other words, it is a basic fact.

'To generate the hard problem of consciousness, all we need is the basic fact that there is something it is like to be us.' (Chalmers 2018, 49-50)

We partially disagree with Chalmers on this point. The fact of experience is sufficient only for MHP. HP as Chalmers understands it requires more work. The conceptual layer is given by the fact that conceivability scenarios are coherent, which provides justification for the fact that experience cannot be explained in terms of structure and function. Call this the Double-Layered Hard Problem (DHP) .

(DHP) There is something it is like to be us and we need to explain this fact. Since conceivability scenarios are coherent, a mechanistic explanation in terms of physical processes is insufficient for this, therefore we need an alternative explanation.

Obviously, DHP contains MHP, but adds further epistemic claims about how an explanation should look like. The HP of 'Why does physical processes give rise to experience at all?' requires this layered view.

'For any physical process we specify there will be an unanswered question: why should this process give rise to experience? Given any such process, it is conceptually coherent that it could be instantiated in the absence of experience. It follows that no mere account of the physical process will tell us why experience arises. The emergence of experience goes beyond what can be derived from physical theory.' (Chalmers 2010, 14)

Construed like this, HP is part of a more comprehensive system of metaphysical (e.g. zombies, inverts) and epistemological (knowledge, epistemic asymmetry) arguments against a materialist theory of consciousness in terms of some physical process. However, as our analysis shows, the problem itself is generated via argumentative mechanisms within the system, but we will not explore this aspect here. A brief remark though: any complete theory of subjective experience must either accept the constraints imposed by this argumentative system or dismantle it piece by piece.

## 4. IIT and the Hard Problem

Now the question is: does IIT address HP? If so, which of them? We will argue that IIT indirectly gives an explanation to MHP, and both directly and indirectly denies that there is a DHP. Insofar indirect refutation is concerned, there are multiple ways to interpret Tononi's claims along these lines.
We could take the integrated theorist as explicitly claiming that HP (i.e. DHP) does not arise for a theory like IIT. Tononi (2015) seems to acknowledge that this is an obstacle only for 'bottom-up' theories, i.e. those

attempting to infer the existence of consciousness from physical processes. Only when a theory posits some physical mechanisms and then tries to get to conscious experience DHP arises. By contrast, IIT proceeds in a top-down fashion: by capturing the essential properties of experience in its axioms, the theory infers the postulates describing the physical properties that a system must exhibit in order to be conscious. IIT makes the subjective character an assumption ("axiom") rather than an explanandum. In this sense, one could make a case that IIT does not directly address HP because there is no such problem for a theory of its type. However, one could also argue that, by indirectly accounting for experience, which is taken as fundamental, is a way of explaining it, even though it is not further reducible to something else. In this further sense, IIT indirectly addresses the Monolayered Hard Problem (MHP). Thus, a first conclusion is that IIT directly denies that there any DHP for a theory like it to address, but indirectly attempts to solve MHP.

More can be said about the relation between the integrated information view of consciousness and DHP when considering conceivability scenarios. Note that DHP's conceptual layer presupposes the conceivability of zombies or inverts. These are entities that are "molecule for molecule" (Chalmers 1996) replicas of entities in our world, which nonetheless lack or have different subjective experiences. These are known as 'philosophical' zombies. If such entities are conceivable, then they are metaphysically possible; if they are metaphysically possible, then consciousness is nonphysical; ergo materialism is false.

There is an important difference between 'philosophical' and 'perfect'/'true' zombies (Oizumi et al. 2014, Tononi 2015). One of the corollaries of IIT is that sheer functional complexity does not entail consciousness. There are sophisticated, yet unconscious systems (e.g. those displaying a feed-forward architecture) that can perform functions identical with those performed by a conscious system having high integrated information. This marks the difference in zombies: IIT's 'perfect' zombies are not physically identical to their counterparts, only functionally identical. However, 'philosophical' zombies needed both structural and functional identity – zombies were perfect replicas. A possible interpretation here is that, if further pressed, Tononi would reject the conceivability step: according to IIT, 'philosophical' zombies are not conceivable, only 'perfect' ones are. A considerably more speculative remark could be made that an integrated theorist would take nomic or natural possibility as the fundamental modality rather than the metaphysical or logical one, but this is not obvious from the present state of IIT. Thus, a second conclusion is that, under a certain interpretation, IIT indirectly denies the conceivability of zombie scenarios, thereby denying the justification for the second layer, thus DHP itself.

Further, there is another possible interpretation for the indirect rebuttal of DHP by IIT via negating the possibility of conceivability scenarios. This has to do with the fundamental identity posited by the information integration theorist between an experience and a maximally irreducible conceptual structure (MICS). Tononi (2015) is more explicit in this regard: "if the postulated identity […] is true, a system of elements in a state that specifies [a] conceptual structure has the corresponding experience necessarily and cannot be a zombie". If the argument is sound and thus the identity true, then, by necessity of identity, a high Φ physical system in a state specifying a conceptual structure has the corresponding conscious experiences necessarily. Therefore, zombies are perhaps conceivable, but not metaphysically possible. As such, a third conclusion is that, under a more evident interpretation, IIT flatly denies the possibility of conceivability scenarios, thus the justification for the second layer, ergo DHP itself.

## 5. Conclusion

We have shown that, if the layered analysis of the Hard Problem that we proposed is true and the standard Hard Problem is the Double-Layered Hard Problem, then the relation between it and Integrated Information Theory depends on the theory's stance on conceivability scenarios. Firstly, regarding the Monolayered Hard Problem, IIT takes the road of nonreductive fundamental explanation and thus can be said to indirectly attempt to solve it – explaining how come, but not why. Secondly, IIT can be said to directly deny that there is a Double-Layered Hard Problem for it to answer to due to its methodological choice of a top-down approach. Thirdly, IIT indirectly denies that there is in general a Double-Layered Hard Problem, either by allowing only for functionally, but not physically identical zombies (no conceivability), or by holding the necessary identity between an experience and its MICS (no possibility). If our argument is sound, then IIT and HP in their current state cannot be both true: one of them needs to be revised or rejected. Our bet is on the rejection of the HP, but this needs further argument beyond the fact that it is incompatible with the best player in the game of Consciousness.

Balduzzi, D., & Tononi, G. (2008). Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework. PLoS Comput Biol, 4(6), e1000091. http://dx.doi.org/10.1371/journal.pcbi.1000091
Bucci, A., & Grasso, M. (2017). Sleep and Dreaming in the Predictive Processing Framework. In T. Metzinger & W. Wiese (Eds.), Philosophy and Predictive Processing. Frankfurt am Main: MIND Group.
Chalmers, D. (1996). The Conscious Mind. Oxford Paperbacks.
Chalmers, D. (2004/2010). Consciousness and Its Place in Nature. In The Character of Consciousness. New York: Oxford University Press.
Chalmers, D. (1995/2010). Facing Up to the Problem of Consciousness. In The Character of Consciousness. New York: Oxford University Press.
Chalmers, D. (2018). The Meta-Problem of Consciousness, 25(9-10), 6–61. Journal of Consciousness Studies.
Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. Nat Rev Neurosci, 17(5), 307-321. http://dx.doi.org/10.1038/nrn.2016.22
Nagel, T. (1974). What is it like to be a bat?, 165-180. http://dx.doi.org/10.1017/cbo9781107 341050.014
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), e1003588. http://dx.doi.org/10.1371/journal.pcbi.1003588
Tononi, G. (2004). An information integration theory of consciousness. BMC Neurosci 5: 42.
Tononi, G. (2015). Integrated information theory. Scholarpedia, 10(1), 4164. http://dx.doi.org/10.4249/scholarpedia.4164
Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere?. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1668), 20140167-20140167. http://dx.doi.org/10.1098/rstb.2014.0167
Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. Nat Rev Neurosci, 17(7), 450-461. http://dx.doi.org/10.1038/nrn.2016.44

## SECOND SESSION: General Philosophy of Science

**Eugenio Petrovich** (University of Siena)

*Bridging the Gap between General Philosophy of Science and Scientometrics: Towards an Epistemological Theory of Citations*

Citations are a crucial aspect of contemporary science (Biagioli, 2016). Citation-based indicators such as the JIF (Journal Impact Factor) are commonly employed by scientists to choose the publication venue for their articles, whereas indicators such as the h-index are used (and frequently misused) by university administrators to monitor and evaluate the research performance of academics (Biagioli, 2018; Rijcke, Wouters, Rushforth, Franssen, & Hammarfelt, 2016). The implementation of performance-based research evaluation systems in many European and extra-European countries has further speeded up the proliferation of metrics in which citations are often a crucial component (Whitley & Gläser, 2007; Whitley, Gläser, & Engwall, 2010). Thus, scientometrics and bibliometrics, the disciplines that investigate the quantitative dynamics of citations and articles in science, have risen from relative obscurity as statistical provinces of information science to playing a major, and often much criticized, role within the social and political processes of the science system (De Bellis, 2014; Mingers & Leydesdorff, 2015).

Unfortunately, citations have mostly escaped the attention of philosophers of science, maybe because they are relegated to the "context of discovery" of science (Leydesdorff, 1998). The discussion around a comprehensive theory of citations in science has witnessed important contributions by scientometricians and sociologists of science, but not by philosophers of science (searching on the database Web of Science for the articles appeared in leading philosophy of science journals that address scientometric topics in the interval [1980-2018], I found only one article which discusses citation analysis in the light of structuralism: (Massey, 2014). [Search date: 5.12.2018]).

This paper aims at beginning to close the gap between scientometrics and philosophy of science, first by advancing an epistemological theory of citations as a bridge between the two fields, and then by highlighting several scientometric phenomena that are in need of an epistemological interpretation.

In the first part of the paper, I will present the two main competing theories of citation developed in the sociology of science: the normative theory, inspired by the normative sociology of science of Robert K. Merton, and the socio-constructivist theory, grounded in the social constructivist approaches in the sociology of science (Bruno Latour, Karin Knorr-Cetina, David Bloor amongst others) (Bornmann & Daniel, 2008). I will show that, even if these theories advance conflicting claims about the role of citations in science, they share the same explanandum as a target: they both assume that the key aim of a theory of citation is to uncover the motivations that scientists have for citing. Thus, both theories can be considered as theories of the citing behavior.

I propose to shift the focus from the behavior of the scientists to the epistemological function of citations within the scientific documents (such as books, papers, and so on) in which they appear (Petrovich, 2018a). The basic idea is to consider the citations as information channels between the citing document and the cited texts (Amsterdamska & Leydesdorff, 1989; Small, 1978). Thus, a network of scientific documents connected by citations (the so-called citation network) can be considered as an information structure in which scientific information flows. In this way, the focus is no more on the motivations of scientists for citing (sociological perspective), but on the dynamic of scientific information that is made visible by citations (epistemological perspective).
I will claim that the transformation of scientific information into scientific knowledge can be studied by analyzing the dynamics of the citation network of scientific documents (Petrovich, 2018c, 2018b). To do that, I will draw on the Khunian distinction of pre-paradigmatic, normal, and revolutionary science, and I will argue that different citation structures characterize each of these phases. Thus, I will argue that citation analysis can be used to investigate the transition between a pre-paradigmatic to a normal-scientific period and to shed light on the phase in which the mere accumulation of information converts into the accumulation of scientific knowledge.

In the last part of the paper, I will sketch a research program at the crossroad between general philosophy of science and scientometrics. In particular, I will highlight the need to provide an epistemological interpretation of the following scientometric phenomena:

a) The skewed distributions of citations and papers: few authors collect most of the citation and produce most of the papers (Katz, 1999).

b) The aging of the scientific literature: citations are not distributed homogeneously in time, but tend to accumulate close to the year of publication of the citing text (the so-called immediacy effect): how is the aging process connected to the growth of knowledge? (Larivière, Archambault, & Gingras, 2008);

c) Visualizations of the citation network of science (the so-called science maps): what is the meaning of the clusters of documents we find in such visualizations? Should they be considered Kuhnian paradigms, Lakatosian research programs, or something else (Chen, 2013; Massey, 2014; Small, 2003; Waltman & van Eck, 2014)?


Amsterdamska, O., & Leydesdorff, L. (1989). Citations: Indicators of significance? Scientometrics, 15(5–6), 449–471. https://doi.org/10.1007/BF02017065

Biagioli, M. (2016). Watch out for cheats in citation game. Nature, 535(7611), 201–201. https://doi.org/10.1038/535201a

Biagioli, M. (2018). Quality to Impact, Text to Metadata: Publication and Evaluation in the Age of Metrics. KNOW: A Journal on the Formation of Knowledge, 2(2), 249–275. https://doi.org/10.1086/699152

Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. Journal of Documentation, 64(1), 45–80. https://doi.org/10.1108/00220410810844150

Chen, C. (2013). Mapping scientific frontiers: the quest for knowledge visualization (Second Edition). London: Springer.

De Bellis, N. (2014). History and Evolution of (Biblio)Metrics. In Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact (pp. 23–44). London: MIT Press.

Katz, J. S. (1999). The self-similar science system. Research Policy, 28(5), 501–517. https://doi.org/10.1016/S0048-7333(99)00010-4

Larivière, V., Archambault, É., & Gingras, Y. (2008). Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). Journal of the American Society for Information Science and Technology, 59(2), 288–296. https://doi.org/10.1002/asi.20744

Leydesdorff, L. (1998). Theories of citation? Scientometrics, 43(1), 5–25. https://doi.org/10.1007/BF02458391

Massey, T. (2014). Structuralism and Quantitative Science Studies: Exploring First Links. Erkenntnis, 79(S8), 1493–1503. https://doi.org/10.1007/s10670-013-9579-4

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. European Journal of Operational Research, 246(1), 1–19. https://doi.org/10.1016/j.ejor.2015.04.002

Petrovich, E. (2018a). Accumulation of knowledge in para-scientific areas: the case of analytic philosophy. Scientometrics, 116(2), 1123–1151. https://doi.org/10.1007/s11192-018-2796-5

Petrovich, E. (2018b). Forms, Patterns, Structures. Citation Analysis and the History of Analytic Philosophy. Journal of Interdisciplinary History of Ideas, 7(13), 1–21. http://dx.doi.org/10.13135/2280-8574/2843

Petrovich, E. (2018c). Reply to Wray. Scientometrics. https://doi.org/10.1007/s11192-018-2871-y

Rijcke, S. de, Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—a literature review. Research Evaluation, 25(2), 161–169. https://doi.org/10.1093/reseval/rvv038

Small, H. (1978). Cited Documents as Concept Symbols. Social Studies of Science, 8(3), 327–340.

Small, H. (2003). Paradigms, citations, and maps of science: A personal history. Journal of the American Society for Information Science and Technology, 54(5), 394–399. https://doi.org/10.1002/asi.10225

Waltman, L., & van Eck, N. J. (2014). Visualizing bibliometric networks. In Measuring scholarly impact: Methods and practice (pp. 285–320). Springer.

Whitley, R., & Gläser, J. (Eds.). (2007). The changing governance of the sciences: the advent of research evaluation systems. Dordrecht, the Netherlands: Springer.
Whitley, R., Gläser, J., & Engwall, L. (Eds.). (2010). Reconfiguring knowledge production: changing authority relationships in the sciences and their consequences for intellectual innovation. Oxford ; New York: Oxford University Press.

**Alejandra Casas Munoz** (University of Bristol)
*An Inferential Conception of Scientific Explanation*

## 1. Introduction

This paper aims to provide an account of scientific explanation that emphasizes the role of inferential considerations in explanatory contexts. The account should accommodate two crucial desiderata of a scientific explanation: to make sense of the role played by theories in explaining the phenomena, and identify the kinds of things that explain the phenomena in question. We expect that a proper account of scientific explanation should address both desiderata.
To motivate the proposal, we consider two accounts: one that emphasises the role of theories in explanatory contexts, thus yielding theoretical explanations (Hughes [2010]), and another account that highlights the role of structures, understood as objective features of the world, in producing the relevant phenomena, thus yielding structural explanations (McMullin [1978]).
We argue that, on their own, none of these proposals provides a complete and proper account of scientific explanation, since they miss part of the desiderata for such explanations. On the one hand, theoretical explanations overlook the role of structures and their relations, as objective features of the world, in the explanation of the phenomena. Structural explanations, on the other hand, despite being orientated in the right direction, wobble at the crucial distinction between the phenomena (objects, structures, relations, and processes) in the world, and the theories formulated to frame and describe them. What is needed is an account that integrates both proposals.
However, an additional constraint also needs to be met. Both theoretical and structural explanations, given the ways in which they are formulated, assume that the phenomena under consideration have well-defined identity conditions. As a result, the resulting accounts are unable to accommodate phenomena at the nanoscale (specifically those at the lower level of this scale) and quantum phenomena, given that, on a significant interpretation, they lack such identity conditions (French and Krause [2006]). A proper account of scientific explanation needs to accommodate phenomena independently of where they stand regarding their identity.
To address of these constraints, we offer the inferential conception of scientific explanation as an alternative: it extends the inferential conception of the application of mathematics (Bueno and Colyvan [2011], and Bueno and French [2018]) to explanatory contexts and has the resources to implement the needed integration. It provides a framework that brings together the positive features of the theoretical and the structural accounts while avoiding their shortcomings.

## 2. Theoretical Explanation

Central to a theoretical explanation is to identify a feature of the world to be explained and to present a theoretical model of the relevant part of the world to explain it. This is achieved by establishing that the feature of the world, which is not explicit in the model, in fact, corresponds to a specific feature of the model (Hughes [2010], p. 210). In other words, a theoretical explanation is mainly guided by theories and models mediated by those theories. R.I.G. Hughes distinguishes three types of models: constitutive, analogue, and foundational. Their common aspect is that they all represent and such representation is mediated by theories (Hughes [2010], pp. 217 and 229).
Hughes, however, understands representation, in a very specific way, in terms of his DDI account, which highlights that representation is a matter of denotation, demonstration, and interpretation (Hughes [1997]). Crucial for this account is denotation, which is understood model-theoretically as particular functions from the objects in the domain to the relevant model, and such functions, formulated in set-theory, require the identity of the relevant objects. Denotation thus involves identifying a system, whose parts have well-defined

identity conditions, and creating a theoretical model in which the relevant features of the system are formulated.
In order to provide an explanation, Hughes states, theoreticians, provide a representation of the phenomena being studied. It is ultimately in terms of the relevant models that the explanation of the phenomena is achieved. This means that in order to be explained, parts of a system need first to be denoted, but this requires, we note, that the system's parts have identity conditions. As it turns out, this is not the case for the lowest-level phenomena at the nanoscale and quantum phenomena more generally, on a significant interpretation. Thus, it is unclear that one can provide theoretical explanations of such phenomena.
Furthermore, the theoretical model that represents, and which on the theoretical explanation account is necessary to explain, is related to the phenomena or empirical set up only via denotation (understood model-theoretically). But this is too narrow a conception, and it overlooks additional mappings that can be established between the empirical setup and the model, such as isomorphism, homomorphism, partial isomorphism, partial homomorphism, etc. (Bueno [2006]).
Finally, Hughes's account addresses the issue of the evaluation of the explanation. On his view, the evaluation is determined by pragmatic terms (Hughes's understands explanation as a perlocutionary act); in particular, the explanation needs to be appropriately delivered to the intended audience. Moreover, a theoretical explanation depends on how adequate the theories it invokes ultimately are. While the adequacy of the relevant theories is certainly an important component in the assessment of an explanation, it does not seem to us that the accuracy of an account of explanation should depend on how it is delivered to a target audience, since this is a matter not of the adequacy of the explanation per se, but of its accessibility to the relevant community.
Arguably what ultimately explains the phenomena are relevant features of the world (structures, objects, relations, events). Thus, the structural explanation comes to the scene.

## 3. Structural Explanation

Structural explanations invoke relevant structures, understood as objective features of the world, to explain the behaviour and properties of the phenomena (McMullin [1978]). Structures are understood as a "set of constituent entities or processes and the relationships between them" (McMullin [1978], p. 139). Structural explanations are causal whenever the structures are the cause of the phenomena under study.
On the structural explanation view, what explains is not the theory but the relations among the objects and structures themselves. The theory will bring propositions that describe those objects and structures, but the explanatory work is based on the very phenomena produced by the actual relations among them. For example, it is well known that the properties of gold nanoparticles vary depending on the scale: At 2-3 nm, gold is considerably magnetic and a good catalyst. Under 2-3 nm, gold nanoparticles (AuNPs) are no longer metallic and turn into insulators; furthermore, the cubic structure of gold at a larger scale becomes icosahedral at that size. The explanatory work invokes properties that gold has at a particular scale and not at other scales. The explanatory resources come from the phenomena rather than the theory describing the phenomena.
However, despite the emphasis on the role of structures, as objective features of the world, theoretical considerations are still raised by the structural explanation, but in a way that blurs the distinction between structures, as aspects of the world, and models, as representational devices. As McMullin notes, "the structure underlying such an explanation is often called a physical (or a theoretical) 'model', since the explanation is a hypothetical one" (McMullin [1978] p. 139). By considering the structures underlying an explanation as a theoretical model, the understanding of structures as objective traits of the world rather than a theoretical description of it and as the main source of explanatory power is ultimately undermined.
Moreover, structural explanations face an additional challenge. Structures are also presented as a set of entities or processes and their relations (McMullin [1978] p. 139). But sets, as formulated in classical set theories, satisfy the extensionality axiom (according to which the identity of sets is determined by the identity of their members), which requires that sets can only be formed by objects that have well-defined identity conditions. Hence, structural explanations face the same limitation we identified in theoretical explanations above, namely, they are unable to accommodate explanations involving objects, such as quantum particles and those at the lowest level of the nanoscale, for which identity may not be defined. This provides a significant shortcoming for the proposal.

What is needed is an account that combines both the positives features of structural and theoretical explanations (the former's identification of objective features of the world as the source of explanatory power and the latter's use of theoretical descriptions in the conceptualization of the relevant phenomena), but without the commitment to the identity of the objects and structures invoked in explanatory contexts. What is needed is a proposal that: (a) accounts for the role that theoretical considerations play in scientific explanations without undermining the role of objective structures; (b) identifies the role of structures as the source of explanatory power without disregarding the role of theoretical considerations in the description of the relevant phenomena, and (c) makes room for cases where explanations do not require the identity of the objects whose behaviour is being explained. To address these requirements, we offer the inferential conception of scientific explanations.

4. Scientific Explanation: An Inferential Conception

We offer an extension of the inferential conception of the application of mathematics (Bueno and Colyvan [2011] and Bueno and French [2018]) to scientific explanations. The account invokes structures and relations found in the world as the source of explanation and draws inferences in the model to conceptualize the relevant phenomena; the outcomes of the relevant inferences are then interpreted back in the world. In this way, the inferential conception of scientific explanation integrates both theoretical explanations and structural explanations.
In Hughes's and McMullin's cases we have similar concerns considering the features of the lowest-level nanophenomena. This account of structural explanations goes further of the requirement of identity conditions. It does not require looking at the structures as a domain of objects and family relations; they are not required to be sets. The inferential conception of scientific explanation will allow talking about atoms, electrons, nanoparticles, etc.
Thus, to evaluate our account of explanation, we examine cases of explanations in nanoscience research, involving, for instance, gold nanoparticles, molecular machines, or DNA replication. They provide us with an opportunity to determine how the inferential conception applies not only to structures that can be denoted (such as those placed at the upper level of the nanoscale), and which, thus, have identity conditions, but also to structures that do not have such conditions, such as those at the lowest level of the nanoscale or at quantum level.

Bueno, O. [2006]: "Representation at the Nanoscale", Philosophy of Science 73(5), 617-628.
Bueno, O., and Colyvan, M. [2011]: "An Inferential Conception of the Application of Mathematics", Noûs 45, 345-374.
Bueno, O., and French, S. [2018]: Applying Mathematics: Immersion, Inference, Interpretation. Oxford: Oxford University Press.
French, S., and Krause, D. [2006]: Identity in Physics: A Historical, Philosophical, and Formal Analysis. Oxford: Oxford University Press.
Hughes, R. I. G. [1997]: "Models and Representation", Philosophy of Science 64(4), S325-S336. (Reprinted in Hughes [2010a].)
Hughes, R. I. G. [2010]: "Theoretical Explanation", in Hughes [2010a].
Hughes, R. I. G. [2010a]: The Theoretical Practices of Physics: Philosophical Essays. Oxford: Oxford University Press.
McMullin, E. [1978]: "Structural Explanation", American Philosophical Quarterly 15, 139-147.
Roduner, E. [2006]: "Size Matters: Why Nanomaterials are Different", Chemical Society Reviews 35, 583-592.

**Alberto Corti** (University of Urbino)
*Scientific Realism without Reality? What remains when metaphysics is left out*

There are two broad positions that philosophers take regarding science: scientific realism and scientific antirealism. These approaches can be articulated in many different ways, leading to a plethora of different positions so that, when someone calls herself a realist or an antirealist, it can be difficult to figure out what

exactly she means. Our aim is to propose a different way of framing the debate, by disentangling several different questions that, we will argue, are best kept separate.

## 1. Entangled Debates: Metaphysical and Scientific Realism

Scientific realism, as it is typically presented (for example in Psillos, 2005), is thought to involve a commitment to the reality of the external world, that is to say, an objective and mind-independent reality. In other words, many people assume that a necessary component of any scientific realist position is the claim that there is a world that in no way depends for its existence on human beings or their activities. We argue that, on the contrary, this is not the case. It is perfectly consistent to be a kind of scientific realist without the metaphysical commitment.

We first outline the space of logical possibilities that can be occupied by combinations of positions regarding metaphysical realism and scientific realism. If the debate on metaphysical realism (i.e. the question of whether the world is as it is, independently of how human beings take it to be) is independent from the debate on scientific realism (a possible construal of which is the question of whether scientific unobservable entities exist (Alai, 2009)), then each combination of metaphysical realism (henceforth 'MR') or metaphysical antirealism ('MA') with either scientific realism ('SR') or scientific antirealism ('SA') should be possible. Of course, some combinations will sound more plausible than others, but for our aims we need only to show that each combination can describe at least one coherent view, clearly demarcated from the rest. We acknowledge connections between the intuitions involved in both debates on metaphysical and on scientific realism, while maintaining that they are logically separable.

We have four possible combinations: MR and SR, MR and SA, MA and SA, and MA and SR. In the literature, the first three of these have been widely acknowledged. The combination of MR and SR is classical scientific realism, which covers the majority of its contemporary forms. MA combined with SA results from following extreme skepticism to its limit, and are what forms of solipsism or idealism that are suspicious of science would accept; MR and SA combined results in the instrumentalist view, according to which scientific theories are just useful tools for making successful predictions. The combination of MA with SR is not often dealt with explicitly in the literature, and could appear at first sight to be highly counter-intuitive.

The combination of MA with SR, then, is the corner of our logical space that we need to show can also be occupied by a coherent view. Indeed, it could be taken to describe what a form of epistemic structural realism, perhaps of a neo-Kantian flavor, is claiming (for example: Massimi, 2011). It could also provide a way to understand perspectival realist positions. No matter what the commitments of epistemic structural realism or perspectival realism might be, we still think that the combination of MA with SR results in an attractive approach to these debates. With broad brush strokes, we show how such a position can be imagined and the differences that this would have with full-blown anti-realism, i.e. MA combined with SA. The purpose of focusing on this distinction is to show what the disagreement between scientific realists and anti-realists amounts to, in very general terms.

A scientific realist ought to grant the same ontological status to scientific entities as to whatever else is considered 'real' (such as everyday objects like tables and chairs), but a different status than to what is considered 'unreal' (such as perhaps the content of dreams and illusions). Of course in the case of hardcore reductionists, scientific entities are more real than everyday objects, or perhaps even the only entities that can properly be called 'real'. We do not wish to engage in debates on reductionism here - we take it that the reality of everyday objects would not be undermined if they were reducible to scientific entities. Nonetheless, to say that scientific entities are real in the sense of being non-illusory does not entail a commitment to the external world. That is, we can define contrast classes of entities, to motivate various forms of scientific realism. These contrast classes provide us with a sense of the reality of scientific entities which is not lost by removing the metaphysical commitment. Such distinctions may themselves be considered as objective, for example, in the sense of being the same for all agents at all times.

Our primary motivation for removing the metaphysical claim as a necessary condition for scientific realism is to encourage focus on other, arguably more interesting and decisive, aspects of the debate, so as to avoid contamination from extreme skepticism and purely speculative metaphysics. Moreover, an analysis of how scientific enquiry is conducted makes our way of separating these debates look plausible. Our discussion here is intended to be very general; different areas of science will be more or less heavily engaged in varying combinations of the processes described. It is of no doubt that empirical data are of vital importance to the

scientific enterprise. Scientists consider collections of data of various forms, model regularities exhibited by this data, and attempt to produce explanations and predictions using deductive reasoning based on these models. So, a position in the scientific realism debate ought to take due consideration of the nature of empirical data, of mathematical models and of the processes involved in generating predictions and explanations.

It is commonly assumed that empirical data must be attributed to a mind- independent, perspective-independent, external world. That is to say, the metaphysical claim mentioned above is a widespread assumption, however it is one that we believe not to be necessary for elucidating scientific realist positions. Scientific realists often claim that true scientific theories are objective, in a robust sense. Even though the existence of an external world would provide an indubitably strong sense of objectivity, we are inclined to think that it is not the only way for achieving it. Objectivity can be achieved in the sciences without this assumption, by means of the objectivity of deductive reasoning, empirical support, or through the objectivity of certain 'theoretical virtues' employed in scientific research (see Agazzi (2014), especially ch. 2). This results in the possibility for defining positions which are in a certain sense realist about science yet antirealist or agnostic about the external world. We do not endorse such a view, but only intend to show that it is tenable and, therefore, the debates on scientific realism and metaphysical realism can be separated.

Our analysis takes into account the interconnectedness of empirical data and developments in modelling techniques and in pure mathematics. We briefly present a relational view of data (similar to that put forward by Sabina Leonelli, 2018), where the scientific process allows for the enlargement and gradual alteration of the set of available empirical data, and the subset of that data that we take to be reliable changes as theories develop. The same goes for the mathematics used in modelling. None of these processes, we argue, require or entail that phenomena exist in an external, mind-independent reality. Put differently, the scientific enterprise itself is neutral on the metaphysical claim: the phenomena studied by scientists may or may not come from a mind-independent world. There are no parts of scientific practice, as we see it happening, that either support or refute the reality of a mind-independent external world.

The arguments typically presented in scientific realism debates, such as pessimistic meta-induction and no-miracles, can also be thought to have a bearing only on the part of the debate that is relevant to science, and not on the metaphysical claim we have sectioned off. Pessimistic meta-induction compares past with present epistemic states in the sciences, and makes an inductive inference to future states: our best scientific theories of the future will probably contradict those of the present, therefore, so the argument goes, our current best theories are probably wrong (or at least will be considered wrong according to the standards set by the scientists of the future). Notice that the comparison is between our best science at different times, and at no point does the argument need to make reference to an external world to carry its force. The argument is entirely neutral on the metaphysical claim: if there is an external world, current science probably describes it incorrectly, and if there is not, then science does not achieve its required objectivity and is incorrect in this sense. The same goes for the no-miracles argument. Very briefly, this argument rhetorically asks the anti-realist, "if science is not correct, why is it so successful for making predictions?" Again, correctness need not be understood as corresponding to a mind-independent, external world. The no-miracles argument, at best, provides reasons for taking unobservable scientific entities to be as real as other entities that we consider real (for example everyday objects); it in no way helps us to decide in what sense they are real (i.e. if they really belong to a mind-independent external world).

To summarize so far, we show that the metaphysical debate about the reality of a mind-independent external world is separate and should be dealt with separately from the debate on scientific realism. Firstly, all four possible combinations of MR/MA with SR/SA can be occupied by coherent positions, so the two debates are logically disentangled. Secondly, we motivate this separation of the debates by showing that the metaphysical claim is not and cannot be accessed epistemically by any part of the scientific process, and is in fact irrelevant to understanding the claims made by scientists, contrary to what is commonly assumed. We would like to add that this work allows for metaphysical 'agnosticism', so to speak, to be combined with SR or SA. Such a position is quite similar to that advanced by Fine (1984); still, our dissatisfaction with Fine's proposal will be clarified in the second part of the paper.

We now focus on what remains of the scientific realism debate now that the metaphysical dispute has been separated.

## 2. Entangled Questions: Aims and Commitments

Now we have shown that the debates on metaphysical and scientific realism are best dealt with separately, we turn our attention to the questions that remain regarding the debate on science. We wish to point out that, looking at the structure of the positions advanced in the literature, they can be divided into two groups: one related to the aims, as well as possible and actual achievements, of science, which we call a ‘stance towards science’, and the second regards the sorts of ontological commitments we ought to hold, which we call a ‘theory on science’.

As far as we understand it, a position in the debate can be understood as what we call a ‘theory on science’, that is to say a philosophical thesis that explains what our commitments should be in the face of scientific practice or given a scientific theory. In other words, a ‘theory on science’ should provide an answer to several difficult questions, such as: Are scientific statements true? What does it mean that they are true? Do scientific entities exist, and, if affirmative, what does it mean that they exist? Which scientific entities exist if not all of them? and so on and so forth. An example of a ‘theory on science’ is the characterization of scientific realism given by Psillos (2005), as a threefold thesis involving metaphysical, semantic and epistemic commitments.

Some authors have instead characterized scientific realism as being a stance toward what science aims to do (Van Fraassen, 1980; Dorato and Laudisa, 2014). Such a conception has been criticized by some philosophers (Kitcher 1993; Chackravatty, 2007) as too weak a definition because this way of defining scientific realism is compatible with science having failed to achieve any purposed aims; therefore, it is not what (many) scientific realists want to claim, since they usually want to concede that science has achieved, at least partially, some of its aims. For example, scientific realists who accept the metaphysical claim usually want to assert also that scientific theories are approximately true, in the sense of achieving partial correspondence with the external world. What we call a 'stance toward science' answers the question of what science aims to do, taking into account ideas about what it can possibly accomplish. In the literature, the majority of the stances we have come across are realist. These do not spell out their ontological commitments but take some characterization of science's aims to be sufficient for defining their position. It is less common for those who call themselves anti-realists to be satisfied with a description of aims.

We assert that realists and anti-realists alike (as well as those who take a 'middle ground' position) will, either implicitly or explicitly, assume a stance towards science. That is, in order to put forward what we are calling a ‘theory on science’, one must first have in mind some idea of what the scientific enterprise can possibly achieve as well as what it concretely aims to do.
We agree that confining oneself to ‘stance-talk’ evades the questions associated with a 'theory on science' which, we think, are intimately bound up with the question of scientific realism and cannot be avoided. Our dissatisfaction with Fine’s (1984) proposal, for example, is then soon explained: his natural ontological attitude seems to imply an acceptance of scientific theories without any philosophical analysis; in this way, we think, he evades the aforementioned questions about truth, commitments and so on. Although he gives a sensible suggestion for how to think about science in general, he does not offer any tools for demarcating those parts of a scientific theory we ought to take seriously from those we should not.

No serious contender in the debate on scientific realism can be defined only as a stance, since this misses what we believe to be the main points of contention. But while we agree with those who think that the stance talking is not enough to fully characterize a clear position in the debate, we disagree that it is entirely irrelevant. Indeed, the questions of aims and achievements are still relevant and important, and have a noticeable impact on how one might respond to the questions we associate with a ‘theory on science’. Adopting a stance towards science, in many cases, constrains the ‘theory on science’ that one can endorse. We do not think that the debate on scientific realism must be spelt out in terms of either ‘stance’ or ‘theory on’ alone, but that a full position will describe commitments and be at least partially based on what one takes the aims and possible achievements of science to be. We think that outlining such connections between the stances implicitly assumed and the theories explicitly endorsed will provide a more comprehensive analysis of the disagreement which the participants of the debates are involved in.

To make this distinction between aims and commitments ('stance towards' and 'theory on' science), and the relationship between them, a little clearer, consider van Fraassen's constructive empiricism (1982) as an example. Van Fraassen makes a distinction between observable and unobservable entities, and claims that we ought to believe only in observable scientific entities. The majority of realists would not accept that such a distinction is a useful guide to ontology. We think that such a disagreement follows from the stances adopted: van Fraassen's (1982, p. 12) stance is roughly that "science aims to give us theories which are empirically adequate". 'Empirically adequate', for van Fraassen (1982, p. 18) is synonymous with the statement, 'what the theory says about what is observable (by us) is true'. It is not surprising that, with such a stance, he refuses to be committed to the existence of unobservable entities. If science aims to describe only the observable world, the unobservable entities posited are merely directed towards this aim, and so we ought not to be committed to their reality. On the contrary, scientific realists usually think that the aim of science is to describe the external world. So, the disagreement that constructive empiricists and scientific realists have about the existence of unobservable scientific entities seems to be grounded in the stance assumed rather than on the commitment (the reality of unobservables) itself.

We think that investigating 'stances' and the ways in which these influence the 'theories on science' is a project worth pursuing, since this framework provides a nice way to taxonomize the positions involved in the debate showing were lie the roots of the commitments taken by each philosophers; as a consequence, it is also helpful for clarifying many of the standard disagreements regarding scientific realism. Finally, it sets out some conditions that a position in the debate should answer to, in order to show what assumptions it rests on.

Agazzi, E., (2014): Scientific Objectivity and Its Contexts, Springer.
Chackravatty, A., (2007): "Six Degrees of Speculation: Metaphysics in Empirical Contexts", in Monton Images of Empiricism: Essays on Science and Stances, with a Reply From Bas C. Van Fraassen. Oxford University Press. pp.183–208.
Dorato, M. and Laudisa, F., (2014): "Realism and instrumentalism about the wave function. how should we choose?", arXiv preprint arXiv:1401.4861.
Fine, A., (1984): "The natural ontological attitude", The philosophy of science, pp.261-277.
Kitcher, P., (1993): The Advancement of Science: Science Without Legend, Objectivity without Illusions, Oxford: Oxford University Press.
Leonelli, S.,(2016): Data-centric biology: a philosophical study, University of Chicago Press.
Massimi, M., (2011): "Structural Realism: A Neo-Kantian Perspective", in Alisa Bokulich and Peter Bokulich (Eds.), Scientific Structuralism, Springer Science+Buisness Media, pp. 1-23.
Psillos, S., (2005): "Scientific realism", Encyclopedia of Philosophy, Gale Macmillan Reference.
Van Fraassen, B. C., (1982): The scientific image, Oxford University Press.

## THIRD SESSION: Classical and Non-Classical Logics

**Stefano Bonzio** (Marche Polytechnic University), **Tommaso Flaminio** (Artificial Intelligence Research Institute, IIIA — Spanish National Research Council, CSIC), **Paolo Galeazzi** (University of Copenhagen)
*Sure-wins under coherence*

In a series of seminal contributions [1, 2], Bruno de Finetti provided a rather general justification for the probabilistic representation of rational beliefs. To this end, he identifies degrees of belief, for an event to occur, with the price of gambles in a suitably defined betting situation, described below (see also [4]).

Let us fix a finite Boolean algebra **A** and a finite subset $\Phi = \{a_1, \dots, a_n\}$ of $A$, the set of events. A *bookmaker* B publishes a *book*, i.e., a complete assignment $\beta\colon \Phi \to [0,1]$. A *gambler* G chooses stakes $\sigma_1, \dots, \sigma_n \in \mathbb{R}$ and, in order to bet over the event $a_i$ (with $i = 1, \dots, n$) pays $\sigma_i \cdot \beta(a_i)$ to B: G, in the possible world $v$ (a homomorphism from **A** to $\{0,1\}$), will get $\sigma_i$ (from B), provided that the event $a_i$ occurs (i.e. $v(a_i) = 1$), and 0 otherwise. Notice that stakes may be negative. A negative stake $\sigma_i$ means *reversing* the bet (or, betting against $a_i$), namely receiving (from B) $-\sigma_i \cdot \beta(a_i)$ and paying $-\sigma_i$ in case $a_i$ takes place.

The book $\beta$ is said to be *coherent* if G has no choice of (real-valued) stakes $\sigma_1, \dots, \sigma_n$ such that, for every valuation $v$

$$\sum_{i=1}^{n} \sigma_i\big(\beta(a_i) - v(a_i)\big) < 0\,.$$

The left-hand side of the above inequality expresses the bookmaker's balance. Therefore, a book is coherent if it prevents B from what in the literature is known as a *sure-loss*.

Recall that a finitely additive *probability measure* over a Boolean algebra **A** is a map $P\colon \mathbf{A} \to [0,1]$ such that $P(1) = 1$ and $P(a \vee b) = P(a) + P(b)$, provided that $a \wedge b = 0$. De Finetti's theorem states the following.

**Theorem 1** (de Finetti [1]). *Let* $\Phi$ *be a finite subset of a Boolean algebra* **A**, and $\beta\colon \Phi \to [0,1]$ *a book. T.F.A.E.*

1. $\beta$ *is coherent*;
2. $\beta$ *extends to a finitely additive probability measure over* **A**.

The rationale behind de Finetti's theorem shows that a bookmaker can prevent himself from a sure-loss, i.e. from going bankrupt if and only if the betting quotes he fixes for the events are chosen accordingly with Kolmogorov's axioms of (finitely additive) probability theory.

In this paper, we will be concerned with the problem of establishing if a gambler, playing on two or more books $\beta_1, \dots, \beta_m$ on the same set of events has a *sure-win strategy*, i.e., if there exists a choice of stakes which, once suitably placed, ensures him a strictly positive gain in all possible worlds. A sure-win strategy, which will be precisely defined below, trivially exists under the hypothesis that at least a book amongst the $\beta_i$'s is incoherent. However, as we will show below, sure-win strategies also exist although assuming that *all* the $\beta_i$'s are coherent.

**Definition 2**. Let $\beta_1, \dots, \beta_m$ be coherent books on $\Phi = \{a_1, \dots, a_n\}$. We say that a gambler has a *sure-win strategy on* $\beta_1, \dots, \beta_m$ if for each event $a_i \in \Phi$ there is a book $\beta_{w(i)}$ amongst $\beta_1, \dots, \beta_m$ such that the book $\beta_w\colon a_i \mapsto \beta_{w(i)}(a_i)$ is incoherent.

In other words, a gambler has a sure-win strategy if there exists a map $w\colon \{1, \dots, n\} \to \{1, \dots, m\}$ and stakes $\sigma_1, \dots, \sigma_n \in \mathbb{R}$ such that in every possible world $v$,

$$\sum_{i=1}^{n} \sigma_i \left(\beta_{w(i)}(a_i) - v(a_i)\right) < 0$$

If $\beta_1, \dots, \beta_m$ are coherent books, we say that they are *jointly coherent* if no gambler has a sure-win strategy on them. In case $m = 2$, we will also say that $\beta_1$ is *jointly coherent with* $\beta_2$ if $\beta_1, \beta_2$ are jointly coherent.

It is worth to recall that de Finetti's notion of coherence admits an equivalent geometrical characterization. Indeed, Theorem 1 can be restated by saying that a book $\beta$ is coherent if and only if $\beta \in \mathcal{C}_\Phi$, where $\mathcal{C}_\Phi \subseteq [0,1]^n$ is the convex hull of the points corresponding to the valuations of formulas in $\Phi$ (possible worlds).

The following example gives a first geometrical glimpse on coherence and joint coherence.

**Example 3**. Consider a set with two events $\Phi = \{a_1, a_2\}$, where $a_1 = p$ and $a_2 = p \vee q$ in a language with two propositional variables $p, q$. Thus, the free algebra **A** (generated by the events) has four homomorphisms to 2, namely those maps $h_1, h_2, h_3, h_4 : \{p, q\} \to \{0,1\}$ which assign, respectively, to $p$ and $q$ the values $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$. Therefore, we obtain the following points $q_1, \dots, q_4 \in \mathbb{R}^2$:

$$
\begin{aligned}
q_1 &= \langle h_1(a_1), h_1(a_2)\rangle = \langle h_1(p), h_1(p \vee q)\rangle = \langle 0,0\rangle \\
q_2 &= \langle h_2(a_1), h_2(a_2)\rangle = \langle h_2(p), h_2(p \vee q)\rangle = \langle 0,1\rangle \\
q_3 &= \langle h_3(a_1), h_3(a_2)\rangle = \langle h_3(p), h_3(p \vee q)\rangle = \langle 1,1\rangle \\
q_4 &= \langle h_4(a_1), h_4(a_2)\rangle = \langle h_4(p), h_4(p \vee q)\rangle = \langle 1,1\rangle
\end{aligned}
$$

Since $q_3 = q_4$, we have:

$$\mathcal{C}_\Phi = co(\{q_1, q_2, q_3\}) = co(\{\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,1\rangle\})$$

depicted as in Figure 1 below, where $co(\{q_1, q_2, q_3\})$ denotes the convex hull of the points $\{q_1, q_2, q_3\}$.

Consider the following coherent books:

1. $\beta_1(a_1) = 1/2$ and $\beta_1(a_2) = 2/3$ ;
2. $\beta_2(a_1) = 1/4$ and $\beta_2(a_2) = 2/3$ ;
3. $\beta_3(a_1) = \beta_3(a_2) = 1/3$ .

Fixing $m = 2$, i.e. considering the joint coherence of two books over $\Phi$, it is easy to check that $\beta_1$ is jointly coherent with $\beta_2$, which is jointly coherent with $\beta_3$. On the other hand, $\beta_1$ is *not* jointly coherent with $\beta_3$. Indeed, the book $\alpha : a_1 \mapsto \beta_1(a_1) = \frac{1}{2}, a_2 \mapsto \beta_3(a_2) = \frac{1}{3}$ is incoherent because, from de Finetti's theorem, there is no probability measure which maps $P(p) > P(p \vee q)$. Therefore, a gambler who is allowed to choose, for each event, which book to bet with has a sure-win strategy if the book into play are $\beta_1$ and $\beta_3$. Clearly, if we consider $m = 3$, the books $\beta_1, \beta_2, \beta_3$ are not jointly coherent.

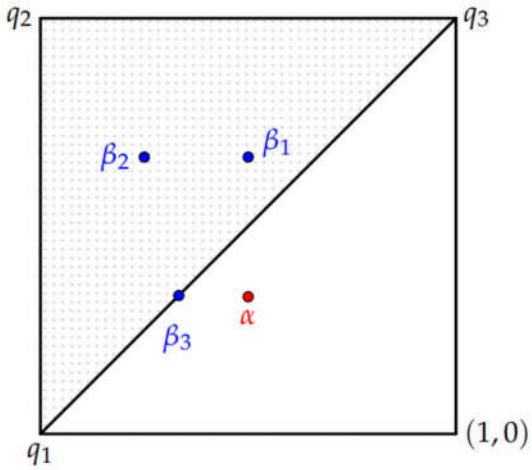



Figure 1: The convex hull $\mathcal{C}_\Phi$ (dotted); the coherent books $\beta_1$, $\beta_2$, $\beta_3$ and the incoherent book $\alpha$ considered in Example 3.

The above example shows also that, when two books are considered, the relation of joint coherence is reflexive and symmetric, but not transitive.

Taking advantage of the geometric version of de Finetti's theorem, we will provide a geometrical characterization of joint-coherence books. In detail, for every book $\beta : \Phi \to [0,1]$ and for every $i = 1, \dots, n$, let $\delta_i = (d_i^+, d_i^-) \in \mathbb{R}^2$ be such that:

1. $d_i^{\pm} \geq 0$ ;
2. the books $\beta_{d_i^+} = \langle \beta_1, \dots, \beta_{i-1}, \beta_i + d_i^+, \beta_{i+1}, \dots, \beta_n \rangle$ and $\beta_{d_i^-} = \langle \beta_1, \dots, \beta_{i-1}, \beta_i - d_i^-, \beta_{i+1}, \dots, \beta_n \rangle$ are coherent.
3. for all $\varepsilon > 0$, $\langle \beta_1, \dots, \beta_{i-1}, \beta_i + d_i^+ + \varepsilon, \beta_{i+1}, \dots, \beta_n \rangle$ and $\langle \beta_1, \dots, \beta_{i-1}, \beta_i - d_i^- - \varepsilon, \beta_{i+1}, \dots, \beta_n \rangle$ are incoherent.

Let us hence define the rectangle

$$\mathcal{R}_\beta = \{\gamma \in \mathbb{R}^n | \ (\forall i = 1, \dots, n)\ d_i^- \leq |\gamma_i - \beta_i| \leq d_i^+\}$$

and

$$\mathcal{C}_\beta = \mathcal{C}_\Phi \cap \mathcal{R}_\beta$$

Obviously $\mathcal{C}_\beta$ is nonempty if (and only if) $\beta$ is coherent. Further, the following result holds.

**Proposition 4**. Let $\beta: \Phi \to [0,1]$ be a book. Then, $\gamma: \Phi \to [0,1]$ is jointly coherent with $\beta$ if and only if $\gamma \in \mathcal{C}_\beta$.

**Example 5**. Consider gain the set of events $\Phi = \{p, p \vee q\}$ and the convex hull $\mathcal{C}_\Phi = co(\{q_1, q_2, q_3\}) = co(\{\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,1\rangle\})$, given in Example 3. Recall that the book $\beta_1$ was such that $\beta_1(a_1) = 1/2$ and $\beta_1(a_2) = 2/3$. The space $\mathcal{C}_{\beta_1}$ of books that are jointly coherent with $\beta_1$, is depicted in Figure 2, according to Proposition 4.

In general, if $\beta_1, \dots, \beta_m$ are coherent books on $\Phi$, the following theorem characterizes the situation in which there is no sure-win strategy for any gambler.

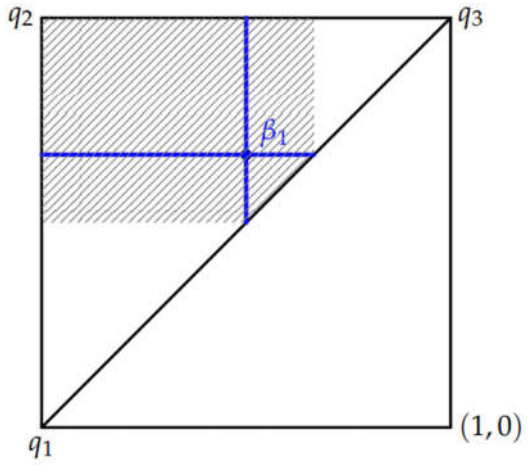



Figure 2: The coherent book $\beta_1$ of Example 5 and the space $\mathcal{C}_{\beta_1}$ of books which are jointly coherent with it (in dashed lines).

**Theorem 6**. Let $\beta_1, \dots, \beta_m$ be coherent books. Then a gambler has no sure-win strategy iff

$$\beta_1, \ldots, \beta_m \in \bigcap_{i=1}^{m} cC_{\beta_i}$$

In the final part of the talk, we will also provide a logical characterization of the notion of joint coherence. This is funded on the fact that coherence of a book can be expressed in various equivalent ways (see [3]), one of which is purely logical form and consists of checking whether an implicative formula is a theorem of Lukasiewicz propositional logic.


[1] B. de Finetti, "Sul significato soggettivo della probabilità", *Fundamenta Mathematicae* 17 (1931), 289-329.
[2] B. de Finetti, Theory of Probability, Vol. 1, John Wiley and Sons, New York, 1974.
[3] T. Flaminio, "Three characterizations of strict coherence o infinite-valued events", submitted.
[4] T. Flaminio, L. Godo, H. Hosni, "On the logical structure of de Finetti's notion of event", *Journal of Applied Logic* 12(3) (2014), 279-301.



**Michele Pra Baldi** (University of Padua)

*The lattice of logics of variable inclusion*

Give a logic $\vdash$ there always exist two sublogics that can be defined by means of a different *variable inclusion* principle, as follows:

$$\Gamma \vdash^{l} \varphi \Leftrightarrow \text{there is } \Delta \subseteq \Gamma \; s.t. Var(\Delta) \subseteq Var(\phi) \text{ and } \Delta \vdash \varphi$$

and

$$\Gamma \vdash \text{^}\{r\}\, \varphi\} \Leftrightarrow \begin{cases} \Gamma \vdash \varphi \text{ and } Var(\varphi) \subseteq Var(\Gamma) \text{ or} \\ \Sigma \subseteq \Gamma\,. \end{cases}$$

where $\Sigma$ is an antitheorem of $\vdash$, i.e., a set of formulas such that $\Sigma \vdash \varphi$ for any formula $\varphi$.

The logics $\vdash^{l}$ and $\vdash^{r}$ are called he left and the right variable inclusion companions of $\vdash$, respectively. The most representative examples arise when $\vdash$ is classical logic. In this case, $\vdash^{l}$ is known in the literature as Paraconsistent weak Kleene logic PWK ([3, 6]), while $\vdash^{r}$ corresponds to Bochvar Logic $\boldsymbol{B_3}$ [2].

These logics can be semantically defined by using so-called Weak Kleene tables (**WK**)

| $\wedge$ | 0 | $n$ | 1 |
|---|---|---|---|
| 0 | 0 | $n$ | 0 |
| $n$ | $n$ | $n$ | $n$ |
| 1 | 0 | $n$ | 1 |

| $\vee$ | 0 | $n$ | 1 |
|---|---|---|---|
| 0 | 0 | $n$ | 1 |
| $n$ | $n$ | $n$ | $n$ |
| 1 | 1 | $n$ | 1 |

| $\neg$ | |
|---|---|
| 1 | 0 |
| $n$ | $n$ |
| 0 | 1 |

and the following logical matrices:

- $\langle \mathbf{WK}, \{1\} \rangle = \boldsymbol{B_3}$
- $\langle \mathbf{WK}, \{1, n\} \rangle = \text{PWK}$

Logics of variable inclusion finds application in different fields, and the philosophical debates is one of them. The first philosophical reason of interest is the "infectious" or "contaminating" behaviour that the element $n$ has over the two classical truth values. It is indeed easy to see that every operation $\delta(n, \vec{c})$ with $\vec{c} \subseteq \{0, n, 1\}$ on **WK** in which $n$ really occurs in such that $\delta(n, \vec{c}) = n$, so, in this sense, $n$ contaminates every sentence in which it occurs. As it emerges from [1] and [2], the main philosophical problem is how to interpret the value $n$ in a suitable way.

Recently, moreover, logics of variable inclusion have also been investigated from the point of view of abstract algebraic logic (AAL). The work in [3] consists in an algebraic analysis of PWK, while [4, 5] provide a general framework to model an arbitrary logic of (left and right) variable inclusion, describing its matrix models and an appropriate Hilbert calculus. Such investigations

extend also to second order AAL, by determining the structure of the Suszko reduced models of a logic of variable inclusion, as well as its location in the so-called Leibniz hierarchy.

The general theory of AAL dictates that, given an algebraic language $\mathcal{L}$, the set of logics of type $\mathcal{L}$ can be equipped with a lattice structure, and such lattice is a complete one [7]. In this paper, given a logic $\vdash$, we investigate the structure of the lattice of sublogics that can be obtained by applying the above-mentioned variable inclusion principles.

In general, given a logic $\vdash$ with a left and right partition function (see [4, definition 16] and [5, definition 25]), we show the following theorems:

**Theorem 1.** Let $\vdash$ be a logic with a partition function and with antitheorems. Then, the lattice of sublogics of variable inclusion of $\vdash$ has at most 8 elements.

**Theorem 2.** Let $\vdash$ be a logic with a partition function and without antitheorems. Then, the lattice of sublogics of variable inclusion of $\vdash$ has at most 5 elements.

Interestingly enough, the structure of the lattice of sublogics of variable inclusion of $\vdash$ deeply depends on the presence of antitheorems of $\vdash$.

We also provide a full characterization of the order relations occurring among all the different sublogics of variable inclusion. This allows for a transparent description of the lattice of sublogics of variable inclusion of a given logic $\vdash$. The following figure 1 represents the lattice of sublogics of variable inclusion of an arbitrary logic $\vdash$ that does not have antitheorems, while figure 2 describes the situation in the case that $\vdash$ possesses antitheorems:

**Figure 1**

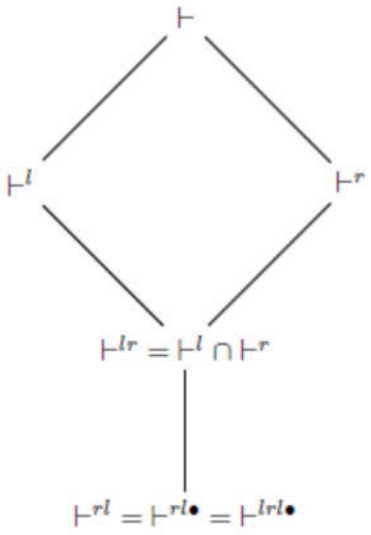
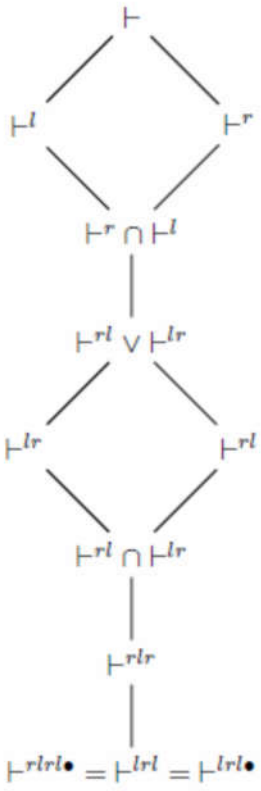



**Figure 2**




(where $\bullet$ is any sequence of elements among $\{l, r\}$)

At last, we apply the results to the particular case of Classical Logic, in order to study the members of its lattice of sublogics of variable inclusion. With this respect, we also investigate the algebraic counterparts of the members of the lattice.

**References**

[1] J. Bell, "Off-topic: a new interpretation of weak-kleene logic", *The Australasian Journal of Logic* 13(6), 2016.
[2] D. Bochvar, "On a three-valued calculus and its application in the analysis of the paradoxes of the extended functional calculus", *Mathematicheskii Sbornik* 4, 1938, 287-308.
[3] S. Bonzio, J. Gil-Férez, F. Paoli, L. Peruzzi, "On paraconsistent weak-kleene logic: axiomatization and algebraic analysis", *Studia Logica* 105(2), 2017, 253-297.
[4] S. Bonzio, T. Moraschini, M. Pra Baldi, *Logics of left variable inclusion and Plonka sums of matrices*, Submitted manuscript, 2018.
[5] S. Bonzio, M. Pra Baldi, *Containment logics and Plonka sums of matrices*, Submitted manuscript, 2018.
[6] S. Halldén, *The Logic of nonsense*, Lundequista Bokhandeln, Uppsala, 1949.
[7] R. Wójcicki, *Theory of logical calculi. Basic theory of consequence operations*. Vol. 199 of Synthese Library. Reidel, Dordrecht, 1988.

**Sara Negri** (University of Helsinki), **Edi Pavlovic** (University of Helsinki)
*DSTIT modalities through a sequent calculus*

Dstit (deliberately seeing to it that) is an agentive modality usually semantically defined upon indeterminist frames - a semantics that builds upon a combination of Prior-Thomason-Kripke branching-time semantics and Kaplan's indexical semantics - enriched with agency. The temporal structure for branching time (BT) is given by trees with forward branching time, corresponding to indeterminacy of the future, but no backward branching, corresponding to uniqueness of the past. Moments are ordered by a partial order, reflecting the temporal relation, and maximal chains of moments are called histories. The trees are enriched by agent's choice (AC), a partition relative to an agent at a given moment of all histories passing through that moment (a partition since, intuitively, an agent's choice determines what history comes about only to an extent).

In such (BT+AC) frames, formulas are evaluated at moments in histories. Specifically, an agent a deliberately seeing to it that A holds at the moment m of a history h, holds iff (i) A holds in all histories choice-equivalent to h for the agent a, but (ii) doesn't hold in at least one history that m is a part of. In simple terms, the agent sees to it that A if their choice brings about those histories where A holds, but nonetheless it could have been otherwise (i.e. an agent can't bring about something that would have happened anyway).

While the semantics for stit modalities and logics built upon them is well established, their proof theory has been largely restricted to axiomatic systems (starting with [18] and [2]) with just a few exceptions, namely a treatment of multi-agent deliberative stit logic through labelled tableaux in [15] that builds upon Belnap's original semantics, and of the related logic of imagination in [17] that exploits a newly defined neighbourhood semantics, introduced in [16].

As for the meta-theoretical properties of stit logics, as for other logics, completeness is usually established through the method of canonical models for axiomatic systems and through exhaustive proof search for tableaux [17]. Decidability, on the other hand, has been achieved through filtration methods [18, 1].

Our aim in this work is to lay down the bases for the development of systems of deduction that cover the stit modalities presented by Belnap et al. [2], starting with dstit, in a way that respects all the desiderata of good proof systems, in particular to achieve a direct proof of decidability though a bound on proof search in a suitable analytic proof system.

Here the method of labelled sequent calculi developed since [6] is utilized: relatively complex truth conditions can be transformed into rules with the help of auxiliary modalities, as in the treatment of Lewis' counterfactuals [13], and additional properties for the characteristic frame conditions are expressed as sequent calculus rules following [9, 10]. The result is a G3-style labelled sequent calculus which is shown to possess all the structural properties, including being contraction- and cut-free.

Moreover, we demonstrate multiple applications of the system. We prove the impossibility of delegation of tasks among independent agents, the interdefinability of Dstit with an agent-relative modality cstit and an agent-independent modality 'settled true,' as well as the treatment of refraining from [2] and [14]. Finally, we demonstrate the meta-theoretical properties of our system, namely soundness, completeness and decidability via a bounded proof search.

[1] Balbiani, P., A. Herzig and N. Troquard, 'Alternative axiomatics and complexity of deliberative STIT theories', Journal of Philosophical Logic, vol. 37, pp. 387-406, 2008.
[2] Belnap, N. D., M. Perloff and M. Xu, Facing the Future: Agents and Choices in Our Indeterminist World, Oxford: Oxford University Press, 2001.
[3] Chellas, B. F., Modal Logic: An Introduction, Cambridge University Press, 1980.
[4] Dyckhoff, R., and S. Negri, 'Geometrization of first-order logic', The Bulletin of Symbolic Logic, vol. 21, pp. 123-163, 2015.
[5] Girlando, M., S. Negri, N. Olivetti, and V. Risch, 'Conditional beliefs: from neighbourhood semantics to sequent calculus', The Review of Symbolic Logic, vol. 11, pp. 736-779, 2018.
[6] Negri, S., 'Proof analysis in modal logic', Journal of Philosophical Logic, vol. 34, pp. 507-544, 2005.
[7] Negri, S., 'Proof theory for non-normal modal logics: The neighbourhood formalism and basic results', IfCoLog Journal of Logics and their Applications, vol. 4, pp. 1241-1286, 2017.
[8] Negri, S., and N. Olivetti, 'A sequent calculus for preferential conditional logic based on neighbourhood semantics', In H. de Nivelle (ed) Automated Reasoning with Analytic Tableaux and Related Methods (TABLEAUX2015), Lecture Notes in Computer Science, vol. 9323, pp. 115-134, 2015.
[9] Negri, S., and J. von Plato, 'Cut elimination in the presence of axioms', The Bulletin of Symbolic Logic, vol. 4, pp. 418-435, 1998.
[10] Negri, S., and J. von Plato, Structural Proof Theory, Cambridge University Press, 2001.
[11] Negri, S., and J. von Plato, Proof Analysis, Cambridge University Press, 2011.
[12] Orlandelli, E., 'Proof analysis in deontic logics', In F. Cariani et al. (eds) Deontic Logic and Normative Systems (DEON2014), Lecture Notes in Artificial Intelligence, vol. 8554, pp. 139-148, 2014.
[13] Negri, S., and G. Sbardolini, 'Proof analysis for Lewis counterfactuals', The Review of Symbolic Logic, vol. 9 , pp. 44-75, 2016.
[14] von Wright, G. H., Norm and Action: A Logical Enquiry, Routledge & Kegan Paul, 1963.
[15] Wansing, H., 'Tableaux for multi-agent deliberative-stit logic', In G. Governatori, I.M. Hodkinson, Y. Venema (eds.), Advances in Modal Logic 6, pp. 503-520, 2006.
[16] Wansing, H., 'Remarks on the logic of imagination. A step towards understanding doxastic control through imagination', Synthese, vol. 194, pp. 2843-2861, 2017.
[17] Olkhovikov, G. K., and H. Wansing, 'An axiomatic system and a tableau calculus for STIT imagination logic', Journal of Philosophical Logic, 47(2), pp. 259-279, 2018.
[18] Xu, M., 'Axioms for deliberative STIT', Journal of Philosophical Logic, vol. 27, pp. 505-552,1998.

**Davide Dalla Rosa** (University of Padua)
*In which sense is Kant's categorical syllogistic non-classical?*

This talk aims at offering a reconstruction of Kant's theory of categorical syllogistic in his general logic.
The reconstruction shall highlight the differences between Kant's syllogistic and a possible formalization of it in classical predicate logic. From some problematic issues involved in this reconstruction, it will be held that Kant's theory of categorical inferences of reason as it is, namely as it is placed in the context of his general logic, seems at a first glance to not match some desiderata that are related to the traditional extensional interpretation of terms in the Aristotelian syllogistic, and in general to some deductions allowed in traditional syllogistic. It is claimed that the mismatch between traditional syllogistic and Kant's syllogistic

shall be related at least in principle to three main issues, not taken into account by the interpreters: a) to the intensional character of the semantic of concepts in Kant's general logic; b) to the difference between Kant's categorical judgements and traditional categorical propositions, which is related to the to the issue of quantification in an intensional reading of concepts and c) thirdly to the heavy restrictions on the inferential rules that Kant admits. It will be concluded that there are substantial reasons for considering the restriction on the inferential rules as the core issue on which the mismatch is based.

The initial hypothesis is that Kant's general logic shall be analyzed in the framework of traditional syllogistic, together with a further assumption, namely that traditional syllogistic is equivalent to first-order predicate logic with existential import for the subject term of categorical propositions.

This hypothesis is on the same line of an existing debate on the reduction of syllogistic figures in Kant's philosophy, that seems to implicitly rely on a classical extensional interpretation of the semantics of terms and that presents Kant's syllogistic as somehow incomplete or mistaken. Amongst the interpreters that take part to the debate from different points of view we could find Capozzi (1980), Kirk Wilson (1973), Myrstad (2008), and Glei (2013). A partial explanation for these contrasting patterns of analysis is that they mirror important differences in the structure of Kant's works on logic. A core point that crosses these interpretations is the sub-debate on the problematic status of two valid syllogistic moods, that have been labelled as "non utiles" (useless) by Kant in the Dohna-Wundlacken Logik in this quotation: "The modi come altogether to 64 various kinds from the 4 vowels a — e — i — o. For each there are 16 different inferences. But 28 are left out according to the rule: "Ex puris negativis nihil sequitur. 18 are left out according to the rule: Conclusio sequitur partem debiliorem. 8 are excluded according to the rule that a negative conclusion cannot follow from merely affirmative premises. There remain no more than 10 modi, of which only 8 modi are utiles" (I. Kant 1992, p.508). This debate shall not be taken as particularly meaningful in itself for the aims of this talk, and the aforementioned quotation can also be seen as not completely reliable, but, nonetheless, it will be taken as a useful tool of analysis.

Although they have not been fully identified as the two "non utiles" moods, I will focus myself in particular on the syllogistic moods Baroco and Bocardo, even though I will mention Disamis and some other valid moods too. In traditional syllogistic, the former two moods cannot be proved to be valid through the rules of conversion, but they require a reductio ad absurdum or, only in the case of Bocardo, a further procedure of reduction called ecthesis. Since Kant almost does not mention and features these methods of proof in his general logic, it remains unclear how to account the reduction for these syllogistic moods, also comparing it with what Kant explicitly says about reduction for syllogisms in each syllogistic configuration.

The structure of the talk will be then the following. It will deliver firstly a fairly general assessment of Kant's syllogistic and a reconstruction of Kant's version of the so-called dictum de omni et de nullo that is based on his published works (§1); then it will be examined in more details some reconstructions of Baroco's and Bocardo's possible reduction to the perfect syllogistic moods that do not employ a reductio ad absurdum, but that seem to require different kinds of alternative devices not mentioned by Kant (quantification in concepts, obversion, transposition of premises) as well as the reduction of syllogistic moods that seemingly make use of the immediate inference known as conversio per accidens, that implies a relation of subalternation between terms which does not hold in classical first-order predicate logic; (§2). I claim that the issues connected with these accounts, amongst the others the ones related to the existential import of singular and universal judgements in Kant's theory (§3), could support in principle the problems that we encounter in giving a satisfactorily account of Kant's categorical syllogistic and of the difficulties we face in proving some specific moods.

It will be concluded however that neither subalternation, nor the exclusive negation involved in the reductio proof, seem to be so problematic from the point of view of Kant's theory of categorical syllogisms in his general logic, but rather that this unaccountability has be to be ascribed to the restriction on inferential rules theorized by Kant.

Following this argumentative line shall come out that for these reasons Kant's syllogistic as a logical theory shows a behaviour which is somehow divergent from its possible formalization in classical first-order predicate logic, being in a way non-classical.

CAPOZZI M.(1980), Osservazioni sulla riduzione delle figure sillogistiche in Kant. In: "Annali della Facoltà di Lettere e Filosofia dell'Università di Siena", 1, 79–98.

GLEI R. (2013), Die merk-würdigen Modi Baroco und Bocardo, Zur Axiomatik der Syllogismen bei Aristoteles, Boethius und den moderni (mit einem Ausblick auf Kant). In: "Bochumer Philosophisches Jahrbuch für Antike und Mittelalter", Band 16, pp.163-84.
KANT I, (1992), Lectures on Logic, (eds. M.Young), Cambridge University Press, Cambridge. MYRSTAD J.A., Kant's Treatment of the Bocardo and Barocco Syllogisms. In Valerio Rohden et al. (ed.), Recht und Frieden in der Philosophie Kants: Akten des X. Internationalen Kant-Kongresses. 4.–9. Sept. 2005 in São Paulo, Berlin 2008, 5, 163–174.
WILSON K.D. (1975), The Mistaken Simplicity of Kant's Enthymematic Treatment of the Second and Third Figures. In "Kant-Studien", 66, 404– 417

## FOURTH SESSION: Philosophy and Foundations of Physics

**Frida Trotter** (University of Lausanne)
*The [un]observability of the entangled state*

The limits of the observable in quantum mechanics (QM henceforth) are not dictated exclusively by structural limits of our empirical relationship with the world, such as our perceptual abilities and our technological means. In fact, the boundaries of observability seem to be "located" already at a purely theoretical level: it is within the theory that we find, in the characterization of physical systems, the reasons why some elements of the quantum realm are in principle unobservable. The general claim advanced in this abstract is that the theoretical characterization of quantum systems, and especially of their states, seems to set what can be empirically observed of these very systems, and what instead is by definition left out of the range of observability.
For supporting this claim, I will consider the example of the entangled state. In order to tackle the argument with major precision, I will consider an account of observation according to which the entangled state is simply unperceivable, but not unobservable. Such account is Peter Kosso's interaction-information account, which considers observation essentially as a process of information-gathering. My argument does not intend to take a rigid adversative position against Kosso's account as a whole, but it rather suggests that it has been inadequately applied to the case of the entangled state.
The structure of the abstract will thus be as follows. In section 1, I will spell out the definition of observation I am going to question, and its application to the entangled state. Section 2 contains three criticisms that it is possible to move to Kosso's claim for the observability of the entangled state. Finally, section 3 contains the conclusion that I derive from the argument that this abstract schematically displays.

1. Observation as detection that carries information.

The account of observation considered here is based on a interaction between detection and information-gathering. According to Kosso's "interaction-information account of observability" (Kosso, 1989), observation in science is the process through which we gather information about the object observed. Kosso individuates two constraints for observability, the first of which is that "observability is a two place predicate." (31) This means that the things observed are not objects or properties in isolation, but always either objects with certain properties, or properties of an object. Second, the transfer of information is conveyed through a physical interaction between the object observed and the observer. The definition of observable that he provides is the following:
The ordered pair < object x, property P > is observable to the extent that there can be an interaction (or a chain of interactions) between x and an observing apparatus such that the information "that x is P" is transmitted to the apparatus and eventually conveyed to a human scientist. (Kosso, 1989, p. 32. Underlined in the original)
It emerges clearly that this definition is rather permissive. It is crucial for the study of fundamental entities that perception does not play a major role in the interaction between observer and object observed. Moreover, no limits are posited to the type of technology that can be used for obtaining information from the object under investigation.
On the basis of this definition, Kosso operates a distinction among things that are unobservable in principle, unobservable or unperceivable in fact but not in principle, and perceivable (39-41). He collocates quantum state functions in the second category, namely, of things that are unperceivable in fact, but still observable.
In the section dedicated to such case study (72-80), Kosso first displays some of the particular properties of the quantum state. He notices that the state is defined for an ensemble, and not for individual systems, and that, per hypothesis, the wave function describing the state of the ensemble contains all the information concerning measurable properties of this latter (73). Since the value of the quantum state is not associated with any singular physical property, information about it can be gathered empirically by considering individual eigenvalues of the properties associated to the state, and by measuring the probability distributions of these values (74). Both can be gathered only after the interaction between entangled system and apparatus.
Kosso considers an example of experiment conducted with a Stern-Gerlach apparatus, by means of which it is possible to measure properties of some of the systems of an ensemble once they interact with the detectors; from the information provided by samples of the system it is possible to make inferences about the entangled

state of the non-measured part of the ensemble (76). There are some important features that characterize the experiments for measuring the entangled state which are not seen by Kosso as compromising factors for the state's observability. First, the empirical basis for making inferences about the entangled state is constituted by a measure of probability of the values of the registered outcomes. The information that needs to be gathered concerns more values of the same property (for instance, both the probability of measuring spin up and that of measuring spin down along different axes are considered. (78)). Moreover, and this is the crucial point for the present argument, measurements carried out on an ensemble of entangled particles cause the disruption of the entangled state. But Kosso argues

> […] observation destroys what it observes so that in the end one has information of the state of that part of the ensemble with which one has not interacted. One can describe this alternatively as having information of the state of the entire ensemble immediately before the observation. (78)

Nevertheless, this section is concluded with the claim that the entangled state is indeed observable, and that experiments such as the one described in these pages are one way to observe it.

> the observability of $\psi$ it can be located along the dimensions of observability. The state of a quantum ensemble is unperceivable in fact but not unobservable in principle. The Stern-Gerlach example demonstrates how an apparatus can correlate ensemble state information to apparatus state information. (79)

The justification of this claim derives from the notion of information: since correlations in the measurement outcomes are the consequence of an experiment that prepares the ensemble studied in an entangled state, by observing these very correlations, and hence by obtaining reliable information about these, one is justified in making inferences about the entangled state of the unsampled portion of the system. If by observing the interaction between the ensemble and the apparatus one gets information about the outcomes and, derivatively, about the entangled state of the ensemble, one thereby can claim that they are observing the entangled state of the ensemble.

This is the part of Kosso's argument that I consider fallacious. The structure of my objection is very simple: the entangled state, as described by QM has the constitutional feature that it cannot be preserved unaltered in an interaction with the measuring apparatus. The elements of the ensemble, described by a unique wave function that assigns a range of values to the considered properties before measurement, behave as singular systems when they interact singularly with the apparatus, and what is measured is a definite value for the property considered. The objection is simply that if the act of measurement causes the disappearance of the entanglement relation among the elements of the ensemble, what is observed consists only of the values registered upon measurement, and not of the entangled state that describes the elements of the ensemble before measurement. It seems reasonable to claim that information provided by the measurement outcomes might warrant some type of inference regarding information about the ensemble before measurement. But I do not think that this ipso facto justifies concluding that thereby we are also obtaining observational data of the ensemble before measurement too. My contention is that what is observed is only what can be measured, and that the inferences that can be drawn about the entangled state before-measurement have only theoretical nature.

## 2. Three criticisms

The elements at play in the present abstract are an account of observation based on the connection between observation and information gathering, and the properties of the quantum description of the entangled state for an ensemble of particles. I have already briefly hinted at one objection I have regards Kosso's analysis. In this section, I would like to outline three more criticisms I have with respect to what has been schematically exposed above, and draw a possible conclusion in the light of these criticisms.

First, the interaction between entangled system and apparatus is a characteristic feature of QM normally referred to as the "measurement problem". This expression refers to the fact that the empirical significance of the quantum description of systems emerges when such description is supplied with a notion of measurement. By means of specific mathematical procedures it becomes possible, given the quantum description of systems, to predict which measurement outcomes are going to obtain and with which probability value. The aspect that is important for the present argument is that before the measurement is carried out, the entangled state that describes an ensemble of particles formally assigns more than one possible value for the same property at the same time. In other words, the entangled state contemplates

always the presence of a superposition of the different values that could be measured, none of which is prevalent over the others before the measurement is carried out. It could be possible to say that the value of the state before measurement is indefinite, as it is constituted by the coexistence at the same time of many possible values. The interaction between the entangled system and the apparatus leads to the instantiation of one possible outcome which, in a sense, was not there before the measurement was carried out. When the system under consideration is in an entangled state, the system does not have one definite value for the property (e.g. spin) that will be measured. Turning the expression around, when, upon measurement, the apparatus registers definite values through the different detectors, the system is not in an entangled state any longer. What is measured, through a physical interaction with the system, are the definite values obtained after the disruption of the entanglement relation. Hence, as the presence of the latter is possible only when the relation is disrupted, it is not possible that measurement outcomes represent an instance of observation of the entangled state. They can convey information about it, and are surely considered evidence of the presence of an entangled system. But they cannot be considered an instance of observation of it.

The second criticism regards a notion of reality connected to the entangled state. It has been recently formulated a mathematical proof of the fact that if one considers the quantum state to be merely a representation of the information we have of a physical aspect of the world, predictions are obtained that contradict those of quantum theory. Hence, the quantum state must be considered as a physical property of the system (Pusey, Barrett, & Rudolph, 2012). This demonstration, also called "PBR theorem", applies also to the entangled state, which therefore can be considered to be as real as the very measurement outcomes obtained through measurement. Nevertheless, the nature of the reality described by the entangled state is far from being less obscure. The observational data connected to entangled systems are the correlations among the measurement outcomes obtained in experiments where two or more particles are prepared in an entangled state. Contention of my first criticism is the fact that the correlated measurement outcomes cannot be considered as observational data of the entangled state before measurement. My second contention is that, since the entangled state is a physical property of the system, it seems reasonable to think that if the correlated measurement outcomes were data about that state itself, its physical nature would not be so obscure.

My third and final criticism regards more of a logical aspect of the observability of the entangled state as described by Kosso. It is undeniable that the observed correlations are predicted with high accuracy by quantum theory, and that it is possible to derive very precise probabilities of the outcomes based on the very mathematical notion of the entangled state. Hence, the fact that the observational data correspond with the predictions made by means of the notion of entanglement may imply that this latter is a strong and correct part of the theory. But the fact that the theoretical description of the entangled state yields observable predictions does not bear any implication concerning the observability of the state in itself. Rather, on the basis of what has been briefly argued above, it is possible to derive the following conclusion. Since the production of the measurement outcomes requires the "dismemberment" of the entangled system, either the system is in an entangled state, or it is not anymore, as when observational measurement outcomes are obtained. Therefore, in the hypothetical eventuality that it was possible to observe the entangled state, it seems not logically possible that observational data of it would be represented by the data considered by Kosso, as the presence of the entangled state implies the absence of definite values as those obtained upon measurement.

## 3. Conclusion

In the present abstract I have tried to outline schematically the structure of an argument against the claim that we can observe the entangled state, and that correlated measurement outcomes obtained from experiments involving entangled systems are instances of observation of the state itself. To accomplish this, I have first summarized the main features of Peter Kosso's "interaction-information account of observability", according to which the entangled state is an example of thing that is unperceivable in fact, but observable. I have argued that pairing the notions of observation and of information as he does, and justifying a claim about observability on the basis of an inference about information is a fallacious move. In section 2, I have outlined three criticisms to the idea that there are instances of observation of the entangled state, and I have advanced the idea that in fact this latter might represent an example of thing that is unobservable in principle.

In the light of this, I believe that it could be possible to advance also the following conclusion. It seems reasonable to require that the account of features of the world contained in scientific theories is directly

supported by empirical claims, but it seems that the notion of entangled state itself lacks such support. Hence, it seems at least very problematic to make claims about the theory-independent ontological nature of entanglement, which, rather, seems to fit very well the category of theoretical entity. At the very least, it seems that it might be difficult to draw a different conclusion, based on the state of the theory and of the empirical data that are available at the present moment.

Jaroszkiewicz, G. (2017). Quantized Detector Networks; The Theory of Observation. Cambridge UK: Cambridge University Press.
Kosso, P. (1989). Observability and observation in physical science. Dordrecht: Kluwer Academic Publishers.
Pusey, M. F., Barrett, J., & Rudolph, T. (2012). On the reality of the quantum state. Nature Physics, 8, 475-478.
Rynasiewicz, R. (1984). Observability. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association (pp. 189-201). The University of Chicago Press.
Shapere, D. (1982, December). The Concept of Observation in Science and Philosophy. Philosophy of Science, 49(4), 485-525.
Wallace, D. (2008). Philosophy of Quantum Mechanics. In D. Rickles, The Ashgate Companion to Contemporary Philosophy of Physics (pp. 16-98). Aldershot UK: Ashgate Publishing limited.

**Ivan Chajda** (University of Olomouc), **Davide Fazio** (University of Cagliari) **Antonio Ledda** (University of Cagliari)
*The Generalized Orthomodularity Property: Configurations, Pastings and Completions*

Quantum logic generalizes classical logic yielding a framework for foundation of quantum theory. This research area, and its name, originated in a famous article by G. Birkhoff and J. von Neumann, who were attempting to reconcile the apparent inconsistency of classical logic with the facts concerning the measurement of complementary variables in quantum mechanics, such as position and momentum. Within this theory, a key role is played by the concept of orthomodular lattice (OML), which provides an algebraic abstraction of the lattice of closed subspaces of any Hilbert space.
In particular, OML's were introduced by K. Husimi [13], and simultaneously they were studied by G. Birkhoff and J. von Neumann [3], when they were trying to develop the logic of quantum mechanics by inquiring into the structure of the lattice of projection operators on a Hilbert space. However, the term was coined by I. Kaplansky, when it was realized that the concrete lattice of projectors, albeit ortho-complemented, is not modular. A wide number of articles and several monographs are devoted to the subject (for an extensive account see [1, 15, 2, 5, 9, 16, 14]). In 1968, many years after the introduction of the concept of OML, it was realized that a more appropriate formalization of the logic of quantum mechanics could be obtained by relaxing the lattice conditions to the weaker notion of orthomodular poset (OMP) [4, 16]. The reason that motivates such weakening is that, in the logic of quantum mechanics, the disjunction of two sentences exists only in case such propositions are indeed orthogonal, but in the general case need not be de- fined.
Actually, albeit endowed with a weaker order, OMP's share interesting features with OML's suggesting, perhaps, these properties are independent of the lattice order. By this fact, one is naturally brought to the following question: which properties of orthomodular structures are independent of lattice operations? In order to provide a possible answer, we introduce a class of orthoposets, namely, complemented posets in which meets, and joins may not exist at all, satisfying a generalized orthomodularity condition (GO-property) defined in terms of LU-operators (see e.g. [6]), and then analyzing their order theoretical properties.
Roughly speaking, orthoposets having the GO-property (GO-posets) can be regarded to as complemented posets whose image in their Dedekind-MacNeille completion satisfies the orthomodular law. This notion turns out to capture important features of the set of subspaces of a (pre-)Hilbert space, the concrete model of sharp quantum logic (see [7]). Furthermore, the class of GO-posets turns out to include many well-known structures such as modular complemented posets, granting a quite general framework ordered sets of prominent importance for the foundation of quantum mechanics can be represented in.

This work is structured as follows. After dispatching all basic notions, we define the concept of GO-poset providing a number of equivalent characterizations and natural applications. In the same section, we study commutator theory in order to highlight the connections with Tkadlec's Boolean posets. Actually, it will turn out that GO-poset are nothing but pastings of Boolean posets. Furthermore, it will be highlighting that main properties of the commutativity relation as well as distributivity in Foulis-Holland sets do not depend on lattice operations. Finally, we apply those results to provide a completely order-theoretical characterization of generalized orthomodularity in terms of orthogonal elements, generalizing Finch's celebrated achievements [8].
Subsequently, we study forbidden configurations with the aim of providing Dedekind- Birkhoff's type Theorems, that fully describe GO-posets by means of particular sub-posets (LU-subsets) which cannot occur in their order structure, generalizing what J. Rachůnek and one the present authors proved in [6]. Interestingly enough, it will turn out that whenever the largest framework of GO-posets is taken into account, orthomodularity and paraorthomodularity (see e.g. [10]) are no longer equivalent. Moreover, taking advantage of these results, we propose a novel characterization of atomic amalgams of Boolean algebras (cf. [1, Chapter 4.4]). In particular, a development of our arguments will yield Greechie's celebrated Theo- rems as corollaries [11, 12] by putting into relationship "critical" atomic loops and forbidden configurations for GO-posets.
Finally, making use of the techniques developed, we will prove that Effect Algebras (see e.g. [7]) whose induced partial order has the GO-property are exactly OMP's. This fact has some relevance for the theory of orthoalgebras, in that it allows to conclude that an (orthosummable) orthoalgebra has a (orthoalgebraic) Dedekind- MacNeille completion if and only if its induced poset is orthomodular, and it can be completed to an orthomodular lattice. To the best of our knowledge, these results are new and subsume, under a unifying framework, many well-known facts sparsely scattered in the literature [17, 18].

[1] Beran L., Orthomodular Lattices: Algebraic Approach, Riedel, Dordrecht, 1985.
[2] Bruns G., Harding J., "Algebraic Aspects of Orthomodular Lattices", In: Coecke B., Moore D., Wilce A. (eds) Current Research in Operational Quantum Logic. Fundamental Theories of Physics, vol 111. Springer, Dordrecht, 2000.
[3] Birkhoff G., von Neumann J., "The logic of quantum mechanics", Annals of Mathematics, 37, 1936, pp. 823-843.
[4] Chajda I., Kolarík M., "Variety of orthomodular posets", Miskolc Mathematical Notes, 15, 2, 2014, pp. 361-371.
[5] Chajda I., Länger H., "Coupled Right Orthosemirings Induced by Orthomodular Lattices", Order, 34, 1, 2017, pp. 1-7.
[6] Chajda I., Rachunek J., "Forbidden configurations for distributive and modular ordered sets", Order, 5, 1989, pp. 407–423.
[7] Dalla Chiara M. L., Giuntini R., Greechie R., Reasoning in Quantum Theory–Sharp and Unsharp Quantum Logic, Kluwer Dordrecht, 2004.
[8] Finch P. D., "On orthomodular posets", Bulletin of the Australian Mathematical Society, 2, 1970, pp. 57-62.
[9] Foulis D. J., "A note on orthomodular lattices", Portugaliae Math., 21, 1962, pp. 65-72.
[10] Giuntini R., Ledda A., Paoli F., "A New View of Effects in a Hilbert Space", Studia Logica, 104, 2016, pp 1145-1177.
[11] Greechie R. J., "On the structure of orthomodular lattices satisfying the chain condition", Journal of Combinatorial Theory, 4, 1968, pp. 210-218.
[12] Greechie R. J., "Orthomodular lattices admitting no states", Journal of Combinatorial Theory, 10, 1971, pp. 119-132.
[13] Husimi K., "Studies on the foundations of quantum mechanics, I", Proceedings of the PhysicsMathematics Society of Japan, 19, 1937, pp. 766-789.
[14] Kalmbach G., "Orthomodular logic", Z. Math. Logic Grundlagen J. Math., 20, 1, 1974, pp. 395-406.
[15] Kalmbach G., Orthomodular Lattices, London Math. Soc. Monographs, vol. VIII, Academic Press, London, 1983.
[16] Matouˇsek M, Pt´ak P.,"Orthocomplemented Posets with a Symmetric Difference", Order, 26, 1, 2009, pp. 1-21.

[17] Navara M., Rogalewicz V.,"The Pasting Constructions for Orthomodular Posets", Mathematische Nachrichten, 154, 1991, pp. 157-168.
[18] Riecˇanov´a Z., "MacNeille Completions of D-Posets and Effect Algebras", International Journal of Theoretical Physics, 39, 2000, pp. 859-869.

**Silvia Bianchi** (IUSS, Pavia)
*Introducing Thin Objects in Mathematical Structuralism: Ontological Dependence and Grounding for a Weak Approach*

The main purpose of this paper is to introduce what will be called Weak Mathematical Structuralism (WMS) as further position within the mathematical structuralist debate. WMS provides a more moderate understanding of Shapiro's (1997) non-eliminative ante rem structuralism and is considerably based on the notions of ontological dependence and grounding, in line with Linnebo (2008) and Wigglesworth's (2018) analyses. WMS is worth endorsing because it avoids some problems of ante rem structuralism without abandoning its main intuitions.
Whereas Shapiro's (1997) account is committed to a background ontology of abstract mathematical structures and reduces the nature of individual objects to mere positions in these structures, WMS applies a non-eliminative approach to both individual objects and abstract structures. On the one hand, individual objects are admitted as thin objects, which play a more significant role in the structural ontology; on the other hand, the priority of structures is retained, in accordance with ante rem structuralism.
The relevant conception of thin mathematical objects should be distinguished from Linnebo's (2018) idea of thin objects as entities which "do not make a substantial demand on the world", that are based on fregean abstraction principles. The articulation of WMS largely presupposes the idea of thin objects as appealed to in the philosophy of science, where weak forms of structuralism have been already presented as alternatives to Ontic Structural Realism, that eliminates individual objects tout court.

Shapiro understands mathematical objects in terms of a 'places-are-objects' perspective (distinguished from the weaker 'places-are-offices' perspective), in which the difference between a place and an object is a relative one and (empty) places may qualify as legitimate mathematical objects. However, one may argue that places/objects of this kind are still too structurally interpreted, thus resulting in a position where there are no objects at all – similarly to what happens with OSR in the scientific domain. I will propose a variation of the 'places-are-objects' perspective in which mathematical objects are something more than mere position, but something less than the thicker objects which occupy these positions in the concrete systems. I will define this position Weak Mathematical Structuralism, and such objects thin mathematical objects.

The difference between Shapiro's purely structural entities and thin mathematical objects can be firstly investigated by considering their structural and non-structural properties.
In fact, structural objects are typically described as possessing structural properties only, which determine their essential identity: what they really are as opposed to all the other objects in the same structure. On the contrary, thin objects are endowed with both essential-structural properties and non-essential, non-structural properties. Such properties, even though they do not determine their essential identity as individuals, will be useful to introduce them as numerically distinguished relata, conceivable independently of the structure they belong to.
In line with Linnebo (2008), structural properties can be described as the properties that can be inferred through a process of abstraction (e.g. Dedekind's abstraction) or, similarly, as the properties that are shared by every system that instantiates the structures.
Still, this definition is subject to different counter-examples, concerned with non-structural (nonintrinsic) properties of objects. Linnebo (2008, pp. 5-6) presents such properties as follows:

«the number 8 has the property of being my favourite number. It also has the property of being the number of books on one of my shelves. And it has non-structural properties such as being abstract and being a natural number. In fact, the property of being abstract seems to be a very important property of natural numbers».

Here, different non-structural properties are mentioned: intentional properties (e.g. “being my favourite number”), applied properties (e.g. “being the number of books on one of my shelves”), metaphysical properties (e.g. “being abstract”) and kind properties (e.g. “being a natural number”). The analogy with scientific Weak Structural Realism (WSR) and thin objects in the philosophy of science will clarify which of them can better support the present account.

I focus on those formulations of WSR which appeal to ontological dependence. This notion displays some features which are crucial in defending a non-eliminative approach to individual objects. French’s (2010) describes WSR as follows (p.17):

The identity of the putative objects/nodes is (asymmetrically) dependent on that of the relations of the structure.

According to French, the asymmetrical notion of dependence at play is adequately captured by Lowe’s (2005) identity dependence (ID):

(ID): x depends for its identity upon y = df There is a two-place predicate “F” such that it is part of the essence of x that x is related by F to y.

WSR is distinguished from Moderate Structural Realism (MSR) and Ontic Structural Realism (OSR). MSR introduces a mutual relation of dependence between objects and structures, that are ontologically on a par. As explained by French (2010, p. 16):

the identity of the objects/nodes is (symmetrically) dependent on that of the relations of the structure and viceversa.

This conception includes objects in the ontology, but it is open to criticism with respect to the circularity of the notion of dependence at hand.

Ontic Structural Realism (OSR) more radically reduces objects to their structural features – they solely exist if the relevant structure exists and there is nothing to them (identity, constitution, etc.) which can be defined independently from the structure. French (2010, p. 18) outlines OSR as follows:

the very constitution (or essence) of the putative objects is dependent on the relations of the structures.

OSR appears seriously controversial, because it introduces relations without relata. A non-eliminative stance towards objects has been supported by a reflection on ontological dependence itself: as argued by Wolff (2012), dependence is a non-reductive notion, since to say that B depends upon A means that B is less fundamental than A, and not that it is to be eliminated. For this reason, I will claim that – even though objects are less fundamental than structures – both the relata of the relation should exist, consistently with WSR.

Let us now rehearse WSR’s conception of thin objects, which provides an interesting interpretation of quantum entanglement, by focusing on Esfeld (2004) and Wollf’s (2012) proposals.
Esfeld (2004, p. 11) motivates a non-eliminativist metaphysics of relations for quantum particles in the following terms:

«relations require things that stand in the relations (although these things do not have to be individuals and they need not have intrinsic properties».

This idea allows interpreting physical theories as referring to things or entities (not to the very individual objects) that may exist independently from the relations in which they stand.
On this basis, a first definition of thin objects can be drafted:

1.a) thin objects are things/entities whose essential identity depends upon the relevant structure, but whose existence is to be acknowledged if relations are to be posited.

I submitted the existence of thin objects as not reduced to their essential-structural properties, since it also results in their non-essential, non-structural properties (e.g. state-independent properties of quantum particles).
As illustrated so far, these properties can be interpreted as intentional, applied, metaphysical and kind properties.
Wollf's (2012, p. 3) position includes a plausible explanation of the more precise nature of nonstructural properties as kind properties:

«Particles qua individuals are thin objects. To the extent that we understand their identity as individuals, we understand it in terms of the state they are in. This leaves unaffected their 'kind identity', that is, their identity as electrons rather than muons. Which kind of particles they are does not depend on any particular state the particles are in».

This interpretation seems more promising than its alternatives: in fact, if metaphysical properties are very important properties – actually appearing as non-structural but essential properties of objects – intentional and applied properties seem too contingent for distinguishing thin objects from entirely structural entities.
This leads to a second definition of thin objects:

2.a) thin objects are things/entities that – in addition to their structural properties – possess also non-essential, non-structural kind properties (the properties that qualify quantum particles as electrons, muons, etc.).

Let us consider the permutation of quantum particles in a single state as a more specific example. Quantum particles are indistinguishable in isolation: they can be permuted while leaving the relevant state unchanged. Therefore, solely quantum entanglement structure grounds their identity as individuals. On the other hand, relations of quantum entanglement require things to stand in the relations, and existing metaphysically prior to them (definition 1.a).
As a consequence, the relevant particles cannot collapse in a single one, because they also possess state-independent properties that allow them to be considered – if not as individuals – as numerically distinguished relata; in fact, even though non-structural properties cannot distinguish particles of the same kind, they are able to distinguish particles that belong to different kinds, e.g. electrons, muons, etc. (definition 2.a).
Definitions 1.a) and 2.a) concerning thin physical objects will be useful to describe thin mathematical objects as well. Objects so understood raise two main worries, to which we will come back when addressing WMS:

i) are thin objects substantial enough to avoid resulting in a "no-objects-at-all" position?
ii) are thin objects weak enough to preserve a structuralist framework?

In analogy with WSR in the philosophy of science, I will now delineate Weak Mathematical Structuralism (WMS), focusing on the formulations of non-eliminative mathematical structuralism in terms of ontological dependence and grounding: namely, Linnebo (2008) and Wigglesworth's (2018) proposals.
Linnebo (2008) examines mathematical structuralism through Lowe's (2005) essential dependence. He introduces two different dependence claims: objects depend for their identity upon other objects (ODO, 'objects depend on objects') and upon their structures (ODS, 'objects depend on structures').
Wigglesworth (2018) articulates the relation between objects and structures in terms of metaphysical grounding. In line with Linnebo's essential dependence, two grounding assumptions are introduced: the identity of each object is partially grounded in the identity of all the other objects in the same structure (ODO) and fully grounded in the identity of the structure (ODS).
In both cases, I will leave aside (ODO), that accounts for the (symmetrical) interdependence between objects of the same structure, and I will focus on (ODS), that addresses the (asymmetrical) relation between objects and the structure they belong to.
In this perspective, WMS can be defined as follows:

The identity of the putative objects/nodes is (asymmetrically) dependent on / being grounded in that of the relations of the structure

In analogy with Wollf's (2012) position in the philosophy of science, ontological dependence and grounding can be seen as non-eliminative notions that – for their very metaphysical features – assume the existence of both objects and structure. For this reason, I believe they supply a direct argument in favour of weak structuralism and thin objects also in the mathematical framework.
The comparison between mathematical structuralism and graph theory (cf. Ladyman & Leitgeb, 2008; Wigglesworth, 2018) allows grasping thin mathematical objects more in detail. In this conception, structures are identified with unlabelled graphs, and graphs are formed by nodes and edges between nodes, which can be added or removed. Within these structures/graphs, individual objects can be understood as unlabelled and edgeless nodes in graph, as illustrated in the following figure:

G': ○ ○

In my account, these nodes seem comparable with quantum particles in entanglement states. They are interchangeable because they can be permuted while leaving the graph unchanged; hence, their identity as individuals is solely determined by the relevant graph G'. Nevertheless, the existence of relata is necessary in order to posit the relations themselves, and this is consistent with the non-eliminative framework suggested by the reference to ontological dependence and grounding.
On this basis, thin mathematical objects can be firstly defined as follows:

1.b) thin mathematical objects are things/entities whose essential identity depends upon the relevant structure, but whose existence is to be acknowledged if relations are to be posited.

Significantly, the nodes in question – though interchangeable – cannot collapse into one another, thus resulting in a different (smaller) graph. Exactly as quantum particles, they appear as numerically distinguished relata that are discernible as far as their non-essential, non-structural properties are concerned. As I have suggested, in the context of scientific structuralism nonstructural properties plausibly qualify as kind properties: this interpretation may work also for mathematical structuralism, where thin mathematical objects are endowed with kind properties such as "being a natural number".
Thus, a second definition of thin mathematical objects can be outlined:

2.b) thin objects are things/entities that – in addition to their structural properties – possess also non-essential, non-structural kind properties (the properties that qualify numbers as natural, relative, rational, etc.).

I will now turn to the metaphysical issues (i - ii) presented so far in the philosophy of science, that concern thin mathematical objects as well.
I will face the first issue (are thin objects substantial enough?) by investigating how thin mathematical objects respond to a typical criticism to traditional non-eliminative structuralism – concerning the individuation of objects and its alleged circularity – levelled by Hellman (2001) and MacBride's (2006) against Shapiro's ante rem structuralism. The authors state that even though the identity of objects depends upon the relevant structures, structures presuppose relata having already been individuated or numerically distinguished.
This objection is related to the debate about whether the Principle of Identity of Indiscernibles (PII) can be maintained within the structuralist ontologies. This issue emerges when considering structures that display non-trivial automorphisms (isomorphisms from the structure to itself that are not the identity mappings): they are composed by distinct mathematical objects that – if interpreted as mere positions or empty places – appear as structurally indiscernible. For instance, + 1 and -1 in the relative number structure and +i and - i in the complex number structure.
First, thin mathematical objects as to points 1.b) and 2.b) seem to be substantial enough to partially avoid such problems: in WMS, thin objects have been introduced as things existing metaphysically prior to the structure and being numerically distinguished in virtue of their non-structural kind properties.

This seems to hold in some cases of non-trivial automorphisms as well: in the relative numbers structure, the numbers +1 and -1 appear discernible because +1 belongs to the natural numbers kind, that is a subset of the relative numbers kind.
This solution is not unproblematic, but it has the advantage of not involving either a primitive notion of identity (cf. Ladyman and Leitgeb), or reference to a weak form of PII (cf. Ladyman, 2005; Saunders, 2006) where objects are individuated by the symmetric and irreflexive relations holding between them (e.g. + 1 is the additive inverse of -1). In fact, the former is not completely convincing, as the notion of primitive identity is controversial in the structuralist literature. The latter, according to MacBride (2006), does not actually face the objection, since irreflexive and symmetric relations still presuppose the numerical diversity between objects.

Second, thin mathematical objects are also weak enough, thus responding to the second issue (are thin objects weak enough?). In fact, their introduction in the ontology does not commit to eliminative structuralism, according to which the existence and the individuation of abstract structures depend on the concrete systems instantiating them. By contrast, thin objects, conceived of as unlabelled and edgeless nodes in a graph, are consistent with an ante rem individuation of structures, in which no concrete system is required, as compellingly demonstrated by Wigglesworth (2018).
As required by the asymmetry of WMS, the identity of structure/graphs does not depend on the very identity of objects – which can be permuted while leaving the graph unchanged – but rather on the operation of adding or removing an edge, that would result in a different graph. In a nutshell, the identity of graphs is determined by their isomorphism classes.
This idea is connected to Shapiro's (1997, p. 93) definition of structures:

«We stipulate that two structures are identical if they are isomorphic. There is little need to keep multiple isomorphic copies of the same structure in our structure ontology, even if we have lots of systems that exemplify each one».

On this respect, thin mathematical objects, though substantial enough to be legitimate relata of structural relations, are also weak enough to retain an ante rem individuation of structures.

To sum up, I show that Weak Mathematical Structuralism (WMS) can be introduced in the philosophy of mathematics in analogy with a specific conception of WSR and thin physical objects in the philosophy of science. In particular, the articulation of mathematical structuralism according to ontological dependence and grounding makes the introduction of WMS worth endorsing as a more moderate variety of Shapiro's ante rem structuralism. WMS is presented as a middle-ground position which attempts to overcome some difficulties of non-eliminative mathematical structuralism (circularity in the individuation of objects) without abandoning its main intuitions (priority of abstract structures). This proposal has been developed by elaborating a variation of Shapiro's 'places-are-objects' perspective, so as to obtain thin objects as something more than mere positions but something less than the concrete, "thick" objects which occupy these positions.

Esfeld, M. (2004), 'Quantum Entanglement and a Metaphysics of Relations', Studies in the History of Philosophy of Physics, 35B, 601-617.
French, S. (2010), 'The Interdependence of Structure, Objects and Dependence', Synthese, 175: 89109.
Ladyman, J., (2005), 'Mathematical structuralism and the identity of indiscernibles', Analysis, 65: 218–221.
Ladyman, L., Leitgeb, (2008), 'Criteria of Identity and Structuralist Ontology', Philosophia Mathematica, 16: 388-396.
Linnebo, Ø. (2008), 'Structuralism and the Notion of Dependence', Philosophical Quarterly, 58 (230), 381-398.
Linnebo, Ø. (2018), Thin Objects, an Abstractionist Account, OUP.
Lowe, E. J. (2005), 'Ontological Dependence', in Zalta, E. N. (ed.), Stanford Encyclopedia of Philosophy.
MacBride, F. (2006), 'What Constitutes the Numerical Diversity of Mathematical Objects?', Analysis, 66 (1): 63-69.
Saunders, S. (2003), 'Physics and Leibniz's Principles', In K. Branding, E. Castellani (ed.), Symmetries in Physics: Philosophical Reflection, CUP, 289-307.
Shapiro, S. (1997), Philosophy of Mathematics: Structure and Ontology, Oxford University Press.

Wigglesworth, J. (2018), 'Grounding in Mathematical Structuralism', in Bliss, R. and Priest, G. (eds.) Reality and its Structure: Essays in Fundamentality, OUP.
Wolff, J. (2012), 'Do Objects Depend On Structures?', British Journal for the Philosophy of Science, 63 (3), 607-625.

**Andrea Oldofredi** (University of Lausanne)
*An Internal Realist Interpretation of the Primitive Ontology Programme*

Generally the standard formulation of Quantum Mechanics (QM), albeit extremely empirically successful, is not considered ontologically satisfying, being affected by several conceptual conundrums and technical difficulties. Against this background, the Primitive Ontology (PO) programme has been advanced in the first place to overcome these issues. According to it, physical theories - quantum and classical - must connect their mathematical and physical structures to the macroscopic ontology specifying which theoretical entities represent real and fundamental objects in the world, and how these dynamically behave in space and time. For this reason, it is typically assumed that the PO approach aims to restore a realist view in the context of quantum physics.

Recently, new developments of this perspective have been proposed by Allori (2017), who argues that the PO provides the means to overcome the Pessimistic Meta-Induction (PMI), and by Esfeld and Deckert (2017), who endorse an atomistic theory-independent ontology of the natural world which should be valid in all physical domains, from the classical regime to the Planck scale (and beyond).

The aim of this paper is twofold: firstly, I will argue that classical antirealist arguments as underdetermination of theories by empirical evidence and the PMI cannot be overcome invoking the notion of PO. Secondly, it will be shown that the an internal realist interpretation of this approach is more apt to faithfully represent the ontological commitment - and its limits - implied by the endorsement of a given PO theory, capturing the original scope of the PO programme. As I will explain, this characterization offers a number of advantages, among which the possibility to maintain a realist attitude within particular theoretical frameworks without a commitment to any fundamental ontology of the natural world, especially taking into consideration the available physical knowledge concerning the inherent structure of matter, space and time provided by Quantum Field Theory (QFT) and Quantum Gravity (QG), and the current degree of development of the PO theories.

Furthermore, since there seems to be a very few indications in physics that ontology should be scale-invariant, the internal realist view admits the possibility to have different ontologies at diverse energy/length scales, giving also ontological robustness to the physical content of effective theories.

- Primitive Ontology and (Anti)Realism

Notoriously, the PO programme endorses a realist view about the existence of theoretical entities represented by primitive variables in the languages of physical theories Ti, since these directly refer to real objects in the world. Nonetheless, it is useful for our discussion to highlight two aspects of this approach which seem to be often left implicit by its supporters: (i) one's ontological commitment is subordinated to the acceptance of a given theory, (ii) such a commitment depends on the domain of validity of the endorsed theory.

- The Problem of Underdetermination

Discussing the problem of underdetermination, let us firstly consider the notion of empirical equivalence. Given two or more theoretical frameworks Ti, they are empirically equivalent iff every Ti predicts the same results for every experiment performed, so that experimental outcomes corroborate a set of theories. Remarkable cases of empirically indistinguishable theories come from quantum physics: Bohmian Mechanics (BM) is empirically indistinguishable from QM, in turn, they are both equivalent to the Many Worlds interpretation. Even if GRW theory turns out to be the correct description of microphysical regimes one cannot escape empirical underdetermination in virtue of its variants. Hence, we cannot decide

experimentally among the plethora of theories available which is the correct description of the physics at the quantum length scales. The antirealist would then conclude that we cannot know whether the theoretical entities postulated by a particular theory T exist, or more precisely we cannot know whether T provides the correct physical description of the world at a certain scale, although T is observationally well-confirmed.

To overcome this problem, one usually refers to meta-empirical virtues in order to evaluate and compare theories. For instance, albeit BM, the Everettian interpretations (i.e. the many worlds and many mind interpretations) and standard QM are empirically equivalent, philosophers and physicists turn to metaphysical criteria such as ontological clarity, unification, explanatory power, internal coherence, formal simplicity, etc. to evaluate the pros and cons of these theoretical frameworks and prefer one over the other. Interestingly, many consider BM superior to the other options for its conceptual clarity and explanatory power, while others reject this conclusion since BM currently has not yet been successfully extended to the relativistic regimes, or because QM is formally simpler. Still, other prefer the MWI over QM or BM, and so on.
The debate concerning which is the correct (or best) interpretation of quantum mechanics, however, has not reached a widely accepted conclusion among experts working on foundational issues, since the selection of which meta-empirical virtues are relevant in order to compare and evaluate rival theories is still a subjective decision (The term subjective has a double reference in this context: it refers to individuals and their preferences, and to groups of researchers part of a sub-community in the field of quantum foundations.), lacking de facto a hierarchy of metaphysical criteria accepted by the overwhelming majority of the researchers involved in this field (To this regard, the reader may find sociological evidence for this claim in surveys as Schlossauer et al. (2013), Norsen and Nelson (2013), and Sivasundaram and Nielsen (2016) concerning the attitudes of professional physicists and philosophers with respect to foundational issues in quantum theory.).

One faces an analogous situation considering solely the most important PO theories - i.e. BM, GRWm, GRWf - among which is not trivial to establish which one is superior over another. To this regard, one can argue that BM is preferable with respect to GRW theories since the latter have less explanatory power due to their peculiar ontology (see notably Esfeld (2014)). However, many physicists and philosophers are inclined to prefer them for the possibility to test their predictions against those of QM, contrary to the case of BM. Another argument to prefer GRW-type theories comes from their promising relativistic extensions, while others may claim that BM is more adequate to obtain a clearer theory of the classical limit, which is harder to get in the GRW context. As above, the choice of which metaphysical criteria should be employed in order to establish which theory is metaphysically superior is left to subjective decisions, since there are sound arguments to prefer each of these proposals. Hence, the epistemological
problem posed by underdetermination seems to be left untouched by the set of metaphysical virtues possessed by PO theories. Moreover, in the context of PO theories there exist also cases of proper metaphysical underdetermination. An example is given by the several proposals for the ontology of the wave function in BM, whereas some authors regard it as a physical object, other conceive it as a nomological entity, others as a parameter figuring in the dynamical structure of the theory. All these alternative formulations of the ontology of BM are characterized by notable meta-empirical features, so that it is difficult to provide a knock-out argument to rule out one of them. A second example comes from the spontaneous collapse theories. Suppose that GRW theories would be the correct description of the physical world, then GRWf and GRWm seem to be metaphysically underdetermined theories being both mathematically and physically coherent, in principle falsifiable, both solve the quantum measurement problem and admit relativistic extensions. In sum, underdetermination can be characterized as a fundamental limitation of science and it cannot be neither eliminated, nor attenuated appealing to PO theories and their meta-empirical virtues.

- PO and the Pessimistic Meta-Induction

Recently Allori (2017) proposed a reply to the PMI within the context of the PO approach, arguing that in scientific theories there are entities which persist through scientific revolutions, namely the primitive variables of physical theories. Then, in order to block the PMI, one should be realist about the set of objects maintained in theory changes, i.e. the PO. This argument can be stated as follows:

1. To defeat the PMI it is sufficient to show that some structures of scientific theories are invariant under theory change;
2. The PO carry over through scientific revolutions;
3. The PO is primarily responsible for the success of a given theory;
4. Therefore, there is something (the PO) preserved in physical theories that blocks the PMI and guarantees the empirical successfulness of scientific theories.

Allori considers the transition from classical mechanics to BM since the particle content of the former persists in the latter. However, one should characterize more explicitly what is implied by (2): to claim that a particular PO is actually maintained in every theory change is not sufficient to block the PMI, one should explicitly show that it is the case. If it is true that in the transition from classical to Bohmian mechanics a particle ontology is preserved, we cannot conclude that such ontology will be preserved in deeper theories as QFT, QG or a final unified theory. In fact, a particle ontology is not trivially extendible to QFT also in the context of the Bohmian framework, with the consequence that this ontology will not necessarily survive in future theories. Furthermore, if a GRW theory were the correct description of the quantum world, there would be an ontological discontinuity between the classical and the quantum scales. In this case Laudan's argument trivially applies, since this more fundamental theory will tell us that our former ontological commitment towards the entities of classical mechanics is wrong. Finally, considering the current state of the art in QG, we are open to a variety of ontological possibilities which are discontinuous with respect to any currently proposed PO. Thus, taking into account these theories it seems difficult to carry over Allori's argument (Since PO theories in their original reading are always defined in space and time, to preserve the PO according to Allori's argument becomes remarkably difficult in theories where space and time are emergent phenomena such Loop quantum gravity or causal set theory.). In sum, although there exist examples of theory changes in which a particular PO is preserved, there are several cases (also within PO theories) supporting the opposite claim, showing the restricted validity of Allori's proposal, and the consequent survival of the PMI. From this brief discussion, it may not be implausible to maintain an agnostic position towards the question of what is the correct description for the ontology of the quantum regimes.

- An Internal Realist Interpretation

The suspension of the judgement concerning what is the correct ontology for the natural world does not necessarily imply tout court an anti-realist view.
To maintain a realist ontological commitment towards theoretical entities appearing in the languages of physical theories without being committed to a fundamental, scale-invariant ontology, one can adopt an internal realist interpretation of theories. Given a particular physical theory T, this form of realism allows to be uniquely committed to the existence of those entities in T, in the domain of validity of T. One's ontological commitment, then, becomes strictly theory-dependent and contextual to the preferred theoretical framework. The similarity between this form of realism and the PO programme is evident, since both suggest that one's ontological commitment depends on the acceptance of a given theoretical framework at a certain scale. To better characterize the proposed internal realist view, one may recall Carnap's distinction between ontological questions asked internally or externally a given linguistic framework, as proposed in Carnap (1950). My suggestion is to consider the available PO theories as different, rival linguistic frameworks among which individuals may prefer one over the other on the basis of extra-linguistic criteria as simplicity, explanatory power, falsifiability, unification, fruitfulness, pragmatical utility etc; in this manner it will be possible to be ontologically committed exclusively to the physical content of a specific theory, and to remain agnostic with respect to ontological questions which lie outside its scope. Against this background, Carnap proposed a dissolution of the traditional metaphysical debate concerning realism, stating that the question of ontological commitment is dependent upon the acceptance of a given language, which in turn depends on pragmatic factors as those mentioned above. In his essay, in fact, he claims that in matter of ontology there are two kinds of question one can ask: questions internal or external to a given framework. The first class has to do with existence questions regarding a given set of objects asked within a particular linguistic framework. Interestingly, Carnap provided also a criterion for reality: to be real means nothing more than to be an element of a certain framework or language, therefore, the adoption a given linguistic framework implies as a consequence the acceptance of its entities. For instance, we are committed to the existence of the electromagnetic field "if we agree to understand the acceptance of the reality, say, of the electromagnetic

field in the classical sense as the acceptance of a language LT and in it a term, say 'E', and a set of postulates T which includes the classical laws of the electromagnetic field (say, Maxwell equations) as postulates for 'E'" (Carnap (1956), p. 45).

From this quotation appears clearly that theoretical terms have a functional role for the explanation of a given set of phenomena, and that acquire their meaning only after (i) being inserted within a set of axioms which provides laws constraining the behaviour of the entity in question, and (ii) being well connected with empirical observations through correspondence rules. One's commitment to a given theoretical term, then, is always dependent on the acceptance of the given theoretical framework containing it. On the contrary, external questions concern the ontological status per se of a given set of objects, independently of the acceptance of any linguistic framework. Carnap considered meaningful only the first class of questions, since they can always be formulated and answered within a particular framework, contrary to the external ones.

At this point we may look at the primitive ontology programme in Carnapian terms, in order to evaluate its ability to provide answers to internal and external questions. In the first place, it seems correct to claim that in recent interpretations of this programme one is engaged with external questions either postulating an ontology which is independent of any theoretical framework, or claiming that the PO can defeat the PMI being preserved in future theory changes. Although I certainly reject the extreme Carnapian conclusions for which such projects are devoid of scientific interest, or that are not philosophically meaningful, I also think that given (i) that the traditional anti-realist arguments remain untouched by the PO programme, and (ii) that the current knowledge available of the physics beyond QFT is still tentative and speculative - not to mention the disagreement within the debate about the ontology of QFT - one can affirm that presently the external question about what the fundamental ontology of the natural world is cannot receive answer, or equivalently, that a unique primitive ontology cannot be maintained from the classical regime to the Planck scale.

For these reasons, therefore, I propose to weaken the realist import of the PO perspective and to concentrate solely on the ontology of particular theories valid at specific energy/length scales, as the case of PO theories, so that ontological questions can be meaningfully answered via a careful consideration of their internal structure. In other words, I take the PO theories to be effective theories as well, with the ability to provide a coherent ontology for certain regime. It is worth noting, furthermore, that according to the theory of local beables, a physical theory to be well-defined, although not fundamental, should provide a clear ontological picture for the domain in which it is a reliable description of physical phenomena. In Bell's and Bohm's writings fundamentality is seldom mentioned, so that the internal reading proposed in this paper is perfectly adherent with the original intentions of this programme. Taking into account singular instances of PO theories, answers to internal questions are available after the analysis of their structures, in particular focussing on the specification of their primitive ontologies. It is obvious, for instance, to answer these questions within the frameworks of BM, GRWm or GRWf. Accepting a given theoretical framework - which comes from extra-linguistic factors-, one is committed to the entities figuring in its first principles, with the consequence that one believes and accepts exclusively the statements of the adopted theory. Thus, one is able to save her realist commitment towards a particular class of entities without endorsing a stronger realism, which would not be supported by the available theoretical knowledge.

In conclusion, it is useful to stress which are the advantages of the proposed "Carnapian" interpretation of the primitive ontology theories:

1. The realist attitude of the PO programme is carried over, since the decision concerning the adoption a given framework implies the acceptance of the entities postulated by this framework;
2. To establish a fundamental ontology for the natural world is a different project with respect to the original scope of the theory of local beables which is instead deeply theory dependent, i.e. namely to construct ontologically well-defined quantum theories. Thus, as already said, the interpretation presented in this paper is closer to the initial aims of the primitive ontology programme;
3. The current physical knowledge indicates that ontology may be scale-dependent, or better, not scale-invariant. Approaching the Planck scale ontological reflections become speculative and tentative, so that it is possible to expect that physics will be remarkably different at those energy/length scales with respect to the classical or non-relativistic quantum regimes;
4. Adopting a Carnapian stance would provide more effective answers to the underdetermination, since the adoption of a certain framework is dictated by solely subjective pragmatical considerations concerning its virtues over other rival theories. Moreover, this reading of

the PO theories admits the strength of the PMI, i.e. that it is not defeated by this programme, nor by this new internal realist interpretation.

Allori, V. (2017). Scientific realism and the quantum: Primitive ontology and the pessimistic meta induction. APA Conference, Pacific Division, pages 1 – 11.
Carnap, R. (1950). Empiricism, semantics, and ontology. Revue Internationale de Philosophie, 4:20 – 40.
Carnap, R. (1956). The methodological character of theoretical concepts. Minnesota Studies in the Philosophy of Science, 1(1):38–76.
Esfeld, M. (2014). The primitive ontology of quantum physics: guidelines for an assessment of the proposals. Studies in History and Philosophy of Modern Physics, 47:99–106.
Esfeld, M. and Deckert, D.-A. (2017). A minimalist ontology of the natural world. New York: Routledge.
Norsen, T. and Nelson, S. (2013). Yet Another Snapshot of Foundational Attitudes Toward Quantum Mechanics. https://arxiv.org/pdf/1306.4646v2.pdf.
Schlossauer, M., Kolfer, J., and Zeilinger, A. (2013). A Snapshot of Foundational Attitudes Toward Quantum Mechanics. http://arxiv.org/pdf/1301.1069v1.pdf.
Sivasundaram, S. and Nielsen, K. (2016). Surveying the Attitudes of Physicists Concerning Foundational Issues of Quantum Mechanics. https://arxiv.org/pdf/1612.00676.pdf.

## FIFTH SESSION: Philosophy of Social Sciences

**Giulia Miotti** (Sapienza University of Rome)

*Imperfect Knowledge And Non-Equilibrium In Finance: The Efficientist Approach In The Light Of Fallibilism*

In this paper, I propose an analysis of the theory of financial markets known as efficient market hypothesis and I point out how it meets some critical shortcomings from an epistemological viewpoint. More specifically, I focus on two problematic assumptions of the efficientist approach represented by the assumptions of perfect knowledge of rational agents and of market equilibrium . By means of the notions of fallibility and reflexivity, I show how the efficient market hypothesis, being built on these two assumptions, is theoretically and epistemologically weak and I also claim that taking into account the notions of fallibility and reflexivity would help in overcoming both theoretical and descriptive shortcomings of the efficientist approach.

Following the two tenets of perfect knowledge and market equilibrium, in fact, the efficientist approach provides a unrealistic description of agents' cognitive abilities (as rational agents they are assumed to display rather strong cognitive and computational abilities that allow for a rational behaviour), of the exact epistemological nature of information (which is assumed to be exhaustive, which means that it is fully reflected in prices and objective, which means that it is formed outside of the market itself and therefore independent of market's internal dynamics) and of the possibility of market equilibrium as an outcome of agents' rationality and perfect knowledge (equilibrium amounts here to a condition of "symmetry" between the quantity of goods offered in the market and the quantity of goods required by the market). Furthermore, the assumptions according to which agents make decisions within a context of perfect knowledge and move within a market in constant equilibrium are not only theoretically but also empirically implausible, since they a-priori rule out the (actual) possibility of agents' manipulative abilities on markets and the (empirically not unusual) condition of out-of-equilibrium markets and extreme financial events.

The three assumptions of agents' rationality, of perfect information and of market equilibrium lie at the core of the efficient market hypothesis and are strongly intertwined: the plausibility of each assumption is guaranteed by the plausibility of the other two. I explain this interdependence as follows: I argue that, for instance, the availability of exhaustive information is a necessary condition for agents to behave rationally. The existence of objective sets of data fully descriptive of the environment allows agents to act in a context of certainty, therefore enabling them to maximize their choices. Information, when fully reflected in between the two sides of demand and offer. In its turn, a market in equilibrium provides the agents with prices that can be taken as reliable parameters in order to behave as rationally as possible . A stable market, in fact, automatically rules out all those choices (bundles of offerdemand prices) which are out-of-equilibrium; thus relieving the rational agent of the burden of an almost infinite set of possibilities among which to select and therefore it enables and facilitates the rational agent in detecting the most rational choices.

In this context the possibility of perfect knowledge and consequently of market equilibrium are guaranteed by the ability of rational agents to act on objective information.

This interdependency, in fact, signals the theoretical depth and completeness of the efficient market hypothesis; on the other hand, it also marks its boundaries, since the dismissal of one of its assumptions entails a critical weakening of the whole hypothesis.

I claim that finance, notwithstanding its complex mathematical apparatuses and powerful models, share some significant epistemological problems with other social sciences.

Following the lead of authors such Merton (1948), Flanagan (1981) and Callon (2007), I refer in particular to the problematic relation between the construction of theories, their explanatory content and the phenomena the theories refer to.

As I argued above, in order to show the efficientist approach theoretical and epistemological deficiencies with respect to the problems of knowledge and market equilibrium, I have recourse to the notion of reflexivity as proposed by George Soros (2013) and his "reflexive market hypothesis", in which Soros further develops the above mentioned problems suggested by Merton and Flanagan and provides them with an interesting theoretical framework openly borrowed from the work of Karl Popper, whose notion of "fallibility" is here extended to finance.

According to the theory of reflexivity, knowledge of social phenomena in general and of financial phenomena in particular, is always incomplete. This incompleteness is due to two main principles: the principle of fallibility and the principle of reflexivity; these two principles provide an account of the possibility and characteristics of agents' epistemic efforts and of the effects that such efforts exert on the reality the observers intend to study. The principle of fallibility states that the outcome of any cognitive effort will never consist in a perfect and complete knowledge of the object studied. No scientific statement, in fact, can completely evaluate and describe the state-of-affairs with which it is confronted since it will be biased by "interferences" due, on the one side, to "outside" constraints connected to the complex and multi- layered structure of the world and, on the other, to "inside" constraints represented by the physiological structure and the cognitive possibility of our mind, endowed with limited computational possibilities and whose reasoning abilities can be substantially influenced by the interference of emotions or, for example, cognitive and computational biases. Whereas, according to Soros' analysis, the principle of fallibility can be applied to both natural and social sciences, the principle of reflexivity attains only to social sciences. It postulates the possibility of feed- back loops between what might be addressed as the "inside" reality represented by the hypotheses and beliefs formulated by the observers and the "outside" reality represented by facts and phenomena that the observers intend to describe. According to this principle, it could be argued that within social sciences, and finance in particular, the cognitive actions and expectations of agents (even in their role of observers) do not play a neutral role with respect both to theories and to the reality these theories are confronted with. It could be argued, in fact, that whereas in the natural sciences the direction of inquiries moves from facts to facts (even though we are subject to fallibility), in the context of social sciences inquiries move from facts to agents' perceptions and further from agents' decision and hypotheses to facts.

In the light of the description of the structure of the efficientist theory I proposed above, it is clear that the principles of fallibility and reflexivity of theories in finance impair the efficientist hypothesis in its core tenet, i.e. the rationality of agents acting according to perfect knowledge and would therefore rule out the possibility of a market in constant equilibrium.

On the one side, the principle of fallibility takes into account cognitive and computational limitations and biases which would not allow for the attainment of perfect knowledge by market agents; on the other side, the principle of reflexivity undermines the possibility of perfect information since information, as hypotheses and theories, might be subject to feedback loops. In doing so, however, the introduction of the principles of fallibility and reflexivity in finance theories seems able to introduce to a theoretical analysis the crucially important notion of imperfect knowledge and the possibility of out-of-equilibrium markets.

Callon, M. (2007) What Does it Mean to Say That Economics is Performative? In: Do Economists make Markets? MacKenzie, D; Muniesa, F.; Siu, L. (eds.) pp. 311- 357.

Fama, E. (1969) Efficient Capital Market: A Review of Theory and Empirical Work. In: The Journal of Finance, Vol.25, No.2 Papers and Proceedings of the Twenty- Eight Annual Meeting of the American Finance Association. pp. 383- 417.

Flanagan, O.J. (1981) Psychology, Progress, and the Problem of Reflexivity: A Study in the epistemological Foundations of Psychology. In: Journal of the History of Behavioral Sciences: 17. pp. 375-386.

Ippoliti, E. (2015) Dynamic Generation of Hypotheses: Mandelbrot, Soros and Far- From Equilibrium. In: Heuristic Reasoning, Ippoliti, E. (ed.) Springer, Dordrecht.

Merton, R.K. (1948) The Self-fulfilling Prophecy. In: Antioch Review, 8: 193-210.

Morgan, M. (2012) The World in the Model. How Economists work and think, Cambridge University Press, Cambridge.

Popper, K. (1959) The Logic of scientific Discovery, Routledge, London.

Preda, A. (2009) Information, Knowledge, and Economic Life. Oxford University Press, Oxford.

Soros, G. (2013) Fallibility, Reflexivity and the Human Uncertainty Principle. In: Journal of Economic Methodology, 20: 4. pp. 309-329.

**Stefano Vaselli** (University of Turin)
*Is Methodological Individualism Without Ontological Individualism Possible?*

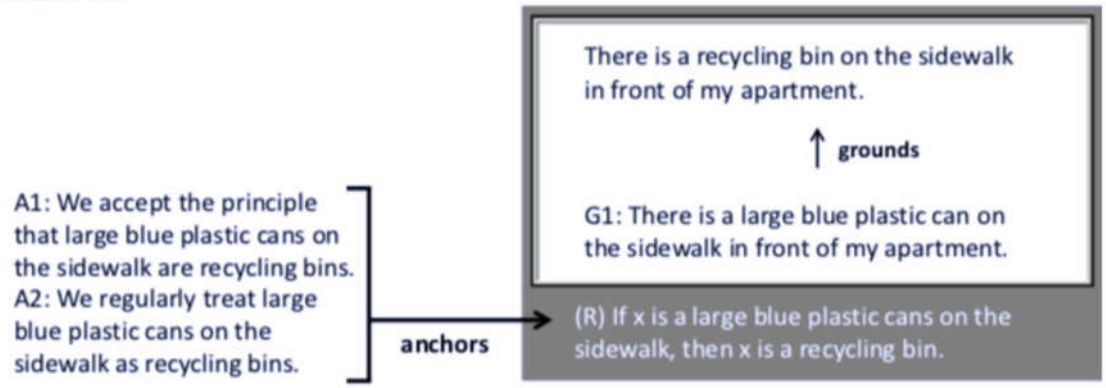
This paper is an attempt to explain how and why any conceptual taxonomy of individualism in social sciences cannot ground on a objective separation from the ontological state of individuals and their properties and relations. In other words: if the (in)existence of a social level is not completely susceptible of a "ontological reduction" or "elimination" (in the Quine' sense of the term) to individualistic level, then none individualism is theoretically possible, methodological individualism included.
Thus, if our view is correct, everyone wants to defend any sort of individualistic interpretation of sociology, history, economics, social psychology must to be ready to state some exemplary of ontological exhaustive reduction, or quinean elimination, of the social to the individual level. But, if ontological individualism has got relevant problems as a metaphysical thesis about the social reality, then this impossibility to separate or split up the explanatory from the ontological view finishes to condemn to fail any attempt to ground in a reliable way any individualistic explanation of the social world, in virtue of being ontological individualism a more general, global and universal premise to consider – in a realist and externalist view of philosophy of social science' ontology – what really exists in the "catalogue of being" (according the ontology/metaphysical distinction furnished by Achille Varzi [2003, 2005]). In our proposal this is clearly the case: the ontological vices of ontological individualism reverberate and reflect on methodological and explanatory individualism with all their criticisms.
This proof proceeds in two parts.
First Parts: Debate on ontological/methodological distinction in analytical philosophy of social sciences:
Actually, a very classical distinction well-grounded in Philosophy of Social Sciences and, at the same time, in contemporary social ontology, is between methodological (explanatory) individualism and ontological individualism. The first thesis is about the methodology of the social sciences: it holds that explanations of social facts or phenomena should be individualistic (Schumpeter 1908 for the first, classical, definition of "methodological individualism", see also Zahle and Kincaid, 2017). The second is a thesis about the nature or metaphysics of social facts, events, objects, and it holds that there is nothing to social fact "over and above" facts about individuals and certain relations between individuals (Hayek, 1948a, b, 1988).
According Brian Epstein (2014a, b) the second doctrine is a thesis in inter-level metaphysics, and we can agree the first thesis without any acceptance of the second. In fact, the social/individual dualism is a claim about the relation between entities at the social level and entities at the individualistic level; i. e. we can speak of "high-level" and "low-level" entities, and draw useful analogies with the mental and the neural, or with the biological and the chemical, or, more deeply, with the chemical and the microphysical. Yet, following Epstein (2015) there are strong reasons to be skeptical about a satisfactory feasibility of hierarchical picture of distinct levels to the social/individual dualism to give a granted explanation of the "supervenience" of social on individual level. Jaegwon Kim (2002), for instance, has argued that we should think of levels as increasingly inclusive sets, where the higher levels include everything at the lower levels and more. William Wimsatt (1994) has argued that the sciences (social sciences included) are too interlinked to be arranged in levels at all, and that the closest we can get is different scales of aggregation. Philip Pettit (1993) has argued that individual attitudes are partly constituted by social entities, and many people have argued that the individual is socially constituted, starting by the classical contributes by Michel Foucault (1970) and others – though often these claims are very vague as to whether this is an ontological or a historical thesis.
Now, all of these statements are severe threats to ontological individualism as above defined. If we cannot distinguish the social from the individualistic in the first place, then we cannot clarify the thesis of ontological individualism. Providing satisfactory treatments of the individualistic base and the social facts amount to distinguishing these levels from one another, and, for this purpose it becomes more crucial to delineate the individualistic so that it is distinct from the social, than to give a complete account of the social. But, if the sceptics like Epstein about levels are correct, these cannot be achieved, and ontological individualism is no thesis at all.
To bypass and circumvent this problem and to save explanatory individualism from the consequences of possible collapse of ontological one, the same Epstein, following an humean insight and some corollary consequences by Searle' theory of Collective Intentionality – given in Searle (1995, 2010) – has formulated the so-called Anchor Individualism claim, which is the thesis that social facts are exhaustively anchored by individualistic facts.

To introduce the idea of anchoring, consider the recycling bin in front of an apartment. In our town, the bin designated for recycling are large green plastic cans. Black cans are for non-recyclables. This is not a natural fact about the cans or a matter of their intrinsic functionality. A black can would work as well for recycling as a green one. But as it is, when we throw recyclables into the black can or non-recyclables onto the green one, we risk a heavy fine from City Police. What makes it the case that green cans are recycling bins, and black ones bins for non-recyclables? Following Searle, the reason being a large green plastic can is the condition for being a recycling bin, is that we (e. g. in our town) collectively accept a constitutive rule for recycling bins. A "constitutive rule", in the Searle' mainstream, is a formula like:

(R) If x is a large blue plastic can on the sidewalk, then x is a recycling bin.

Thus, we can represent an "Anchor Approach" to the problem of supervenience of social on individual as a "grounding explanation" in the following schema #1.

Schema #1.




(Epstein, 2014a, p.19)

So, Epstein (2014a, 15, 16) tries to argue for a grounding individualism metaphysically disentangled by any sort of strong, causal (ontological) commitment with the very far plausible two principal (1-2) thesis of ontological individualism:
1. The social reality globally supervenes on individuals and
2. The social reality globally supervenes on individuals and interactive relations between them.
Effectively, the anchor individualistic explanation doesn't need for any direct supervenience of social reality on individual (Thesis 1) because an "anchor" could be just a constitutive rule of the Searle's sort "X counts as Y in R" as well as a Hume' social habit evolved in a social convention, as well as a D. Lewis' "convention". In all these cases the ontological commitment is really a minimal one, or, this is the real aim of Epstein' proposal, completely reduced to a mental or a linguistic representation.

Second Part: Any sort of distinction between ontological and "simply methodological" individualism is, at least, mistaken.
This last is an apparently counterintuitive claim, but we must underline the adverb "apparently", because, on the contrary, we ought to ascertain the truth of exactly reversal thesis. In history of science the discovery of not existence of some, extraordinarily or theoretically crucial entities as the Ptolemaic Sky of Unmovable Stars, or caloric, phlogiston, (luminous) ether, or the discovery that to consider Pluto a planet of solar system was a physical mistake, and so on, are all clear examples of how discoveries of ontological fail have been fatal to many and many wrong explanations of respective problems in natural sciences as astronomy, physics, chemistry, with consequent increasing progress in our knowledge of physical and chemical nature. The evidence for argue that ontological precedence of the discovery of existence or unreality of one or more entities in a given model of reality is actually a methodological priority for science in relation to epistemological plane of explanation is less weak view than the contrary, classical, "received view" of the "theory laden nature" of theoretical entities in scientific model. Following this pre-theoretical intuition

Maurizio Ferraris (2009, 2012) maintains the fact that, especially in social sciences and in the widest context of "New Realism", epistemology is grounded on ontology and stated the impossibility of the reversal claim. If this is the case, to show the fallacy of ontological individualism becomes an important premises to conclude the problematic status of any other sort of individualism, starting from methodological one.
Furthermore, Epstein' Anchor Individualism, and any similar defenses of distinction between ontological and methodological are, in turn, grounded within a controversial distinction, the intrinsic/extrinsic dichotomy. In fact, claim (1) is not hard to make sense of, because to do so, we only need to distinguish intrinsic from extrinsic properties. For instance, an intrinsic property of Benjamin is some like

3. Benjamin is one metro ninety centimeter tall

On the other hand, among the extrinsic properties we can consider that

4. Benjamin is taller than Matteo Salvini

Or a socio-economic property of a State exemplified in (5)

5. Italy is on the verge of defaulting on its sovereign debt.

In Social Ontology individual's intrinsic properties are those she has in isolation from the rest of the world, properties like her neural states, bodily structure, physical behaviours, etc (of course, a person's neural states might have been caused by external factors, but, in Epstein's view, do not ontologically depend on those). According author as Hodgson and Epstein, when we talk about "supervening on individuals" in thesis (1), we presumably mean on intrinsic properties of individuals. For this reason, Hodgson asserts, in objecting to (1) that intrinsic properties of individuals do not suffice as the supervenience base for social properties, while, on the contrary, Epstein defends (1) from the Hodgson criticism, assuming the (conceptual) existence of anchor individualism, that is intrinsic-properties grounded.
Now is widely plain that in social world the intrinsic/extrinsic distinction is really breakable. In analytic philosophy of language and mind, since the classical wittgensteinian argument against private language it has been impossible demonstrate, even as a matter of principle, the possibility of a realm of semiotic representation as concepts, signs, meaning of words and propositions, completely self-grounded as something of "perfectly intrinsic" one. If a private language is impossible then is impossible the existence of something like a "language intrinsically meaningful and intrinsically provided of references". As the same Wittgenstein remarks:

"If I say of myself that it is only from my own case that I know what the word 'pain' means - must I not say the same of other people too? And how can I generalize the one case so irresponsibly?
Now someone tells me that he knows what pain is only from his own case! - Suppose everyone had a box with something in it: we call it a 'beetle'. No one can look into anyone else's box, end everyone says he knows what a beetle is only by looking at his beetle. - Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. - But suppose the word 'beetle' had a use in these people's language? - If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a something: for the box might even be empty. - No, one can 'divide through' by the thing in the box; it cancels out, whatever it is.
That is to say: if we construe the grammar of the expression of sensation on the model of 'object and designation' the object drops out of consideration as irrelevant."(Wittgenstein, 1953, 1958, §293)

In Wittgenstein's view to define the individual, physical height of Benjamin as an "intrinsic property" in virtue of some privately satisfied property of Benjamin is an absurdity exactly as to pretend a perfect and not circular definition of "meaning of 'beetle' only by looking at the beetle which is contained in the box we have in our own hands, because to define something as a "individual height" we necessarily need for a system of measurement in metros and centimeters, (continental metric system) or feet and inches (American and Commonwealth system), and this system is something of real, objective given in a social world just before the existence of our individual physical height as something of "intrinsically separate" from the social world of measurements. But this is only the first argument against the ontological precariousness of

“intrinsic”. I call this argument the “Private Language Lurking Dependence” of “Intrinsic” Argument. This statement introduces the second, pivotal, claim of our proof, the argument of “Four-dimensional” nature of people (normally called “individuals”) in space-time.
According the “Four-dimensional” claim of personal identity in the space-time (the Einstein-Riemann-Minkowsky Space-Time), also called as “Perdurantism” (Sider, 1996) there is nothing some as a “individual substance” completely enduring in a time separated from space. Perdurantism or perdurance theory is a philosophical theory of persistence and identity which maintains that an individual has distinct temporal parts throughout its existence. Perdurantism is usually presented as the antipode to endurantism, the view that an individual is wholly present at every moment of its existence. Now, when defenders of ontological and methodological individualism refer to “individual” in their theories, they clearly use “individual” in the endurantist sense of “Individual”: a substantial entity wholly present at every, historical and existential, moment of its life. In this way individualist theorist overlook and disregard the eventuality that our personal existence could not be the singular endurance of an individual substance (with her intrinsic properties) but the historical perdurance of a whole collective community of stages, every of which being a “snapshot state of affairs” of the entire process or event “Person”. But if this is the case then the intrinsic substantialism of every kind of individualism could be considered as a complete misunderstanding of real nature of what we really are as human being, and this just before to consider the relevance of methodological individualism as a satisfying exit strategy to explain the emergence or the supervenience of social world in our human structures.

Currie, G. (1984). «Individualism and Global Supervenience». British Journal for the Philosophy of Science, 35, 345-58.
Epstein, B. The Ant Trap: Rebuilding the Foundations of the Social Sciences, Oxford University Press, 2015.
——— «Why Macroeconomics does not Supervene on Microeconomics," Journal of Economic Methodology 21, No. 1 (2014b), 3-18.
——— "What is Individualism in Social Ontology? Ontological Individualism vs. Anchor Individualism," in Rethinking the Individualism/Holism Debate: Essays in the Philosophy of Social Science, ed. by Finn Collin and Julie Zahle, pp. 17-38. Dordrecht: Springer, 2014a.
——— (2008). «When Local Models Fail». Philosophy of the Social Sciences, 38, 3-24.
——— (2009). «Ontological Individualism Reconsidered». Synthese, 166(1), 187-213.
——— (2011). «Agent-Based Models and the Fallacies of Individualism». In P. Humphreys and C. Imbert (Eds.), Models, Simulations, and Representations (pp. 115-44). New York: Routledge.
——— «Social Objects Without Intentions». In A. Konzelmann Ziv, H. B. Schmid and U. Schmid (Eds.), Collective Intentionality and Social Ontology. Berlin: Springer.
Fine, K. (2001). The Question of Realism. Philosopher's Imprint, 1(1), 1-30.
Foucault, M. (1970). The order of things: an archaeology of the human sciences. London: Tavistock Publications.
Hayek, Friedrich August Von (1988) «Scientism and the Study of Society» in Economica, Vol IX, e in The Counter – Revolution of Science, Glencoe, Illinois, The Free Press, 1952.
Hayek, F. A. Von, (1948a) Individualism and Economic Order, Routledge & Chicago University Press, Chicago.
Hayek, F. A. Von, (1948b) «Economics and Knowledge» (1937) in Individualism and Economic Order, cit. 35-56
Hayek, F. A. Von, (1978) New Studies in Philosophy, Politics, economics and the History of Ideas, London-Chicago, Routledge & Chicago University Press,.
Hayek, F. A. Von, (1986) «The Mirage of Social Justice» (1976) «The Political Order of a Free People»(1979) in Law, Legislation and Liberty.
Hodgson, G. (2007). «Meanings of Methodological Individualism» in Journal of Economic Methodology, 14(2), 211-26.
Hume, D. (1777 / 1975). Enquiries Concerning Human Understanding and Concerning the Principles of Morals. In: L. A. Selby-Bigge, ed., Oxford: Clarendon Press.
Kim, J. (1998). Mind in a Physical World. Cambridge: MIT Press.
——— (2002). «The Layered Model: Metaphysical Considerations» in Philosophical Explorations, 5(1), 2-20.

McLaughlin, B. and K. Bennett (2005). «Supervenience». In: E. Zalta, ed., Stanford Encyclopedia of Philosophy.
Pettit, P. (1993). The Common Mind. New York: Oxford University Press.
Putnam, H. (1975). «The Meaning of 'Meaning'». Philosophical Papers. Cambridge: Cambridge University Press.
Searle, J. R. (1995). The Construction of Social Reality. New York: Free Press.
——— (2010). Making the Social World: The structure of human civilization. Oxford: Oxford University Press.
Schumpeter, J «On the Concept of Social value». Quarterly Journal of Economics, volume 23, 1908-9. Pp. 213-232
Sider, Theodore (1996) «All the World's a Stage» Australasian Journal of Philosophy, 74, 433-453;
Sider, T. (1997) «Four-Dimensionalism» in Philosophical Review, 106, pp.197-231.
Thomasson, A. (2003). Foundations for a Social Ontology. ProtoSociology, 18-19, 269-90.
Varzi, A. C., e Casati, R. (2014) «Events» in Stanford Encyclopedia of Philosophy,
in https://plato.stanford.edu/entries/events/#StaDynEve
Varzi, A. C. Ontologia, Bari, Laterza, 2005
Varzi, A. C., Parole, oggetti, eventi e altri argomenti di metafisica, Carocci, 2003
Varzi A.C. e Casati R. (2002), «Un altro mondo?» in Rivista di Estetica, 19, pp.131–59.
Weatherson, B. (2006). Intrinsic vs. Extrinsic Properties. Stanford Encyclopedia of Philosophy.
Wimsatt, W. (1994). «The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets». Canadian Journal of Philosophy, Supplement Vol. 20, 207-74.
Wittgenstein L. (1967), Philosophische Untersuchungen. En. Tr. Philosophical Investigations , Cambridge, (1953 - 1958);
Zahle, J. (2006). «Holism and Supervenience». In S. P. Turner and M. Risjord (Eds.), Philosophy of Anthropology and Sociology (pp. 311-42). Amsterdam: North Holland.
Zahle, J. (2014). «Holism, Emergence and the Crucial Distinction», in J. Zahle and F. Collin (Eds.), Rethinking the Individualism-Holism Debate: Essays in the Philosophy of Social Science (pp. 177-196). Dordrecht, Springer.

Zahle, J. (2016). «Methodological Holism in the Social Sciences», in E.N. Zalta (Ed.) The Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/archives/sum2016/entries/holism-social.
Zahle J. Kincaid Harold (2017) «Why be a Methodological Individualist» in Synthese DOI:10.1007/s11229-017-1523-8

## SIXTH SESSION: Philosophy of Biology and Health Sciences

**Federico Boem** (University of Milan), **Stefano Bonzio** (Marche Polytechnic University) **Barbara Osimani** (Marche Polytechnic University; LMU Münich)

*The Cochrane case: an epistemic analysis on decision-making and trust in science in the age of information*

Introduction

Recently, the Cochrane1, one of the most important independent scientific institutions concerning the review of clinical and health practices, has been invested by a heated controversy which can have important repercussions on both the world of clinical research and the perception of it by the general public.

Last September, during the twenty-fifth Cochrane Colloquium, a meeting of Cochrane representatives dedicated to the discussion of the soundness and solidity of the tools and evidence (i.e. the criteria of Evidence Based Medicine or EBM, according to which decisions, concerning the efficacy and potential harm of drugs and other medical devices, are taken in the biomedical world), one of the members of the Nordic Cochrane Center and co-founder of Cochrane itself, the Danish scientist Peter Gøtzsche (since 2010, professor of Clinical Research Design and Analysis at the University of Copenhagen), was accused, by the leadership of the Cochrane, of misconduct and subsequently expelled from the Cochrane itself. The event actually concludes a long fight between Gøtzsche and the new leaders of the organization. Gøtzsche accused Cochrane board of being increasingly prone to the economic interests related to the business produced by biomedical research and progressively less concerned with the robustness and solidity of scientific work. Roughly speaking Gøtzsche claims that Cochrane seeks money rather than "truth". As a matter of fact, his expulsion is a direct reaction to his critical move.

Peter Gøtzsche is not completely new to this type of critique. Throughout his entire career, Gøtzsche has often raised doubts about methodological and ethical issues concerning biomedical research. He particularly focused on meta-analysis, suggesting ongoing issues in data-extraction [4] and advocating for a broader and more solid perspective in this field [7]. Moreover he also called attention towards scientific misconduct [5]2. Gøtzsche has long history of clashing with the pharmaceutical world (e.g. publicly criticizing the way psychiatric drugs are prescribed and used3). He is also famous for having harshly criticized public health policies, such as mammography screenings [6], generating an intense public debate. More recently, in 2013, he published a book entitled "Deadly Medicines and Organized Crime: How Big Pharma has Corrupted Healthcare", in which he denounces the pharmaceutical industry, both from the scientific stance and in its financial dimension, blaming it for immoral (even illegal) behavior and supporting the need for a radical reform of the entire sector.

In July 2018, Gøtzsche and two other colleagues published an article [10] criticizing a Cochrane meta-analysis [1] (produced by another group), questioning the results concerning the safety of papilloma virus (HPV) vaccines. According to Gøtzsche and his colleagues, that review was unreliable and compromised by different bias (including cherry picking, reporting bias and biased trial designs). Gøtzsche article was also, and above all, a clear accusation of superficiality (or even worse, of misconduct), for the perpetration of several methodological errors and for ignoring almost half of the studies on the HPV vaccines. Moreover, and probably more seriously, it has been also claimed that those studies had already been pointed out to the authors of the review, thus suggesting a deliberate exclusion of them. Lastly, Gøtzsche and colleagues also insinuated that the review presented serious issues concerning conflicts of interest, implying that such aspects were uncritically presented to the public.

The fact that Gøtzsche had advanced these objections on the official Cochrane letterhead (an element that, some argued, could be taken as evidence of authority by anti-vaccine movements) has been severely condemned by the Cochrane leaders. Gøtzsche was thus accused of "bad behavior", responsible of discrediting the Cochrane and potentially contributing to public distrust towards science. Therefore, according to the internal regulations4, he was considered subject to expulsion, which promptly took place. In disagreement with this decision, four other members of the councilor resigned, followed (for technical reasons that allowed the board to remain in office) by two others.

The aim of our contribution is neither to solve this dispute nor to take side in such a delicate issue. This is because the result of this controversy goes beyond simple disagreement among experts. Rather, it could have an impact to the entire world biomedical community and its public perception.

This is due to the fact that the Cochrane is not just a simple organization. It is quite unique and precious, considering the issues and the difficulties of scientific research. In an age of crisis of scientific publishing (a

sector often infested by "predatory magazines", where, by paying, anything can be published) current scientific enterprise is facing the so-called reproducibility crisis [8]. Although misleading according to some scholars [3] the expression describes a situation also worsened by the fact that an immense amount of results, especially negative or unfavorable [12], are likely to remain unknown5). This has the potential of threatening both the efficacy and the trust of science itself. Moreover, the increasing lack of independent funds bends many scientific researches to external interests (not always for the sake of knowledge). Finally, EBM definitely provides crucial information, but it must be integrated with other elements, such as feasibility and economic sustainability, patients' preferences and needs, in order to furnish clinical recommendations useful to physicians and patients.

Because of that, a correct evaluation of research, the soundness of its methodology and the range of its implications, is not an easy task. Even for professional scientists there are too many studies, too many data, too many specializations, too many different areas of investigations, tools, approaches. Bearing this in mind, the Cochrane activity has begun and has been pursued following the idea that science is a collective, collaborative enterprise. The purpose was to combine experts able to collect, select and analyze the data emerged from the different studies published on a given topic, in order to respond to a clinical question with a clear, precise and solid review.

It is not the aim of the present analysis to determine who was/is right in relation to this event. However there is something very important that needs to be addressed.

Indeed, the case of Cochrane clearly exhibits a tension between two fundamental values of contemporary technological society: the right to inform and research freedom. On one hand, science rests on critical thinking. In other words, it is at core of scientific practice the possibility to question its own methods and organization. On the other hand, scientific disagreement and debate, always legitimate, can cause troubles in the way science is effectively pursued and perceived, first by scientific community itself, and by the public. A crisis of this kind concerning the Cochrane can be a serious threat for the world of science itself6.

What is at stake?

The Cochrane issue concerns the epistemic possibility of establishing reliable criteria for the assessment of clinical and scientific evidence. The ways according to which science works and is effective is still a philosophical puzzle in many details. From a practical perspective, there are several aspects that might help to determine a good scientific work. Data must be solid as much as the collection strategies adopted to obtain and to organize them (let's call this scientific methodologies). In the age of information, the need to gather and integrate distinct pieces of different types knowledge (obtained by various approaches, via different procedures and certified by different journals) is particularly demanding and yet necessary. Since it is not something that a single scientist or group can do, the chance to delegate to these capable people these very complicated analyzes, is crucial for several reasons:

1. First, there is the recognition that contemporary science (at least in biomedicine) requires a competence which has to be based on multiple forms of expertise, thus shared and discussed with different kinds of experts and checked against non-experts priorities, needs, expectations.
2. Second, researchers need to be free to discuss their results, to question their methods, practices and conclusions at any time, using reliable, reproducible, controlled criteria (see, for instance [9]).
3. Third, the unavoidable delegation of knowledge should rely on trust. The type of trust that experts' judgment will be based on solid, reproducible, controlled research. Without trust, information as such, something we all need to make informed choices, is not enough.

Information and freedom are two key aspects of scientific research. Yet, one may ask, whether there will be ways to combine them in a rational way.

The proposal

Our epistemic analysis aims at providing an operational frame that might serve as an indication or potential guideline, in cases of conflicts when science and the public are concerned. In this respect we In order to do so we plan to work on 3 different, but intertwined levels in the following manner:

a) First, we will address the issue of information disclosure in a game-theoretic approach 7 (how much, to whom, according to which caveats). This is crucial to determine whether rational solutions can be feasible in principle.

b) Second, we will take into account the question of significance of scientific truths [11] in order to analyze the balance between scientific freedom of research and collective interests.

c) Last, we will provide an epistemic analysis of the weight and the quality of information given the specific public context. (e.g. in some contexts, science-prone, debates and disagreement can be seen as a positive

thing, leading to transparency, accuracy and correctness; while in other ones, suspicious towards science, debates and disagreement can be perceived as evidence for misconduct and biases).

[1] M. Arbyn, L. Xu, C. Simoens, and P. MartinHirsch. Prophylactic vaccination against human papillomaviruses to prevent cervical cancer and its precursors. Cochrane Database of Systematic Reviews, (5), 2018.
[2] S. Bonzio, B. Osimani, and A. Sacco. Science as a signaling game: preregistration, and strategic disclosure of clinical trials. Submitted.
[3] D. Fanelli. Opinion: Is science really facing a reproducibility crisis, and do we need it to? Proceedings of the National Academy of Sciences, 115(11):2628–2631, 2018.
[4] P. Gøtzsche, A. Hr´objartsson, K. Mari´c, and B. Tendal. Data extraction errors in meta-analyses that use standardized mean differences. JAMA, 298(4):430–437, 2007.
[5] P. Gøtzsche, J. P. Kassirer, K. L. Woolley, E. Wager, A. Jacobs, A. Gertel, and C. Hamilton. What should be done to tackle ghostwriting in the medical literature? PLOS Medicine, 6(2):1–4, 2009.
[6] P. Gøtzsche and O. Olsen. Is screening for breast cancer with mammography justifiable? The Lancet, 355(9198):129 – 134, 2000.
[7] P. C. Gøtzsche. Why we need a broad perspective on meta-analysis. BMJ Evidence-Based Medicine, 321(7261):585–586, 2000.
[8] J. P. A. Ioannidis. Why most published research findings are false. PLOS Medicine, 2(8), 2005.
[9] T. Jefferson and L. Jørgensen. Redefining the 'e' in ebm. BMJ Evidence-Based Medicine, 23(2):46–47, 2018.
[10] L. Jørgensen, P. C. Gøtzsche, and T. Jefferson. The cochrane hpv vaccine review was incomplete and ignored important evidence of bias. BMJ Evidence-Based Medicine, 23(5):165–168, 2018.
[11] P. Kitcher. Science, Truth, and Democracy. Oxford Studies in the Philosophy of Science. Oxford University Press, 2001.
[12] I. Nygaard. The importance of publishing trials with negative results. American Journal of Obstetrics & Gynecology, 216(6):541–542, jun 2017.

**Silvano Zipoli Caiani** (University of Florence), **Federico Boem** (University of Milan) **Gabriele Ferretti** (University of Florence)
*Out of our Skull, within our Skin: The Gut Microbiota and the Extended Mind Thesis*

According to a pretty common idea, the mind is realized by the activity of our brain. This is a very intuitive assumption, which might be accepted, without big problems, by scientists, philosophers, and even common people. On the contrary, a very odd idea about the mind states that the mind extends beyond the boundaries of our skull. This idea is cashed out from part of the philosophical literature after the seminal paper by Andy Clark and David Chalmers (1998): the mind and its main activity, cognition, are not always and necessarily segregated to individual brains. Though very odd, this idea has gained an increasing consensus among philosophers and scientists, triggering a heated debate about the effective material underpinnings of cognition (Adams & Aizawa, 2001; Arnau, Estany, Solar, & Sturm, 2014; Clark, 2010; Coleman, 2011; Rupert, 2004; Wilson, 2014). According to this odd view, also known as 'Extended mind thesis', (EMT), the physical states that make up the human cognitive processes can reach beyond the boundaries of the individual brain, so as to include, as its proper parts, different aspects of the individual's body and environment.
The EMT is a close relative of 'Machine state functionalism' (Putnam, 1967), a famous functionalist conception of the mind. According to this view, what makes something a mental state does not depend on its material constitution, but rather on the causal role it plays in the physical system it is a part of. This led several philosophers to state that the same mental state may be realized in a variety of different physical structures (Putnam, 1960), serving as a premise in any argument for the possibility of Artificial Intelligence. Interestingly, this thesis paves the way for the 'Parity principle' argument at the basis of the EMT: "All the components in the system play an active causal role, and they jointly govern behavior in the same sort of way that cognition usually does. […] The external features here are just as causally relevant as typical internal features of the brain" (Clark and Chalmers 1998, 8 f).

Now, it is possible to distinguish between two different versions of this very odd idea about the mind, depending on how much we are willing to extend the mental vehicles outside of our heads:
Widely extended mind (WEM): Cognitive processes can be conceived as extending outside our heads and bodies, involving objects (such as tech devices) and events in the physical environment.
Narrowly extended mind (NEM): We can consider the possibility that the vehicles of our cognitive processes extend out of our heads but do not exceed the boundaries of our body.
While the EMT is an odd idea, the WEM and the NEM might sound crazy in a different way, but they both share the assumption that the human cognitive processes are not realized only by structures of the brain-nervous system, but also by physical vehicles that can reside outside our skulls. In this respect, both the WEM and the NEM are instances of the EMT.
Although the idea that the WEM involves the EEM has attracted most of the attention of prominent scholars in the fields of philosophy of mind and cognitive sciences (e.g., Hurley, 2001; Menary, 2010; Noë, 2004), it has also been the target of critical arguments (Adams & Aizawa, 2001; Rupert, 2004; Sprevak, 2009).
Notably, the idea that our cognitive processes extend beyond the boundaries of our body meets a series of issues that are nowadays considered classical objections to this odd idea in the literature. Among such issues, there is that of understanding how the cognitive process of a certain agent can be extended to parts of the environment that are not always and permanently linked to her natural body. Indeed, among the remarkable consequences of the WEM, there are those for which our own cognitive processes can be instantiated also in physical objects located far from our own bodies, such as the hard-disks of our computers or even a body part that pertains to a different person (see Piredda, 2017 for a review).
The issue here is that, at least prima facie, the external states are not persistently coupled with the agent's body, unlike those that are internally based. Thus, it could be that some external extension is fortuitously broken, so that the integrity of the cognitive system compromised and can eventually re-instantiated only once the coupling with this state is recovered. Now, since it is mostly unobvious to think to a cognitive system as something scattered in the environment, someone may be prone to consider the WEM as an implausible, and even odder consequence of the EMT. Even though defenders of the EMT replied to this and similar challenges many times (e.g., Clark, 2010; Wilson 2014; Menary, 2006), the WEM is still considered a controversial conclusion.
It should be noted, however, that over the past decades few attentions have been addressed to the fact that the EMT also involves the NEM, namely the conclusion that cognition extends to physical realizers that are located outside of our skull, but within our body – this has not to be confused with the idea that there are cognitive states related to body representations whose vehicle is, however, still in the brain (for a review see Alsmith and De Vignemont 2012).
Here is a very interesting way of thinking about a possible formulation of the NEM: internal extension can be provided by focusing on the functioning of the microbioma-gut-brain axis, that is, on the activity of the human gut microbiota in interaction with the human nervous system. The microbiota is the complex community of different micro-organismal species which reside, in a symbiotic relationship, within humans and other eukaryotes. Commensal micro-organisms populate various areas of the human body, forming specific niches, as the body were an ecosystem. These areas, among the others, are the skin, oral mucosa, and the intestines. In this respect, several studies have shown that the relationship between the gut microbiota and humans is not merely that of a non-harmful parasitic coexistence, but rather that of a functional relationship. Traditionally, these functions were normally associated with digestion processes and, more recently, with the immune surveillance. Notably, recent evidence of the existence of a microbiota-gut-brain functional axis has shown that specific cognitive functions that have been always ascribed only to the functioning of the nervous system, are instead shaped also by the biochemical activity of the microorganisms that inhabit our gut (Agustí et al., 2018; Foster, Lyte, Meyer, & Cryan, 2016; Foster & McVey Neufeld, 2013; Sarkar et al., 2018; Sharon, Sampson, Geschwind, & Mazmanian, 2016). At the same time, it has been shown that the functioning of the microbiota may be directly and causally influenced by events occurring in the nervous system, so to outline a two-ways functional role along the microbiota-gut-brain axis. Indeed, while the activity of the gut microbiota casually shapes the activity of the agent's nervous system, the activity of the nervous system causally influences that of the gut microbiota (Cryan & Dinan, 2012; Rieder, Wisniewski, Alderman, & Campbell, 2017). Based on this evidence, our paper has two main goals.
1. The first goal is to show that there are enough empirical arguments to claim that microbiota has a functional role in allowing high-level cognitive activities. Accordingly, following the NEM, we state that the microbiota-gut-brain system should be considered as a physical part of our extended cognitive system.

Indeed, both sides of this axis evolved and developed in conjunction, so that their functional interaction is required for the accomplishment of specific cognitive functions. To this extent, a series of mental processes that are usually ascribed to the nervous system extend outside the boundaries of our skull, involving also the causal events that characterize the interaction between the human gut and its community of residential micro-organisms.

2. Our second goal is to show that extending the vehicle of our cognitive activity to the microbiota-gut-brain system does not suffer from the same issues that are traditionally addressed to the WEM. Indeed, since the microbiota can be considered a genuine extension of our physical body the NEM does not present the counterintuitive consequence of scattering the cognitive agent trough the environment.

Our argument proceeds in the following way. First we introduce evidence showing that the functional activity of the microbiota-gut-brain axis plays a role allowing the agent to accomplish specific cognitive tasks. Then, by drawing on the recent biomedical literature on the gut microbiota, we argue that it is not an exogenous source of functions, but rather a constitutive part of our organism. Based on this, we argue that the functioning of the microbioma-gut-brain axis represents a case of the NEM. Finally, we show that our version of the NEM faces all the issues that are usually raised against the WEM. We conclude that our cognitive processes are not entirely realized by the activity of our brain, but rather extend beyond the boundaries of our skull, though within our skin.


Adams, F., & Aizawa, K. (2001). The bounds of cognition. Philosophical Psychology, 14(1), 43–64. https://doi.org/10.1080/09515080120033571

Alsmith, A.J.T. & de Vignemont, F. Rev.Phil.Psych. (2012) 3: 1. https://doi.org/10.1007/s13164-012-0085-4

Agustí, A., García-Pardo, M. P., López-Almela, I., Campillo, I., Maes, M., Romaní-Pérez, M., & Sanz, Y. (2018). Interplay Between the Gut-Brain Axis, Obesity and Cognitive Function. Frontiers in Neuroscience, 12. https://doi.org/10.3389/fnins.2018.00155

Arnau, E., Estany, A., Solar, R. G. del, & Sturm, T. (2014). The extended cognition thesis: Its significance for the philosophy of (cognitive) science. Philosophical Psychology, 27(1), 1–18. https://doi.org/10.1080/09515089.2013.836081

Clark, A. (2010). Supersizing the Mind: Embodiment, Action, and Cognitive Extension (1 edition). Oxford: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The Extended Mind. Analysis, 58(1), 7–19. https://doi.org/10.2307/3328150

Coleman, S. (2011). There Is No Argument that the Mind Extends. The Journal of Philosophy, 108(2), 100–108.

Cryan, J. F., & Dinan, T. G. (2012). Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. Nature Reviews. Neuroscience, 13(10), 701–712. https://doi.org/10.1038/nrn3346

Foster, J. A., Lyte, M., Meyer, E., & Cryan, J. F. (2016). Gut Microbiota and Brain Function: An Evolving Field in Neuroscience. International Journal of Neuropsychopharmacology, 19(5). https://doi.org/10.1093/ijnp/pyv114

Foster, J. A., & McVey Neufeld, K.-A. (2013). Gut-brain axis: how the microbiome influences anxiety and depression. Trends in Neurosciences, 36(5), 305–312. https://doi.org/10.1016/j.tins.2013.01.005

Hurley, S. (2001). Perception And Action: Alternative Views. Synthese, 129(1), 3–40. https://doi.org/10.1023/A:1012643006930

Menary, R. (2010). The extended mind. MIT Press. Recuperato da https://researchers.mq.edu.au/en/publications/the-extended-mind

Noë, A. (2004). Action in Perception. MIT Press.

Piredda, G. (2017). The Mark of the Cognitive and the Coupling-Constitution Fallacy: A Defense of the Extended Mind Hypothesis. Frontiers in Psychology, 8. https://doi.org/10.3389/fpsyg.2017.02061

Putnam, H. (1960). Minds and machines. In S. Hook (A c. Di), Journal of Symbolic Logic (pagg. 57–80). New York University Press.

Putnam, H. (1967). The Nature of Mental States. In W. H. Capitan & D. D. Merrill (A c. Di), Art, Mind, and Religion (pagg. 1–223). Pittsburgh University Press.

Rieder, R., Wisniewski, P. J., Alderman, B. L., & Campbell, S. C. (2017). Microbes and mental health: A review. Brain, Behavior, and Immunity, 66, 9–17. https://doi.org/10.1016/j.bbi.2017.01.016

Rupert, R. D. (2004). Challenges to the Hypothesis of Extended Cognition. The Journal of Philosophy, 101(8), 389–428.

Sarkar, A., Harty, S., Lehto, S. M., Moeller, A. H., Dinan, T. G., Dunbar, R. I. M., … Burnet, P. W. J. (2018). The Microbiome in Psychology and Cognitive Neuroscience. Trends in Cognitive Sciences, 22(7), 611–636. https://doi.org/10.1016/j.tics.2018.04.006
Sharon, G., Sampson, T. R., Geschwind, D. H., & Mazmanian, S. K. (2016). The Central Nervous System and the Gut Microbiome. Cell, 167(4), 915–932. https://doi.org/10.1016/j.cell.2016.10.027
Sprevak, M. (2009). Extended Cognition and Functionalism. The Journal of Philosophy, 106(9), 503–527.
Wilson, R. A. (2014). Ten questions concerning extended cognition. Philosophical Psychology, 27(1), 19–33. https://doi.org/10.1080/09515089.2013.828568

**Chiara Beneduce** (University Campus Bio-Medico of Rome)
*"complexio". A systemic approach to organism's dynamics*

I. Introduction: the "system root" of Systems Biology and some theoretical issues in bio-medicine
Systems Biology (SB) is concerned with biological entities conceived as complex systems. The definition and aims of SB are widely debated. This debate is due to the pluralistic vocation of SB itself. In a multi-shaded framework, a distinction between two different "roots" or "views" in SB has returned in literature. A SB's "component root" has been counterposed to a SB's "system root". The first being more concerned with the components of the system (i.e., with the large scale studies of molecules) and the second being focused on the system as a whole, and specifically on what emerges from the interaction between the molecules. The "component root" seems to have followed a more "scientific" or "pragmatic" path since it sees SB as an extension of genomics and molecular biology, while the "system root" has been paired with a "theoretical" soul of SB, more attentive towards a theoretical study of the living organisms as systems. The very demarcation between the two roots can be discussed. However, perhaps due its less "scientific" appeal, it is a matter of fact that the "system root" has been less developed than the "component" one. Nonetheless, it is exactly through a systemic approach that organism's dynamics and some emergent theoretical concepts in the bio-medicine in particular, such as "integrative processes", "dynamic unit" or "dynamic stability", could find a more comprehensive framework of understanding.
In order to reinvigorate the "system root" in the understanding of those concepts, my strategy has been to start from a (quite long) time ago, namely from the medieval theory of complexion. More precisely, in this paper, I show some traits of the late-medieval medical and natural-philosophical concept of "complexion" which recall (and invite us to adopt) a systemic approach towards an understanding of organism's dynamics.

II. Complexion
II.1 Complexion in the Middle Ages
"Complexion" ("complexio" in Latin, "κρᾶσις" in Ancient Greek) is one of the pivotal concepts in Galen's medical theory. It is the blend of the primary qualities (hot, cold, wet, and dry) that results from the mixture of the primary elements (earth, air, water, fire). Since Galen incorporated the Hippocratic idea of "humors" into his medical theory, the concept of "complexion" is also linked to the balance of the four humors, i.e., blood, yellow bile, black bile, and phlegm. The word "complexion" and the word "temperamentum" mostly overlapped in medieval scientific literature. However, "complexion" was used as a more technical term referring to the mixture of qualities, while "temperamentum" was most often referred to the humoral blend. The concept of "complexion" forms the idea of "health" as a balanced bodily state. In other words, a balanced complexion implies bodily health, while the imbalance causes a pathological condition in the organism. The Galenic concept of "complexion" appeared in the Latin cultural milieu through the mediation of the Arabic sources and through the translations of Galen's works into Latin. In the Latin scientific world, the concept of "complexion" played an important role both in medical theories and in medical treatments. Theories on complexion were spread in commentaries on Galen- based works (for example, commentaries on Johannitius' Isagoge), but also in independent treatises in theoretical medicine, (e.g., the Conciliatior by Peter of Abano), or in the medical literature of the consilia (medical doctors' written advices on specific diseases and treatments). The notion of "complexion" was also used in the framework of natural philosophy. Theories on complexion are attested in commentaries on Aristotle's De generatione et corruptione (i.e., philosophical texts that discuss primary qualities and elements) and in more strictly biological texts, like commentaries on Aristotle's De anima, Parva naturalia, and De animalibus. The concept of "complexion" was variously interpreted in medieval science, in both fields of medicine and natural philosophy. However, it

is possible to keep some general ideas of what pertains to the notion of "complexion" in the late Middle Ages.

II.2 Complexion and organism's dynamics

As Joël Chandelier and Aurélien Robert had the merit of putting out, medieval physicians, especially late medieval Italian physicians, described their concept of "complexion" in terms of "substantial quality". My claim is that, by means of the concept of "complexion" as "substantial quality", medieval scientists were able to provide natural philosophy and medicine with an idea of organism as a "dynamic unit" and especially to account for a general understanding of the concept of "dynamic stability".

The substantial quality is neither a substance (the essence of the organism as substantial form, or soul), neither an accident (a mere and contingent material occurrence within the organism's body). Complexion is a sort of biological structure or configuration that lays at the interface of form and matter, soul and body, substance and accidents. Complexion emerges from matter but it is not matter. Complexion is a quality emerging from the particulars (the elements and the humors), without being the particulars of the body able to express that quality when isolated from one another. Complexion is the quality resulting from the integration of the elements and humors in a broader scale. At the same time, that more comprehensive scale in which the particulars result is not the formal principle of the body in the same way as the soul is the substantial form of the body. For, complexion is strictly depending on the material aspects of organisms, and it varies for each singular organism and at different moments of organism's life.

Being at the same time substantial and qualitative, organism's complexion gives reason for both the unity and stability of the organism (its identity through space and time) and for the organism's dynamism (its changes over space and time). In other words, complexion can account for the dynamic stability of the organism, meant in general terms, because it keeps together two important assumptions. On the one hand, a) the organism is not a "thing", but a "processual entity" which mutability and dynamism is due to the strict relationship complexion entertains with the material, accidental, and variable aspects it emerges from. On the other hand, and at the same time, b) the organism maintains its unity and stability because its complexion results from an integration of elements and humors and not just from an aggregation or compound of material parts. In fact, if complexion were just an aggregation of parts it would have been as "accidental" as matter, and nothing could have explained its unity and stability. But, complexion is a substantial quality, which means that elements and humours result in a configuration that is not reducible to the (mere) sum of the elemental parts. This substantial configuration gives account for that unity and stability of the organism that otherwise matter alone would not be able to justify.

My second claim is that, with the concept of "complexion", medieval science underlined the relevance of the context-dependencies in the description of organism's dynamics.

As it has been mentioned above, complexions vary for different individuals and for the same individual in different moments of life. As hinted before, the variability of complexion in a organism (which makes it a dynamic organism) is due to the fact that complexion is linked to the materiality of the elements and humors which lays at its basis. However, is a common trait of medieval discussion on complexion to attribute the variability of complexion, its processual and dynamic trait, not only to the mutability of the material "components" of complexion themselves, but to the more comprehensive role of the context in which the material elements and humors are inscribed. In medieval texts, complexion is described as continuously subject to modification due to contextual factors, such as for example nutrition, geographical settings, celestial influences etc. The context-dependence which shapes the organism's dynamic is well expressed by the Latin word "respectiva", which medieval authors often pairs with the word "complexio". Complexiones are respectivae because their changes are intimately connected to the context they are in. Therefore, it is not just the elemental parts of complexion (its "molecular" matter) that can account for its dynamism, but it is the context in which this matter (the elements and humors) is framed that, by modifying matter, makes the organism's dynamic. Complexion would not emerge as a changing structure if the elements it lays on were existing as isolated parts, independently for the context they are in.

III. Conclusions: organism's dynamics in a systemic framework

Organism's dynamics, organism's dynamic unity and dynamic stability are at table as theoretical concepts in current research in the bio-medicine. Those concepts being the explananda in the bio- medical sciences. The medieval theory of complexion is an example of a successful theoretical effort to explore organism's dynamics, and especially the organism's dynamic stability, in a natural philosophical and medical framework. Complexion as "substantial quality" is in fact able to explain, with one and a same idea, the organism's dynamism and its unity and stability. Most importantly, the medieval notion of "complexion"

shows that an account of the dynamic stability of living organisms is possible (only) within a systemic perspective. The medieval theories of complexion adopt a systemic approach when reading complexion not as the sum of elemental components (an aggregation of parts) but as a quality which emerges from the integration of elements and humors, a structure which lays on matter but cannot be reduced to matter, to the extent that it has a substantial trait too. Another clue that medieval scientists were acting in a systemic framework is their insistence on context. Complexion's dynamics and, consequently, organism's dynamics is due to the interaction between the material source of complexion with the context it is embedded it. And when the context matters we have a good reason to think that a systemic perspective is at play. The case of medieval complexion sheds thus light on the relevance of a systemic approach while studying organism's dynamics, and especially when organism's dynamic stability is at discussion. Which ultimately suggests how much SB could take advantage from its more "systemic root".


Beneduce, C., "John Buridan on Complexion. Natural Philosophy and Medicine in the Fourteenth Century", in: C. Beneduce and D. Vincenti (eds), Oeconomia corporis. The Body's Normal and Pathological Constitution at the Intersection of Philosophy and Medicine, MEFISTO Supplement 7, ETS, Pisa 2018, pp. 41-49.

Bertolaso, M., Philosophy of Cancer. A Dynamic and Relational View, Springer, Dordrecht 2016.

Bertolaso, M., "Be-Com-Ing: Cuestiones emergentes en el estudio de la unidad dinámica del organismo", in: H. Velázquez, L. Contreras, F. Mendoza, (eds), La unidad del viviente desde un enfoque interdisciplinario: del origen de la vida a la generación de hábitos. Una aproximación desde la filosofía y las ciencias biológicas, Comares, Granada 2018, pp. 59-76.

Chandelier, J. and Robert, A., "Nature humaine et complexion du corps chez les médecins italiens de la fin du Moyen Âge", Revue de synthèse, IV, 134, 2013, pp. 473-510.

Emerton, N., The Scientific Reinterpretation of Form, Cornell University Press, Ithaca-London 1984, pp. 76-105.

Green, S. (ed), Philosophy of Systems Biology: Perspectives from Scientists and Philosophers, Springer, Dordrecht 2017, esp. pp. 1-23.

Groebner, V., "complexio/Complexion. Categorizing individual nature. 1250-1600", in: L. Daston and F. Vidal (eds), The Moral Authority of Nature, The University of Chicago Press, Chicago 2004, pp. 361-383.

Huneman, P., "Robustness: the explanatory picture", in M. Bertolaso, S. Caianiello, E. Serrelli (eds), Biological Robustness: Emerging Perspectives from within the Life Sciences , Springer, Dordrecht, forthcoming 2018, chapter 5, pp. TBA.

Jacquart, D., La médecine médiévale dans le cadre parisien, XIVe-Xve siècle, Fayard, Paris 1998, esp. pp. 391-402.

Jacquart, D. "De crasis à complexio: note sur le vocabulaire du temperament en latin médiéval", in: G. Sabbah (ed), Textes médicaux latins antiques, Publications de l'Université de Saint-Etienne, Saint-Etienne 1984, pp. 71-76.

Juarrero, A., Dynamics in Action. Intentional Behaviour as a Complex System, MIT Press, Cambridge 1999.

Kaye, J., A History of Balance, 1250-1375. The Emergence of a New Model of Equilibrium and its Impact on Thought, Cambridge University Press, Cambridge 2014, esp. pp. 128-240.

Krohs, U. and Callebaut, W. "Data without models merging with models without data", in: F. Boogerd, F.J. Bruggeman, J.H.S. Hofmeyr, H.V. Westerhoff (eds), Systems Biology: Philosophical Foundations, Elsevier, Amsterdam 2007, pp. 181-213.

Maier, A., An der Grenze von Scholastik und Naturwissenschaft, Edizioni di Storia e Letteratura, Roma 1952.

Maier, A., On the Threshold of Exact Science: Selected Writings of Anneliese Maier on Late Medieval Natural Philosophy, University of Pennsylvania, Philadelphia 1982, esp. pp. 124-142.

McVaugh, M. R., Arnaldi de Villanova Opera Medica Omnia, II, Aphorismi de gradibus, Universidat de Barcelona, Barcelona-Granada 1975, esp. pp. 9-10 and pp. 20-22.

Moreau, E., Elements, Atoms & Physiology: The Medical Context of Matter Theories (1567-1634), Université Libre de Bruxelles & Radboud University, unpublished PhD thesis, 2018.

Moreau, E., "Elements, Mixture and Temperament: The Body's Composition in Renaissance Physiology", in: C. Beneduce and D. Vincenti (eds), Oeconomia corporis. The Body's Normal and Pathological Constitution at the Intersection of Philosophy and Medicine, MEFISTO Supplement 7, ETS, Pisa 2018, pp. 51-58.

Murdoch, J. E., “The Medieval and Renaissance Tradition of minima naturalia”, in C. Lüthy et al. (eds), Late Medieval and Early Modern Corpuscular Matter Theories, Brill, Leiden 2001, pp. 91-132.
Nikolov, S., Yankulova, E., Wolkenhauer, O., Petrov, V., “Principal difference between stability and structural stability (robustness) as used in systems biology”, Nonlinear Dynamics, Psychology, and Life Science, 11(4), 2007, pp. 413-33.
O’Malley, M. A. and Dupré, J., “Fundamental Issues in Systems Biology”, BioEssays, 27 (12), 2005, pp. 1270-1276.
Ottosson, P.-G., Scholastic Medicine and Philosophy. A Study of Commentaries on Galen’s Tegni (ca. 1300-1400), Bibliopolis, Napoli 1984, esp. pp. 127-194.
Siraisi, N. G., Medieval and Early Renaissance Medicine, The University of Chicago Press, Chicago 1990, esp. pp. 101-104.
Thorndike, L., “De Complexionibus”, Isis, 49, 1958, pp. 398-408.
Woods, R. and Weisberg, M., “Interpreting Aristotle on Mixture: Problems about Elemental Composition from Philoponus to Cooper”, Studies in History and Philosophy of Science, XXXV, 2004, pp. 681-706.
Zanier, G., “Il problema della complexio e la nozione del vivente in Marsilio di Inghen”, Esercizi Filosofici/Testi, VI (2002), pp. 69-77.

## SEVENTH SESSION: Foundations of Logic and Mathematics

**Ludovica Conti** (University of Pavia)
*Russell's Paradox ways out*

1. Russell's Paradox

This paper concerns the open question about the explanation (and related solution) of Russell's paradox in fregean contexts. I briefly examine two traditional positions and propose a third one.

There are many different strategies to avoid the same contradiction but, by "solution" of the paradox, I mean the revision of the feature which is considered the proper flaw from which the contradiction is generated. So, to give an "explanation" of a paradox is to offer a (more or less explicit) analysis that is preliminary to the solution, namely it is to identify the non-obvious flaw that makes a specific premise only erroneously acceptable[3].

The formal system of *Grundgesetze der Arithmetik* is, except for minor differences, second-order logic (with an impredicative comprehension's axioms schema – CA), augmented with a single non-logical axiom, Basic Law V (BLV). The principles explicitly involved in the derivation of Russell's paradox are Basic Law V and an impredicative instance of the comprehension's axioms schema; however, the derivation also presupposes, implicitly but necessarily, a logical theorem about the extensions (ET: $\forall X \exists x\big(x = est(X)\big)$[4].

In the debate about this paradox, there are traditionally two main and incompatible proposals: the "cantorian" explanation and the "predicativist" one – with related solutions[5].

The main thesis of the cantorian explanation consists in identifying, as necessary and problematic condition of the paradox, the conditional axiom (BLVb) $\forall X \forall Y\big(est(X) = est(Y) \rightarrow \forall x(Xx \leftrightarrow Yx)\big)$, which is the left-to-right conditional contained in Basic Law V: from this principle follows, by Existential Generalization, $\exists \iota \forall X \forall Y\big(\iota(Y) \rightarrow \forall x(Xx \leftrightarrow Yx)\big)$, namely a proposition which affirms the existence of an injective function from the second-order domain to the first-order domain. So, in the cantorian perspective, the specific fault of BLV consists in (syntactic version) the violation of Cantor's Theorem and (semantic version) in the assumption of an injection between two domains of different cardinalities.

This explanation can be rejected both from a syntactic and from a semantic point of view. Syntactically, the same contradiction can also be derived (Paseau 2015) from a weaker principle (Definable-BLV: $\forall X \forall Y\Big(\forall x(Xx \leftrightarrow \phi x) \rightarrow \big(ext(X) = ext(Y) \leftrightarrow \forall x(Xx \leftrightarrow Yx)\big)\Big)$), that is consistent with Cantor's theorem; then (the standard version of) BLVb is not a necessary condition of the paradox. Semantically, cardinality's difference of the domains concerns only standard models of second-order logic, while BLV is not satisfiable even in secondary models; then, the considerations about cardinality do not explain the paradox.

Correspondingly, the proposed cantorian solutions – consisting in modifying the injectivity of the extensionality function by restrictions of BLVb (Frege 1903, Paseau 2015) – are provably not able to avoid the contradiction (Quine 1955).

On the other side, the main thesis of the predicativist explanation consists in identifying, as necessary and problematic condition of the paradox, the impredicative comprehension's axioms schema $\exists X \forall x\big(Xx \leftrightarrow \exists \phi(x)\big)$: an instance of this schema allows to specify Russell's concept, $\exists X \forall x\big(Xx \leftrightarrow \exists Y(x = ext(Y) \wedge \neg Yx)\big)$, namely a concept which violates the reflexivity of the logical relation of co-extensionality (because

---

[3] Cfr. Sainsbury 1995: a paradox is an argument in which "an apparently unacceptable conclusion (is) derived by apparently acceptable reasoning from apparently acceptable premises".

[4] Russell Paradox.

| | | | |
|---|---|---|---|
| 1. | $\forall X \forall Y\big(est(X) = est(Y) \leftrightarrow \forall x(Xx \leftrightarrow Yx)\big)$ | | (BLV) |
| 2. | $\exists X \forall x\big(Xx \leftrightarrow \exists Y(x = est(Y) \wedge \neg Yx)\big)$ | Call this concept R. | (CA) |
| 3. | $\forall X \exists x\big(x = est(X)\big)$ | | (ET) |
| 4. | $\exists x\big(x = est(R)\big)$ | | (2, 3) |
| 5. | $\neg Rest(R) \rightarrow Rest(R)$ | | (2, 4) |
| 6. | $Rest(R) \rightarrow \exists Y \backslash noteqR\big(est(R) = est(Y) \wedge \neg Yest(R)\big)$ | | (2, 4) |
| 7. | $\neg Rest(R)$ | | (1, 6) |
| 8. | $Rest(R) \leftrightarrow \neg Rest(R)$ | | (5, 7) |

[5] Cfr. Uzquiano forthcoming.

it is predicable of its extension if and only if it isn't[6]). So, the specific fault of CA consists in (syntactic version) the impredicative structure of the comprehension's schema and (as different sematic versions) in the consequent indefinite extensibility of the second-order domain (Dummett 1991) or in the alleged vicious circularity of the impredicative quantifier's interpretation (Russell 1903).

This explanation seems to be correct but incomplete: the Russellian impredicative instance of comprehension's axioms schema is really a necessary condition of the paradox but second-order logic (with the same impredicative comprehension's axiom, also interpreted in standard models with the same second-order domain) is consistent and many other impredicative principles (with the same quantifier's interpretation) do not generate paradoxes.

Correspondingly, the predicativist solutions – consisting in modifying the second-order domain by predicative restrictions of CA (Heck 1996, Wehemeier 1999, Ferreira-Wehemeier 2002) – only partially work: these solutions are sufficient to avoid the contradiction but weaken the original Frege's theory[7].

Then, we are in front of an apparent aporia: while both BLV and CA are necessary conditions of Russell's contradiction, both the proposals seem to be unable to deeply explain and really solve the paradox. The cantorian explanation, identifying the mistake of Frege's system in the injectivity of the extensionality function – then in BLVb – selects an irrelevant feature, namely a condition which underdetermines the contradiction; the predicativist explanation, identifying the mistake in the impredicativity of the second-order domain's specification – then in CA – selects a too general feature, namely a condition which only indirectly takes part to the contradiction.

## 2. Extensionalist Explanation

There is a third, less known, way to explain the paradox – which we'll call the "Extensionalist" explanation.

The main thesis of the Extensionalist explanation consists in identifying, as immediate condition of the paradox, a logical theorem (ET: $\forall X \exists x(x = extX)$) involved in the derivation and, as its root, the theory of quantification and identity involved in the classical axiomatization of second-order logic, from which this theorem follows[8]. ET asserts that the function denoted by symbol $ext$ is defined on the whole second-order domain; so, by this theorem, we obtain, from the existence of Russell's concept, the existence of Russell's extension, namely the object which leads to the contradiction. In the extensionalist perspective, the specific fault of classical logic – whose ET is a consequence – consists in (syntactic version) the unrestricted formulation of quantifier's and identity's rules and (semantic version) in the assumption that every singular term must be denoting – then, that every function is defined on the whole considered domain.

Then, from a syntactic point of view, this proposal identifies a mistake in the interaction of the non-logical axiom BLV with classical second-order logic; this means that the problem does not concern the mere injectivity (or, eventually, functionality) of the correlation directly described by BLV but what the logic itself says about it, namely what its domain contains. Reasoning in a classical framework – where ET is a mere consequence of more basic logical laws – the domain of the correlation coincides with the domain of second-order logic and the only axiom which seems to be liable of the contradiction is CA; this consideration leads predicativist explanation to misunderstand the problem which concerns the definition of function's domain as a problem of the second order domain – drawing semantic conclusions that are not able to explain the contradiction. Although in the classical framework these two domains coincide, the extensionalist explanation (unlike the predicativist one) identifies the root of the contradiction in the only function's domain.

So, this third explanation can be classified, as the predicativist one, in the group of Frege's readings[9] that ascribes Russell's paradox to the too generous intersection between *logicist* aim of deriving arithmetic from

---

[6] Cfr. note 1, line 8.

[7] Predicative second-order fragments of Frege's Grundgesetze are equi-interpretable with Rubinson arithmetic, while Frege's original program ask to recover full second-order Peano arithmetic.

[8] Extensions' theorem.

1. $\forall x(x = x)$ (SOL$^{=}$)
2. $extX = extX$ (SOL$^{=}$: $\forall x\phi \rightarrow \phi t/x$)
3. $\exists x(x = extX)$ (IE)
4. $\forall X \exists x(x = extX)$ (IU)

[9] Cfr. Cocchiarella 1992; Antonelli – May 2005. With different solutions, also Boccumi 2010 and Ferreira forthcoming seem to presuppose this reading of Russell's paradox.

logical axioms and the *extensionalist* aim of reconstructing arithmetic as a theory of the extensions. However, as its own specific feature, this third explanation locate the interaction between second order logic and extension's theory not in CA but in a consequence of quantification's and identity's axioms.

## 3. Free Solutions

Just because this interaction occurs in the paradox by a theorem (ET) and not directly by an axiom (BLV or CA), the correspondent solution consists in a deeper alteration which involves all the principles from which the theorem follows. This solution turns out to be, first, the substitution of classical second order logic with negative free logic. This only change is enough to prevent the standard version of Russell's paradox, namely the contradictory derivation of the Russell's concept (by CA) and of Russell's extension (by ET).

However, in every "free" abstractionist system - theory including axioms of second-order free logic augmented with some abstraction principle – we can derive the existence of specific abstracted objects denoted by complex singular term (obtained by abstraction operator) by the conditional which constitutes the right-to-left reading of the abstraction principle: their existence is a consequence of the reflexivity of the abstraction's relation (not necessary equivalence relation[10]) between concepts[11].

In this framework, the standard version of BLV allows us to avoid the existence of Russell's extension - and the contradiction – only accepting that Russell's concept - just as paradox seems to show – is not reflexively coextensional with itself. Nevertheless, this result violates logical feature of the co-extensionality relation and leads to an undesirable and uninterpretable situation[12]. So, the full solution of the paradox - which follows from the extensionalist explanation - presupposes, other than the adoption of second-order free logic, a correspondent weakening of BLVa.

We can briefly compare three free fregean theories which share the logical axioms and distinguish one other by the different restrictions admitted on right sight of BLV. All these free fregean theories[13] involve, as the logical core of the theory (FL) the axioms of classical second-order logic without identity (SOL) for "un-restricted" quantification, the axioms of non-inclusive negative free logic with identity ($NFL^{=}$) for "restricted" quantification and identity, the comprehension's axioms schema (CA) and, as the only inferential rule *modus ponens* (MP).

---

[10] Cfr. Payne 2011.

[11] Derivation of abstract's existence:

1. $X \sim Y \rightarrow (X) = (Y)$ (AP r-1)
2. $X \sim X$ (refl. $\sim$)
3. $(X) = (X)$ (1, 2 MP)
4. $\exists x(x = X)$ (IE)

[12] If Russell's concept is reflexively co-extensional with itself we derive that its extension exists and, from that result, the contradiction; but, if we refute the existence of Russell's extension, Russell's concept turns out to be reflexively co-extensional with itself and then, again, able to introduce its extension.

[13] The language $L_F$ is a second-order language which involves, as primitive symbols:

a) Logical constants: $\neg, \rightarrow, =$ ;
b) A universal "unrestricted" quantifier FOL $\bigvee$ , which applies to first-order variables – by which is defined a particular "unrestricted" quantifier FOL $\bigwedge$ : $\bigwedge x =_d \neg \bigvee x$ ;
c) A universal quantifier SOL $\forall$, which applies to second-order variables – by which is defined a particular quantifier SOL: $\exists: \exists X =_d \neg \forall X$ ;
d) An infinite list of individual constants: a, b, c, …;
e) An infinite list of n-ary predicate constants: $A^n, B^n, C^n, \ldots$ ;
f) A functional symbol: ext, which applies to monadic predicative variables and constants;
g) An infinite list of variables FOL: x, y, z, …;
h) An infinite list of variables SOL: $X^n, Y^n, Z^n, \ldots$ ;

With this vocabulary, we can also define:

i) The predicative monadic constant $E!: E!\, a =_d \exists x(x = a)$ ;
j) A universal "restricted" quantifier FOL $\forall: \forall x Ax =_d \bigvee x(E!\, x \rightarrow Ax)$ ;
k) A particular "restricted" quantifier FOL $\exists: \exists x Ax =_d \bigwedge x(E!\, x \wedge Ax)$ .

FL:

C1) $\alpha \to (\beta \to \alpha)$;

C2) $(\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$;

C3) $(\neg\alpha \to \neg\beta) \to (\beta \to \alpha)$;

C4) $\alpha \to \bigvee v\alpha$, if v is not free in $\alpha$;

C5) $\alpha \to \forall v\alpha$, if v is not free in $\alpha$;

C6) $\bigvee v(\alpha \to \beta) \to (\bigvee v\alpha \to \bigvee v\beta)$;

C7) $\forall v(\alpha \to \beta) \to (\forall v\alpha \to \forall v\beta)$;

C8) $\bigvee v\alpha$, if $\alpha$ is an axiom;

C9) $\forall v\alpha$, if $\alpha$ is an axiom;

C10) $\bigvee v\alpha \to \alpha(t/v))$ ;

N1) $\forall v\alpha \to (E!t \to \alpha(t/v))$;

N2) $\exists v E!v$;

N3) $s = t \to (\alpha \to \alpha(t//s))$;

N4) $\forall v(v = v)$;

N5) $\Pi\tau_1, ..., \tau_n \to E!\tau_i$ (with $1 \leq i \leq n$);

CA) $\exists X \bigvee x(Xx \leftrightarrow \alpha)$;

From this axioms we can also derive, as theorems:

T1) $\forall x E!x$;

T2) $t = t \leftrightarrow E!t$;

T3) $(\neg E!s \wedge \neg E!t) \to (\alpha \to \alpha(t//s))$.

The weaker one of the three theories (E-FL) consists in FL augmented with, as the only non-logical axioms, two conditionals which jointly represent a weaker version of BLV (all instances of E-BLVa and E-BLVb):

E-BLVa: $\forall X \forall Y\big(\bigvee x(Xx \leftrightarrow Yx) \wedge E!\, ext(X) \vee E!\, ext(Y) \to ext(X) = ext(Y)\big)$ ;

E-BLVb: $\forall X \forall Y(ext(X) = ext(Y) \to \bigvee x(Xx \leftrightarrow Yx))$ .

This system is strong enough to define inductively every natural numbers, as complex singular terms (obtained by the application of extensionality operator ext to predicative constants introduced by CA) interpreted by the correspondent sets in Von Neumann's hierarchy[14].

For example, by an instance of CA we introduce the concept of $\lambda x.\, x \neq x$[15] and, with this, we define the number zero:

**Definition 3.1.** $0 = ext(\lambda x.\, x \neq x)$

Inductively we can also introduce, by instances of AC, every concept related to each natural number and define them as the singular terms obtained by the application of abstraction operator to each of these concepts.

**Definition 3.2.** $1 = ext(\lambda x.\, x = 0)$

[14] Cfr. Boccumi 2010.

[15] We use λ – *operator* to express the concept correspondent to a certain formula.

**Definition 3.3.** $2 = ext(\lambda x.x = 1)$

**Definition 3.4.** $3 = ext(\lambda x.x = 2)$

Following the set definition of natural numbers, we ca define also the concepts of successor, inductive concept and natural number.

**Definition 3.5.** $Sn = ext(\lambda x.x = n)$

**Definition 3.6.** $I(X) = X0 \wedge \forall z\big(Xz \rightarrow X\epsilon(\lambda x.x = z)\big)$

**Definition 3.7.** $\mathbb{N}x = \forall X(I(X) \wedge Xx)$

However, the grammatical correctness of these definitions, in a free axiomatization with E-BLVa, does not allow us to consider their *definienda* as denoting.

Nevertheless, we can derive four of five second-order Peano Axioms.

**Theorem 3.1.** $\mathbb{N}0$[16].

**Theorem 3.2.** $\forall x(Sx \neq 0)$[17].

**Theorem 3.3.** $\forall y\forall z(\epsilon(\lambda x.x = y) = \epsilon(\lambda x.x = z) \rightarrow y = z)$[18].

**Theorem 3.4.** $\forall X\left(X0 \wedge \forall y\big(Xy \rightarrow X\epsilon(\lambda x.x = y)\big) \rightarrow \forall x(Xx)\right)$[19].

However, in this theory (E-FL) it is not possible to derive, as theorem, the Peano's Axiom about successor ($\forall z\exists y\big(y = ext(\lambda x.x = z)\big)$), because this theorem involves an existential quantification over individual variables; on the contrary, all the other theorems concern numbers only as (hypothetical) entities, independently from their existence.

The second one of the three theories (P-BLV) consists in FL augmented with, as the only non-logical axioms, two conditionals which jointly represent a different weaker version of BLV (all instances of P-BLVa and P-BLVb):

---

[16] [Proof] Number 0, as empty extension (extension of an empty concept) satisfies the application's conditions of number concept (def. 3.7).

[17] [Proof]:

| | |
|---|---|
| 1. $\exists y(Sy = 0)$ | (A) |
| 2. $\epsilon(\lambda x.x = y) = \epsilon(\lambda x.x \neq x)$ | (def. 0, def. S) |
| 3. $\forall z([\lambda x.x = y](z) \leftrightarrow \lambda x.x \neq x](z))$, | (E-BLVb) |
| 4. $\forall z(z = y \leftrightarrow z \neq z)$. | $(\lambda - conversione)$ |
| 5. $\forall z E!z(z = y \leftrightarrow z \neq z)$ | (T1) |
| 6. $y = y \leftrightarrow y \neq y \perp$ | (N1) |
| 7. $\neg\exists y(Sy = 0)$ | (1,6) |
| 8. $\forall y\neg(Sy = 0)$ | (7) |
| 9. $\forall y(Sy \neq 0)$ | (8, def. $\neq$) |

[18] [Proof]:

| | |
|---|---|
| 1. $\epsilon(\lambda x.x = y) = \epsilon(\lambda x.x = z)$ | (A) |
| 2. $\forall x(\lambda x.x = y)(x) \leftrightarrow (\lambda x.x = z)(x)$ | (E-BLVb) |
| 3. $(\lambda x.x = y)(a) \leftrightarrow (\lambda x.x = z)(a)$ | (N1, E!a) |
| 4. $a = y \leftrightarrow a = z$ | $(\lambda - conversione)$ |
| 5. $y = z$ | (N3) |

[19] [Proof]: The theorem follows from the definition of number's concept $\mathbb{N}$.

P-BLVa) $\forall X \forall Y \left( \left( \bigvee x(Xx \leftrightarrow Yx) \wedge \left( \phi(X) \vee \phi(Y) \right) \right) \rightarrow ext(X) = ext(Y) \right)$.

where $\phi$ means "predicative", namely specified by a predicative instance of CA;

P-BLVb) $\forall X \forall Y (ext(X) = ext(Y) \rightarrow \bigvee x(Xx \leftrightarrow Yx))$ .

The difference between a predicative restriction of CA and a predicative restriction of BLV, into a framework of free logic, is the "object" under the restriction: in the first case, the restriction selects the second order domain of the theory - so that Russell's concept does not exists; in the second case, the restriction concerns only the extensions - so that exists Russell's concept but not its extension. Obviously. this strategy presupposes the restricted (free) axiomatisation and so also excludes ET - and the standard derivation of the paradox: instead, in a classical framework - where ET is a logic theorem – the same predicative restriction of BLV is not sufficient to avoid the existence of Russell's extension and then the contradiction.

This theory allows us to complete the previous reconstruction of Peano arithmetic, affirming the existence of every natural number. Every natural number, in fact, is definable as the extension of a predicative concept, so it can be introduced in a correspondent instance of P-BLVa from which we can derive its existence:

**Theorem 3.5.** $E!\, 0$[20].

Since we can prove the existence of each number, we can also derive the missing Peano's axiom:

**Theorem 3.6.** $\forall z \exists y \left( y = ext(\lambda x.\, x = z) \right)$[21].

However, this system does not allow us to pursue the original fregean strategy because we can now define numbers, in a fregean way, as extensions of second level concepts but again, by this predicative version of BLV, we cannot prove their existence: if "the number of Xs" is defined - as in *Grundgesetze* - as an extension of the impredicative concept "to be an extension of a concept equinumerous to X" (free logic being neutral about existential assumptions) P-BLVa is not sufficient to prove its existence.

The last one of the three theories (B-FL) consists in FL augmented with, as the only non-logical axioms, two conditionals which jointly represent another version of BLV (all instances of B-BLVa and B-BLVb) weaker then fregean one but stronger than both E-BLV and P-BLV:

P-BLVa) $\forall X \forall Y \left( \left( \bigvee x(Xx \leftrightarrow Yx) \wedge \left( \phi(X) \vee \phi(Y) \right) \right) \rightarrow ext(X) = ext(Y) \right)$.

where $\phi$ means "small", following Boolos definition[22], namely a concept X such that $\forall x \left( Xx \rightarrow [\lambda x.\, x = x](x) \right)$ but not *viceversa*;

P-BLVb) $\forall X \forall Y (ext(X) = ext(Y) \rightarrow \bigvee x(Xx \leftrightarrow Yx))$ .

---

[20] [Proof]:

1. $\bigvee x([\lambda x.x \neq x] \leftrightarrow [\lambda x.x \neq x]) \rightarrow ext([\lambda x.x \neq x]) = ext([\lambda x.x \neq x])$ (P-BLVa)
2. $\bigvee x([\lambda x.x \neq x] \leftrightarrow [\lambda x.x \neq x])$ (refl. co-ext.)
3. $ext([\lambda x.x \neq x]) = ext([\lambda x.x \neq x])$ (1, 2, MP)
4. $E!ext([\lambda x.x \neq x])$ (T2)
6. E! 0 (def. 3.1)

[21] [Proof]:

1. $y = ext(\lambda x.x = z)$ (A)
2. $E!ext(\lambda x.x = z)$ (T 3.5 cfr. note 23)
3. $\exists y(y = ext(\lambda x.x = z)$ (IE)
4. $\forall z \exists y(y = ext(\lambda x.x = z))$ (IU)

[22] Cfr. Boolos 1987.

This system allows us to pursue not only an insiemistic reconstruction of Peano Arithmetic, but also a fregean one[23]: "small" restriction on PBLVa allows us to define numbers as extensions of the first level concept "to be an extension of a concept equinumerous to X" and prove a new result about the relation between extension, number and equinumerosity, namely $\forall X \forall Y ext(Y) \in \#(X) \leftrightarrow \phi(X) \wedge \phi(Y) \wedge Y \approx X$[24]. Given this result, we can derive Hume's Principle and, by full expressivity of impredicative CA, define the concepts of successor, ancestral, weak ancestral, and natural number; then, we can derive - also in a fregean way - second order Peano Arithmetic.


[1] Antonelli, A and May R. (2005). Frege's Other Program, Notre Dame Journal of Formal Logic, Vol. 46, 1, 1-17.
[2] Boccuni, F. (2010), Plural Grundgesetze, Studia Logica, 96, 2, 3015-330.
[3] Boolos, G. (1987). Saving Frege from the Contradiction, Proceedings of Aristotelian Society, 87, 137-151.
[4] Cocchiarella, N. B. (1992). Cantor's power-set Theorem versus Frege's double correlation Thesis, History and Philosophy of logic, 13, 179-201.
[5] Dummett, M. (1991). Frege, Philosophy of Mathematics, Oxford University Press, Oxford.
[6] Ferreira, F. (forthcoming), Zigzag and Fregean arithmetic.
[7] Ferreira, F. and K. F. Wehmeier (2002). On the Consistency of the Fragment of Frege's Grundgesetze. Journal of Philosophical Logic.
[8] Frege, G. (1903) Grundgesetze der Arithmetik, II, Verlag Hermann Pohle.
[9] Heck, R. (1996). The Consistency of Predicative Fragments CA11 of Frege's Grundgesetze der Arithmetik. History and Philosophy of Logic 17, 209–220.
[10] Payne, J. (2013). Abstraction relations need not be reflexive, Thought, 2, 137 - 147.
[11] Paseau, A. C. (2015). Did Frege commit a Cardinal Sin?, Analysis 75 (3), 379–386.
[12] Quine W. V. (1955), On Frege's way out, Mind, 64, 145-159.
[13] Russell, B. (1903). On Some Difficulties in the Theory of Transfinite Numbers and Order Types, in Russell B. (1973). Essays in Analysys, New York.
[14] Sainsbury, R. M. (1995). Paradoxes, Cambridge University Press, Cambridge.
[15] Uzquiano, G. (forthcoming). Impredicativity and Paradox.
[16] Wehmeier, K. F. (1999). Consistent Fragments of Grundgesetze and the Existence of Non-Logical Objects. Synthese 121(3), 309–328.
[17] Zalta, E. (2013). Stanford Encyclopedia of Philosophy, Stanford.



**Claudio Ternullo** (Kurt Gödel Research Center for Mathematical Logic, University of Vienna) **Luca Zanetti** (IUSS, Pavia)


*From Bolzano to Frege: A Cantorian Path*

Frege is credited with formulating an original and distinctive account of concepts (in particular, of the 'number' concept, as we find it in [Frege, 2007]), whereby the latter are seen as mind-independent, objective constructs. However, Frege's scholarship has seldom paid attention to the sources of such a conception, and has also mostly disregarded the importance, for its development, of the contributions of Frege's predecessors and colleagues.

A notable exception is maybe represented by Dummett, who, in his [Dummett, 1991, vii], notices that Frege's account of concepts should be viewed as fundamentally original, but also as potentially indebted, at least in part, to Bolzano's account. However, as Dummett remarks, "there is no evidence whatever that Frege ever read Bolzano".

In ([Tait, 2002]), Tait rebukes Dummett's point of view, by showing that Frege's conception of 'number' is, rather, dependent upon Cantor's and Dedekind's own accounts, and, moreover, that the crucial contributions of both authors to Frege's thought through the theory of infinite sets have not been fully, if at all, acknowledged. Tait also explains that Frege fully thrived on the set-theoretic perspective developed by

[23] About the reconstruction of Frege's definition cfr. Frege 1893, par.34; about the differences between this definition and the original one in *Grundlagen*, cfr. Zalta 2013.

[24] This result implies a more general result about extensions, which here holds in a restricted version: $\forall X \forall x (x \in extX \leftrightarrow \phi(F) \wedge Fx)$. Cfr. Zalta 2013.

Cantor, but that he ignored Cantor's warning about the fact that a concept-based account of numbers might potentially lead to inconsistencies.
While we agree with Rossberg and Ebert ([Ebert and Rossberg, 2009]) that Tait's claim that Cantor had foreboded the inconsistency of Frege's Law V is unwarranted, we believe, however, that Tait's work has, at least, the unquestionable merit of underlining the importance of Cantor's contribution to the development of Frege's doctrines.
Therefore, the main goal of the paper is to look at Frege's account of concepts, mostly through the lens of Cantor's conception. Already Ernst Zermelo, in a footnote to the re-print of [Cantor, 1885] in [Cantor, 1932], had observed that, although Cantor and Frege had very often misunderstood each other, their conceptions were significantly alike. In the paper, we aim to further substantiate and elaborate upon Zermelo's suggestion in the way sketched below.
To begin with, in his [Cantor, 1883], Cantor formulates a theory of concepts which fits in very well with Frege's conception in many respects. Moreover, set-theoretic reductionism, that is, the claim that mathematical objects are sets, is clearly at work also in Frege's account of numbers. Finally, both Cantor and Frege subscribed to mathematical Platonism, a view which has subtle and significant bearings on their account of concepts.
The structure of the paper is as follows. In section 1, we examine the relationships between Cantor's and Bolzano's conception. In section 2 we discuss Dummett's and Tait's views, and assess their relevance. In section 3, we exhibit relevant connections among Cantor, Bolzano and Frege. In section 4, we introduce Zermelo's assessment of Cantor's contribution to the development of Frege's ideas, and in section 5 we discuss the analogies between Cantor's and Frege's accounts of concepts. Finally, in section 6, we point to further striking analogies between Frege's and Cantor's treatment of numbers, as based on the use of a set-theoretic perspective.

Cantor, G. (1883). Grundlagen einer allgemeinen Mannigfaltigkeitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen.B. G. Teubner, Lepzig.
Cantor, G. (1885). Review of Frege's Die Grundlagen der Arithmetik. Deutsche Litteraturzeitung, 6(20).
Cantor, G. (1932). Gesammelte Abhandlungen mathematischen und philosophischen Inhalts. Springer, Berlin.
Dummett, M. (1991). Frege. Philosophy of Mathematics. Harvard University Press, Harvard.
Ebert, P. and Rossberg, M. (2009). Cantor on Frege's Foundations of Arithmetic: Cantor's 1885 Review of frege's Grundlagen der Arithmetik. History and Philosophy of Logic, 30:341–348.
Frege, G. (2007). The Foundations of Arithmetic. Pearson-Longman, New York.
Tait, W. W. (2002). Frege versus Cantor and Dedekind on the Concept of Number. In Jacquette, D., editor, Philosophy of Mathematics. An Anthology, pages 40–64. Blackwell Publishers.

**Matteo Zicchetti** (University of Bristol)
*Truth-theories, Cognitive Projects and Trustworthiness*

What should truth-theories be like? Hannes Leitgeb answered this question providing in [6] a list of adequacy criteria for truth-theories. He showed that three possible equally good options are available and concluded with a challenge: to provide a way to choose the best theory. The aim of my paper is to propose a new requirement for truth-theories, which I will call *trust- worthiness*, and provide an answer to Leitgeb's challenge. The paper has the following structure: In section 1. I will present Leitgeb's results. Section 2. is devoted to the explanation of the new requirement: I will present truth- theories as *cognitive projects* [9] and introduce and explain the *trustworthiness* requirement as an epistemic adequacy condition on truth-theories. In section 3. I will make the notion of trustworthiness formally precise. In section 4. I will show that none of the three options singled out by Leitgeb is trustworthy. After that I will consider the truth-theory *PUTB* [4], show that it is trustworthy and that scores better than the rival truth-theories. At the end of this section I will shortly address the philosophical significance of this result.

1. Leitgeb's challenge

Leitgeb [6] answers the question about what truth-theories should be like by proposing eight adequacy criteria:

Truth should be expressed by a predicate (and a theory of syntax should be available)
If a theory of truth is added to a mathematical or empirical theory, it should prove the latter true
The truth predicate should be type-free
T-biconditionals should be derivable unrestrictedly
Truth should be compositional
The theory should allow for standard models
The outer logic and inner logic should coincide
The outer logic should be classical[25]
Although each requirement is intuitive and desirable, it is known that no axiomatic truth-theory can consistently satisfy all desiderata.

**Theorem 1** (Tarski [7])**.** *If a theory S satisfies (a) + (b) + (c) + (d) + (h) then S is trivial.*
As a response to this, Leitgeb singles out subsets of the desiderata and shows that we are left with three possible sets that can be consistently instantiated by truth-theories:
1. (a) + (b) + (c) + (e) + (f) + (h)
2. (a) + (b) + (c) + (e) + (g) + (h)
3. (a) + (b) + (c) + (f) + (g) + (h)
Subsets 1-3 are realized respectively by the theories of truth *KF*, *FS* and *VF*.[26] Therefore, we have at least three equally good truth-theories *in classical logic* that satisfy different subsets of equally good requirements.[27] This shows that the desiderata aren't sufficient to choose one theory over the others. Leitgeb is aware of this and formulates his challenge:
"Can we rank our eight postulates in a way that would permit us to impose some additional 'order of acceptance' on the class of our maximal satisfiable sets? [...] Which other norms do exist that govern our understanding of truth?" [6]
Here I will address the second of Leitgeb's questions. I will propose a new requirement to be added to the list, which I will call *trustworthiness* and spell it out as an epistemic adequacy condition. In the next section I will introduce and explain the new desideratum. To do that, I will introduce the notions of *cognitive project*, *presuppositions* and *entitlement of* cognitive projects and de- fine truth-theories as *cognitive projects*. Moreover, in this paper I will take the trustworthiness requirement, at least in its *informal* definition, to be non- negotiable.

## 2. Cognitive projects, presuppositions and trust- worthiness

The epistemological theory of cognitive projects was introduced by Crispin Wright in [8], [9]. He defined a cognitive project in the following way:
"A cognitive project is defined by a pair: a question, and a procedure one might competently execute in order to answer it." [9]

An essential part of cognitive projects are what Wright calls *cornerstone propositions*, the presuppositions of the cognitive project. Wright defines presuppositions as follows:

"P is a *presupposition* of a particular cognitive project if to doubt P (in advance) would rationally commit one to doubting the significance or competence of the project." [9]

In addition, Wright spells out the notion of *entitlement of cognitive project*:

"[A]n entitlement of cognitive project [...] may be proposed to be any presupposition P [...] meeting the following additional two conditions:
We have no sufficient reason to believe that P is untrue
The attempt to justify P would involve further presuppositions in turn of no more secure a prior standing." [9]

[25] For an explanation of each requirement see [6].
[26] I must assume familiarity with these theories, because presenting them would exceed the scope of and space for this paper. For a presentation see Halbach [5] and Cantini [1].
[27] Leitgeb doesn't pose any order of importance on the requirements.

Wright concludes claiming that, if (i) and (ii) are met, then we are rationally entitled to *accept P*, i.e., to *trust that P*.

My proposal is to spell any truth-theory *T* as a cognitive project, by defining *T* as a pair (*X*, *Y*), where *X* and *Y* are respectively sets of questions and procedures. In order to spell out the presuppositions of a cognitive project (*X*, *Y*), we need first to specify the set *X* of questions. Here I will understand
*X* to be the following question:

(F) *What statement formulated in the language of a theory S should one accept if one has accepted the axioms and rules of S?*[28]

In our case we let *S* be *Peano arithemtic (PA)*. Provided *that we accept PA*, we extend it to a truth-theory *T*, aiming at answering (F) *via proofs in T* of new statements in the language of *PA* but (possibly) not provable in *PA* itself. This will be our cognitive project. What are its presuppositions? Here, I will argue that this cognitive project has as presupposition the proposition expressing the soundness of *T* (for short *Sound(T)*), i.e., that everything that *T* proves is true. *Sound(T)* is a presupposition of the truth-theory *T*, because doubting *Sound(T)* in advance would commit us to doubting either the significance or the competence of the project. It is clear that doubting *Sound(T)* in advance would commit us to doubting the competence, i.e. the procedure, of the cognitive project for the following reason; the aim of *T* is to provide and answer
(F) using procedures, which are the axioms and rules of *T*, carrying out proofs. Moreover, answering (F) tells us what we should accept, provided that we al- ready accepted the base theory. But now, if we would doubt *Sound(T)*, i.e. the fact that proofs in *T* are true, we would undermine our use of *T* to answer
(F) in the first place. Moreover, we are entitled to *trust* that *Sound(T)* since
Wright's conditions are met; we assume that (i) is met, i.e., that we don't have independent reasons to disbelieve *Sound(T)*. (ii) is met, because an attempt of justifying *Sound(T)* would involve a regress of theories aiming at justifying *Sound(T) ad infinitum*.[29]

For *T* to be trustworthy means then informally to be coherent with the pre- supposition that *Sound(T)*. More precisely:

**Definition 1** (Schematic Trustworthiness)**.** *A truth-theory* T *is trustworthy iff* T *is coherent with a method θ of making the trust that Sound(T) explicit.*

In order to specify **Definition 1**, we need to: (I) have a formally precise formulation of *θ*; (II) provide a precise interpretation of the phrase "to be coherent with *θ*". In what follows I will deal with (I) and (II), by presenting Feferman's argument for reflection principles. With Feferman I will argue that reflection principles are exactly what we need to make *Sound(T)* explicit.

3. Expressing trust via reflection

Feferman investigated in [2] what we call *reflection principles* and showed how iterating these principles over arithmetic yields stronger mathematical theories. In what sense are reflection principles the correct way of making *Sound(T)* explicit? Reflection principles are perfect for our strategy, as Feferman claims:

"In contrast to an arbitrary procedure for moving from $A_K$ to $A_{K+1}$, a reflection principle provides that the axioms of $A_{K+1}$ shall ex- press a certain trust in the system of axioms $A_K$." [2]

The crucial point here is that Feferman's argument has epistemic nature; he argues that a reflection principle for a theory *T* expresses the *trust* in the axioms of *T*. According to Feferman's claim, the trust in the axioms of *T* is to be understood as the trust that *Sound(T)*. However, there are various formulations of these principles: *local (Rfn), uniform (RFN)* or *global* reflection (*GRP*). Given a theory *T*, these principles are formalized as:

---

[28] This question was formulated by Feferman *Reflecting on Incompleteness*.
[29] Familiarity with Gö del's incompleteness theorems and with the fact that (in very general set- ting) soundness implies consistency, is assumed.

- $Rfn_T\ Prov_T \ulcorner \varphi \urcorner \rightarrow \varphi$,
- $RFN_T\ Prov_T \ulcorner \varphi(\dot{x}) \urcorner \rightarrow \varphi(x)$,
- $GRP_T\ \forall x\,[Sent_T(x) \wedge Prov_T(x) \rightarrow T(x)]$,

Where '$T(x)$' roughly means 'x is true', '$Prov_T(x)$' means 'x is provable in $T$' and '$Sent_T(x)$' means 'x is a sentence of $T$'.[30]
Although it is uncontroversial that these are formalized soundness statements, there is no agreement upon which of these principles is the best or correct formalization of soundness: one could for instance be committed to a math- ematical theory, but not to a truth predicate. In that case, $GRP_T$ seems to be the wrong formulation of soundness. In our case, since we want to make *Sound(T)* for a truth-theory $T$ explicit, which already adopts a truth predicate, the most natural way to go is to adopt $GRP_T$. This specifies $\theta$ and sets (I). What about (II)? The informal notion of coherence of $T$ with $GRP_T$ could be spelled out in various ways: in terms of consistency with $GRP_T$, but also as soundness with $GRP_T$ etc. Here I am going to understand the coherence of $T$ with $GRP_T$ in terms of consistency of the outer and inner logic of $T$ with $GRP_T$. The outer and inner logic of a theory $S$ are respectively the set of all $\phi$, such that $S \in \phi$, and the set of all $\phi$, such that $S \in T$ '$\phi$'. This allows us to have a precise definition of trustworthiness:

**Definition 2** (Trustworthiness)**.** *A truth-theory* T *is trustworthy iff the outer and inner logics of* T *are consistent with the addition of* $GRP_T$ *to* T *(the procedure may be iterated).*

Having a definition of trustworthiness enables us to formulate the new adequacy requirement to be added to Leitgeb's list:
(i) Truth-theories should be trustworthy.
In what follows I will show that none of the options singled out by Leitgeb satisfies the trustworthiness requirement. I will show that the theory of positive truth *PUTB* proposed by Halbach in [4] satisfies the trustworthiness requirement and scores better than the three rival truth-theories. At the end of the paper, I will shortly address the question about the philosophical significance of this result.

**Results**
We can now look at the three truth-theories singled out by Leitgeb and ask, whether they are trustworthy or not. Let's look at subset 1. and extend it with
to 1*:
1* (a) + (b) + (c) + (e) + (f) + (h) + (i)

However, *KF* cannot consistently satisfy the (i) and hence 1*:

**Theorem 2** (Fischer [3])**.** *KF is internally inconsistent with* $GRP_{KF}$.

We can now extend subset 2 to 2*:
2* (a) + (b) + (c) + (e) + (g) + (h) + (i)
However, *FS* cannot consistently realize 2*:

**Theorem 3** (Halbach [5])**.** *FS is inconsistent with* $GRP_{FS}$.

Finally, we extend 3 to 3*:
3* (a) + (b) + (c) + (f) + (g) + (h) + (i)
However, *VF* cannot consistently realize 3*:

**Theorem 4.** *VF is inconsistent with* $GRP_{VF}$.[31]

[30] For reasons of space here I cannot specify explain these notions any further.
[31] By Montague's theorem.

From these results we can draw the following conclusion:

**Corollary 1.** *KF, FS and VF aren't trustworthy.*
The philosophical explanation of trustworthiness and the understanding of truth-theories as cognitive projects together with this result has crucial con- sequences; if we understand truth-theories as cognitive projects aiming at answering (F), and if we take trustworthiness to be non-negotiable, then *KF*, *FS* and *VF* are not good enough for the task. However, hope isn't lost. In what follows I will present a truth-theory that satisfies the trustworthiness requirement and scores better that the previous options.

**Trustworthiness of positive truth**
The truth-theory *PUTB*, proposed by Halbach in [4], stands for *Positive Uni- form Tarski-biconditionals* and has axioms of the form:

- $T\ulcorner\varphi(\dot{x})\urcorner \leftrightarrow \varphi(x)$

where the formula $\phi(x)$ is *T-positive*. A formula $\phi(x)$ is T-positive iff the truth predicate occurs in $\phi(x)$ under the scope of an even number of negation symbols. *PUTB* satisfies:

$4^{-}$. (a) + (b) + (c) + (f)
However, Halbach proved in [5] that the inner and outer logics of *PUTB* can coincide, adding the two rules:

$$NEC\ \frac{\varphi}{T\ulcorner\varphi\urcorner} \qquad CONEC\ \frac{T\ulcorner\varphi\urcorner}{\varphi}$$

Moreover, $PUTB^{+} =_{Df} PUTB + NEC + CONEC$ is consistent.
**Theorem 5** (Halbach [5])**.** *There are models* $\mathfrak{N} =_{Df} (\mathfrak{N}, \Gamma(S))$ *(of PUTB) as defined in [5], such that* $\mathfrak{N} \models PUTB^{+}$.

This means that we can extend $4^{-}$ to 4:
4. (a) + (b) + (c) + (f) + (h)
Let's now extend 4 to 4* with (i):
4*. (a) + (b) + (c) + (f) + (h) + (i)

Can $PUTB^{+}$ consistently instantiate 4*? The next theorem shows that we can answer this question affirmatively.

**Theorem 6.** $\boldsymbol{R}(PUTB^{+}) =_{Df} PUTB^{+} + \mathrm{GRP}_{PUTB^{+}}$ *is consistent.*[32]
This implies the following conclusion:

**Corollary 2.** $PUTB^{+}$ *is trustworthy.*

Moreover, the next lemma shows that $\mathbf{R}(PUTB^{+})$ recovers parts of requirement (e), compositionality, for all *T*-positive formulas. **Lemma 1.** *Let (T-* $\dot{\wedge}$*), (T-* $\dot{\vee}$*), (T-*$\forall$*) and (T-*$\exists$*) be respectively the axioms for the commutation of the truth predicate with* $\dot{\vee}$, $\dot{\wedge}$, $\dot{\forall}$, *and* $\dot{\exists}$. $\boldsymbol{R}(PUTB^{+})$ *proves (T-* $\dot{\wedge}$*), (T-* $\dot{\vee}$*), (T-*$\forall$*) and (T-*$\exists$*) for all*
T-*positive formulas.*[33]
These results allow us to state the following:

**Corollary 3.** *Let condition* $(e)^{+}$ *be compositionality for T-positive formulas and ex- tend 4* to* $4^{+} =_{Df}$ *(a) + (b) + (c)* $+ (e)^{+} +$ *(f) + (h) + (i).* $\boldsymbol{R}(PUTB^{+})$ *consistently instantiates* $4^{+}$.

[32] I cannot include my proof here for reason of space. This result holds also for $\omega$-iterations of reflection.
[33] For reason of space I cannot include the proof.

**Philosophical significance**
The aim of this paper was to offer a solution to Leitgeb's challenge and provide a way to pick our best truth-theory. As I showed, adding the trustworthiness requirement provides a solution to the challenge; *PUTB* is trustworthy and scores better than the other options. Moreover, the trustworthiness requirement is a philosophically interesting and epistemically well-motivated way of reasoning about truth-theories; it fits very well with the understanding of truth-theories as cognitive projects, but also with Feferman's view of proof- theoretic reflection principles, providing a new connection between the area of epistemology and philosophy of mathematics. Various questions remain unanswered: are there other good truth-theories in classical logic that are consistent with *GRP*? Are these theories better then *PUTB* and for what reasons? Also, the notion of trustworthiness deserves further investigation: one could try to spell it out also via other axiomatic principles different from the known reflection principles. This might involve an understanding of trustworthiness, which is weaker than the notion of soundness. All this is left open for future investigation.

[1] Cantini Andrea (1990), "A theory of formal truth arithmetically equivalent to $ID_1$", in The Journal of Symbolic Logic, 55:244 – 259.
[2] Feferman Solomon (1962), "Transfinite Recursive Progressions of Axiomatic Theories", in The Journal of Symbolic Logic Vol. 27, No. 3, 259 –316.
[3] Fischer Martin, Horsten Leon and Nicolai Carlo (2017), "Iterated reflection over full disquotational truth", in Journal of Logic and Computation, Volume 27, Issue 8, pp. 2631 – 2651, https://doi.org/10.1093/logcom/exx023.
[4] Halbach Volker (2009), "Reducing Compositional to Disquotational Truth", in The Review of Symbolic Logic Volume 2, Number 4, pp. 786 –798.
[5] ———— (2014), Axiomatic Theories of Truth, Oxford University Press.
[6] Leitgeb Hannes (2007), "What Theories of Truth Should be Like (but Cannot be)", in Philosophy Compass 2/2, 276 – 290.
[7] Tarski Alfred (1936), "The concept of truth in formalized languages2, in Logic, Semantics, Metamathematics, pp. 152 – 278.
[8] Wright Crispin (2004), "Warrant for nothing (and foundations for free)?", in Proceedings of the Aristotelian Society, Supplemental Volume LXXVIII, pp. 167 – 212.
[9] ———— (2012), "Replies Part IV: Warrant, Transmission and Entitlement", in (ed. by Annalisa Coliva) Mind, Meaning, and Knowledge. Themes from the Philosophy of Crispin Wright, Oxford University Press, pp. 451 –486.

**Michele Lubrano** (University of Turin)
*Difference-making and explanation in mathematics*

I would like to present an account of mathematical explanation along the lines of Strevens (2011), namely an account based on the notion of difference-maker. I'm going to illustrate what such an account consists in and why it deserves attention and further research effort.

Mathematical explanation is one of the most interesting aspects of mathematical practice. Professional mathematicians not only want their theorems to be correctly proven, they often want them satisfactorily explained. Only in relatively recent times philosophers have stated to pay attention to the issues of what a mathematical explanation consists in and how it works (see Steiner 1978). The growth of the literature on the topic in the last few years shows that, in the philosophical community, mathematical explanation has stated to be regarded as a key problem. There are two classical views of explanation within mathematics: a local model and a holistic model (I borrow the terminology from Mancosu 2018).

The *local model*, first presented by Steiner (1978), is one in which a proof of a theorem T is explanatory when T is deduced from the essence, or nature, of the mathematical objects involved. Being aware of the philosophical difficulties that we face every time we engage with the notion of essence, Steiner shifts toward the term "characterizing property", namely "a property unique to a given entity or structure within a family or domain of such entities or structures" (p. 143). Now, an explanatory proof is one that "makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is

evident that the result depends on the property" (p.143).

According to the *holistic model* (see Kitcher 1989) a proof of a theorem T is explanatory if it shows that the behaviour of the structures or entities mentioned in T can be subsumed under a general pattern, from which the behaviour of different structures or entities can be deduced. In other words, a proof is explanatory if it has a unificatory power, and if it allows us to "reduce the total number of independent phenomena that we have to accept as ultimate or given" (Friedman 1974, p. 15).

These two models have both virtues and limits, which I'm not going to examine here. What can be said is that there is a general consensus on the fact that the two models work well in some cases and are unsatisfactory in others (see Mancosu 2018). Some theorems are well explained by reference to an essential property of a structure or object the theorem is about, while some others are better explained by subsumption under a general pattern. The panorama looks favourable for a pluralist account of mathematical explanation, but before giving up the struggle for a unique general model, it is worth doing some other attempts and explore other possible ways of understanding the phenomena.

A good suggestion for a different way of understanding mathematical explanation comes from one of the two views just sketched, that of Steiner. He says, in his Steiner (1978), that one of the tasks that an explanatory proof of a theorem T must accomplish is to clearly indicate which are the properties T depends on. The notion of dependence might be the key for a deep understanding of mathematical explanation. The problem with this option is that, while there is an extensive literature on ontological dependence and causal dependence, no precise analysis of mathematical dependence has ever been attempted (as far as I know). The best thing to do in order to understand what "depending on" might mean in a mathematical context is to find a suitable example and see what lesson we can learn from it. The guiding example that I've chosen is offered by an interesting field of mathematics, known as *reverse mathematics*. The reasons why I find it particularly illuminating should become clearer soon.

Reverse mathematics is an important research program initiated by Friedman (1975), whose aim is do the reverse path of the most common mathematical research: instead of going from axioms to new theorems, it goes from already known theorems to their axioms. More precisely, the kind of questions that it aims at answering is: which is the weakest group of axioms that we need in order to prove theorem T of ordinary[34] mathematics? For a surprisingly high number of theorems, this question has a perfectly defined answer. Such an answer is often one of the several subsystems of Full Second Order Peano Arithmetic, in symbols, *Z2*. This theory is expressed in a formal language *L2*, provided by *i*) the standard logical connectives and quantifiers, *ii*) number variables ($m$, $n$, …) ranging over the members of set $\omega$ of natural numbers, *iii*) set variables ($X$, $Y$, …) ranging over subsets of $\omega$, and *iv*) and the symbols 0, 1, + , · , and <. Its axioms are the basic axioms of addition, multiplication, and inequality, plus two powerful axioms:

***Induction***: $[0 \in X \wedge \forall n(n \in X \rightarrow n{+}1 \in X] \rightarrow \forall n(n \in X)$

***Full Comprehension Axiom-Schema***: $\exists X \forall n(n \in X \leftrightarrow \varphi(n))$

The symbol $\varphi(n)$ stands for any formula of *L2*, in which $n$ occurs free. Full Comprehension must always be taken on the proviso that $X$ has no free occurrence in $\varphi(n)$.

Subsystems of *Z2* are theories whose axioms are either axioms or theorems of *Z2*. Several subsystems of *Z2* have been extensively investigated and have been ordered on the basis of their demonstrative power: the closer they are to demonstrating all the theorem demonstrated by *Z2* the higher they are in the hierarchy. The differences between these subsystems are very often entirely due to difference in the strength of Comprehension. In other words, most subsystems differ on the basis of which sets are taken to exist. Here I'll point my attention to two important subsystems: Arithmetic Comprehension Axioms (also known as *ACA0*)

[34] By ordinary mathematics I mean "that body of mathematics which is prior to or independent of the introduction of abstract set theoretic concepts" (Simpson 2009, p. 1). It includes, for instance, geometry, number theory, calculus, differential equations, real and complex analysis, etc.

and Recursive Comprehension Axioms (also known as *RCA0*). The former differs from *Z2* because the statement 'φ(n)' occurring in its Comprehension Axiom- Schema can only be arithmetical, namely quantification over sets is not allowed; 'φ(n)' must be such that its quantifiers bind only number-variables. The latter differs from *Z2* (and from *ACA0*) because the statement 'φ(n)' occurring in its Comprehension Axiom-Schema can only be recursive, which means that, not only quantification over sets is not allowed, but also quantification over number variables is significantly restricted: only occurrences of one kind of quantifier are allowed. Examples: $\forall n \forall m(n = m \leftrightarrow n+1 = m+1)$ is recursive, since there are only occurrences of the universal quantifier, while $\forall n \exists m(m = n+1)$ is not. It is easy to see that *RCA0* is less powerful than *ACA0*, which in turn is less powerful than *Z2*.

Now, this difference in demonstrative power can be made more precise by listing some theorems that can be proven in a subsystem but not in a weaker one. For example, Bolzano-Weierstrass Theorem[35] can be proven in *ACA0*, but not in *RCA0*. It can be shown that *RCA0* is the most powerful subsystem of *Z0* in which Bolzano Weierstrass statement is false and *ACA0* is the weakest in which it is true. Since the only difference between the two lies in the strength of Comprehension Axiom Schema, the most natural way to describe the situation is to say that the strengthening of Comprehensions (i.e. its upgrading from Recursive to Arithmetical) is what *makes the difference* between Bolzano-Weierstrass statement being true and its being false. In the context of a hierarchy of subsystems of *Z2*, Arithmetic Comprehension is what such a statement *crucially depends* on. The method of reverse mathematics is able to individuate such difference-makers in a precise way, for a large number of theorems of ordinary mathematics. The relation of crucial dependence that is in play here and how it is connected with the notion of difference-maker can be illustrated by means of this definition:

**Crucial Dependence**: the truth of a statement *T* crucially depends on axiom *A* if and only if, given a hierarchy of systems of increasing strength (S1, …, Sn),

Si instead of Si-1 proves *T* instead of *not-T*, and Si = Si-1 ∪ {A}.

This definition is a good generalisation of the phenomenon described in the example of *RCA0* and *ACA0*. Indeed, *ACA0* instead of *RCA0* proves Bolzano- Weierstrass statement instead of its negation. Moreover, adding Arithmetic Comprehension to *RCA0*, we get *ACA0*. Therefore, Arithmetic Comprehension is what makes the difference between proving Bolzano-Weierstrass statement or its negation, in the context of subsystems of *Z2*. Bolzano-Weierstrass statement crucially depends on Arithmetic Comprehension, in such a context. The fact that this way of understanding mathematical explanation is fundamentally based on the idea of difference-making is what makes me think that it is a natural development of Strevens' account of scientific explanation (see Strevens 2011) and therefore it might be meaningfully named a *kairetic account of mathematical explanation*.

A couple of things deserve to be noticed. First, the crucial dependence relation is a tetradic one and it is presented in terms that are pretty close to the contrastive theory of causation (see Schaffer 2005). Second, crucial dependence must be carefully distinguished from *generic dependence*. We can certainly say that Bolzano-Weierstrass Theorem partially depends on each of the axioms of *ACA0*, since *ACA0* provides the *minimal* bunch of axioms required in order to prove it. None of such axioms is irrelevant for the truth of Bolzano-Weierstrass statement. If we want to exploit an analogy with grounding, we can say that each *ACA0* axiom is a partial ground for Bolzano-Weierstrass statement. Nonetheless, not all these partial grounds are on the same level. As we have already seen, *RCA0* shares with *ACA0* all the axioms except for Arithmetic Comprehension. But in *RCA0* Bolzano- Weierstrass statement is false. That is to say that that each axiom of

---

[35] Bolzano Weierstrass Theorem is fundamental theorem of analysis stating that every bounded sequence in $\Re^n$ has a convergent subsequence.

*RCA0* is a partial ground for the negation of Bolzano-Weierstrass statement. These axioms are sort of *double agents*:[36] in the absence of Arithmetic Comprehension they contribute to proving the negation of Bolzano-Weierstrass statement, in its presence they contribute to proving it. In the absence of an already settled hierarchy of formal systems (which is not necessarily available in every branch of mathematics), the difference-making axiom can be found by contrasting a bunch of axioms that, relatively to a statement T, play as double agents to the axiom(s) that, once added to that bunch, make(s) T definitely provable. Suppose that Γ is a set of axioms able to prove non-T and it is such that each of them is relevant for such a proof. Suppose again that adding a suitable axiom α to Γ we get Γ', which proves T. Then α is the difference-maker and T crucially depends on α relatively to a Γ-Γ' hierarchy.

Now, we are able to formulate a simple statement of the kairetic account of

mathematical explanation: a proof of a theorem T is explanatory if and only if T is deduced, among other things, from a make-difference axiom in the context of a suitable hierarchy of formal systems.

Finally, let's briefly see what virtues can be ascribed to this account of mathematical explanation. The main virtue is that it gives good predictions on the explanatory power of some proofs. I will show that a simple induction proof that is generally considered (by professional mathematicians) not explanatory is such that the application of the kairetic account makes immediately evident the reason why it is not explanatory: induction is not a make-difference axiom in that context. A proof of the same theorem that is generally considered as explanatory will be examined, showing that it relies on a couple of make-difference axioms. Another virtue is that, if a certain theory is inconsistent it is able to explain which the source of the inconsistency is. In particular it is able to individuate an axiom able to make a difference between a certain theory being consistent and its being inconsistent.

In the end, I think that the kairetic account of mathematical explanation is a promising one and it certainly deserve to be both developed in more details and checked in a number of different cases, in order to see whether its predictions on the explanatory power of proofs are in accordance with the opinion of the majority of professional mathematicians.

Friedman, Harvey 1975. 'Some systems of second-order arithmetic and their use'. Proceedings of the International Congress of Mathematicians, Vancouver 1974, Vol. 1: 235-242.

Friedman, Michael 1974. 'Explanation and Scientific Understanding'. The Journal of Philosophy, 71 (1): 5-19.

Kitcher, Philip 1989. 'Explanatory Unification and the Causal Structure of the World, in P. Kitcher and W. Salmon (eds.), Scientific Explanation, (Minnesota Studies in the Philosophy of Science, Volume XIII), Minneapolis: University of Minnesota Press, 410–505.

Krämer, Stephan & Roski, Stefan 2017. 'Difference-making grounds'. Philosophical Studies, 174 (5): 1191-1215.

Mancosu, Paolo 2008. 'Explanation in Mathematics'. The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), URL =
<https://plato.stanford.edu/archives/sum2018/entries/mathematicsexplanation/>.

Schaffer, Jonathan 2005. 'Contrastive causation'. The Philosophical Review, 114 (3): 327-358.

Simpson, Stephen 2009. Subsystems of Second Order Arithmetic (2nd Edition). Cambridge University Press.

Steiner, Mark 1978. 'Mathematical Explanation'. Philosophical Studies, 34 (2): 135-151.

Strevens, Michael 2011. Dept: An Account of Scientific Explanation. Harvard University Press.

[36] I borrow the term and the underlying idea from Krämer & Roski (2017).

## EIGHTH SESSION: Foundations of Computing and Artificial Intelligence

**Mirko Tagliaferri** (University of Urbino)
*How to Build a Formal Notion of Trust*

Many disciplines recognize the importance of trust for the emergence and development of collaborative behaviours [Arrow 1972, Luhmann 1979]. However, being a multi-faceted concept, trust always defiled a comprehensive analysis that could define its core features and thus identify it as a clear notion. This aspect is highly problematic when the concept has to be modelled and, successively, implemented in formal environments. Therefore, it comes with no surprise that there is little consensus, in the computer science literature (and no consensus at all in the logical literature), on the nature of computational trust and how to properly model it. Even though disagreements in scientific research are not rare and neither exceptionally troublesome in most cases, the lack of a unified conceptualization of the notion of trust is a big issue when it is realized that social interactions are gradually transitioning from the physical realm to the digital one [Botsman 2011, Floridi 2014]. In digital environments, all the trust-relevant biological traits that human beings intuitively identify are missing. Trusting or not can't be a matter of instinct anymore and effective mechanisms to establish trust relationships must be explicitly implemented in the design of the digital systems. Those mechanisms can then aid the users to consciously assess whether to trust or not another user during interactions. Moreover, the same mechanisms might help digital agents (either softwares or machines) to make decisions based on social norms they are not ordinarily programmed to implement. In short, having explicit and formal notions of trust implemented in a digital environment, might help all sort of interactions which take place in this specific environments (i.e.: human-human, human-machine and machine-machine).
The aim of the paper is two-fold. It is, first, a methodological paper, highlighting a plausible procedure to generate formal notions (and subsequent formal implementations) of social concepts, which are commonly thought to be difficult to correctly formalize. Moreover, the paper proposes an actual instance of the methodology introduced, providing a novel formalization of trust through the introduction of a logical language; it is then shown how the language can be employed to implement trust as a computational notion in computational environments.
The paper is thus structured: in the first section, emphasis is placed on how to produce valuable conceptual analyses. The scope is philosophical and the aim is to identify some necessary and sufficient conditions to produce useful conceptualizations of notions. Stress is places on advantages and issues of standard philosophical conceptual analysis [Block & Stalnaker 1999, Chalmers & Jackson 2001, Fodor 1998, Margolis & Laurence 1999]. General results are derived and a first step in the direction of building proper conceptual analyses of social notions is made. The section is concluded by a thorough example of such methodology. The notion of trust is thus conceptually analysed: starting from previous generalist analysis of trust, various discipline-specific studies of trust are summarized and then merged into a unified theory. This theory is then review with respect to laboratory experiments on trust. Specifically, results from biology [Bateson 1988, Trivers 1971], sociology [Barber 1983, Durkheim 1893, Luhmann 1979], economy [Fehr 2009, Granovetter 1985, Williamson 1993] are exposed and the core features of trust are derived. Those features are then compared with the necessities of formal digital frameworks. This is done by isolating the standard paradigms concerning trust in the computer science literature [Jøsang & Ismail 2002, Kamvar et al. 2003] and then deriving from them mandatory conditions on the features of the digital environments in which trust must be modelled. The two sets of features are therefore conflated, obtaining a general set that is consistent with both analyses. In the second section, focus is placed on how to actually transform conceptual analyses into formal representations of social notions. The scope is again philosophical and the aim of that of providing proper tools to transform already analysed social notions into formal versions of them. Again, the methodological part is then followed by an explicit example, which employs the previous analysis of trust and tries to produce a novel formalization for such notion. Specifically, a modal language augmented with a trust operator is presented, whose semantics is given with respect to neighbourhood structures [Montague 1970, Scott 1970, Pacuit 2017]. Decidability results for the language and then proved and general considerations about its expressivity and relation with other formalizations of trust in the computer science literature are made. In particular focus is placed on the relation of the language with Subjective Logic (i.e., one of the standard models for trust in computer science) [Jøsang 2016] and Dempster-Shafer Theory of

Evidence (i.e., one of the standard models for evidence and uncertainty representations) [Dempster 1967, Shafer 1976].
All researchers in the area of computational trust can greatly benefit from the methodological insights presented in this paper on how to construct computational versions of social notions. Therefore, the paper achieves three goals: i) it builds a methodology suited to build formal version of social notions; ii) it provides deep insights on the notion of trust; iii) it sustains the results proposed with a thorough example by presenting an actual logical language employable to reason about trust.

K. Arrow, "Gifts and Exchanges", Philosophy and Public Affairs 1(4), pp. 343-362, 1972.
B. Barber, "The Logic and Limits of Trust", Rutgers University Press, 1983.
P. Bateson, "The Biological Evolution of Cooperation and Trust", in: D. Gambetta (ed.), Trust: Making and Breaking Cooperative Relations, Blackwell, pp. 31–48, 1988.
N. Block, R. Stalnaker, ''Conceptual Analysis and the Explanatory Gap'', Philosophical Review 108(1), pp. 1-46, 1999.
R. Botsman, "What's Mine is Yours: How Collaborative Consumption is Changing the Way We Live", Collins, 2011.
D. Chalmers, F. Jackson, ''Conceptual Analysis and Reductive Explanation'', Philosophical Review 110, pp. 315-361, 2001.
A.P. Dempster, "Upper and Lower Probabilities Induced by a Multivalued Mapping", The Annals of Mathematical Statistics 38(2), pp. 325–339, 1967.
E. Durkheim, "The Division of Labor in Society", MacMillan, 1893.
E. Fehr, "On the Economics and Biology of Trust", Journal of the European Economic Association 7, pp. 235–266, 2009.
L. Floridi, "The 4th Revolution: How the Infosphere is Reshaping Human Reality", Oxford University Press, 2014.
J. Fodor, ''Concepts: Where Cognitive Science Went Wrong'', Oxford University Press, 1998.
M. Granovetter, "Economic Action and Social Structure: the Problem of Embeddedness", American Hournal of Sociology 91, pp. 481–510, 1985.
A. Jøsang, "Subjective Logic", Springer, 2016.
A. Jøsang, R. Ismail, "The Beta Reputation System", Proceedings of the 15th Bled Electronic Commerce Conference, e-Reality: Constructing the e-Economy, pp. 324–337, 2002.
S.B. Kamvar, M.T. Schlosser, H. Garcia-Molina, "The EigenTrust Algorithm for Reputation Management in P2P Networks", Procs. of the 12th International Conference on World Wide Web, pp. 640–651, 2003.
N. Luhmann, "Trust and Power", John Wiley and Sons Inc, 1979.
E. Margolis, S. Laurence, ''Concepts: Core Readings'', MIT Press, 1999.
R. Montague, "Universal Grammar", Theoria 36, pp. 373–398, 1970.
E. Pacuit, "Neighborhood Semantics for Modal Logic", Springer, 2017.
D. Scott, "Advice on Modal Logic", Philosophical Problems in Logic, pp. 143–173, 1970.
G. Shafer, "A Mathematical Theory of Evidence", Princeton University Press, 1976.
R.L. Trivers, "The Evolution of Reciprocal Altruism", The Quarterly Review of Biology 46(1), pp. 35–57, 1971.
O. Williamson, "Calculativeness, Trust, and Economic Organization", Journal of Law and Economics 36(2), pp. 453–486, 1993.

**Sandro Sozzo** (University of Leicester)
*Entanglement and Quantum Structures in Concept Combinations*

We investigate the presence of entanglement outside the microscopic domain of quantum physics, specifically in Bell-type experiments on concepts and their combinations. We present the results of two empirical tests, a cognitive test on human participants and a document retrieval test on corpuses of documents on the web. In both cases, empirical data significantly violate the Clauser-Horne- Shimony-Holt version of Bell's inequalities (CHSH inequality), which is typically accepted to indicate the presence of entanglement between the considered concepts. We also work out a quantum model of the web test, which agrees with a general quantum-theoretic framework, developed by ourselves to identify entanglement in empirical situations

violating Bell's inequalities. We finally represent the collected data in Hilbert space and show that the violation of the CHSH inequality is due to the occurrence of a strong form of entanglement, involving both states and measurements and reflecting the meaning connection between the component concepts. These results fit an emerging research that successfully applies quantum structures, detached from any physical interpretation, to situations where the application of traditional Kolmogorovian structures is problematic.

[1] Aerts, D. and Sozzo, S. (2011). Quantum structure in cognition: Why and how concepts are entangled. Quantum Interaction. Lecture Notes in Computer Science 7052, 116–127.
[2] Beltran, L. and Geriente, S. (2018). Quantum entanglement in corpuses of documents. Foundations of Science, doi 10.1007/s10699-018-9570-2.
[3] Bell, J. (1964). On the Einstein Podolsky Rosen paradox. Physics 1, 195–200.
[4] Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R.A. (1969). Proposed experiment to test local hidden-variable theories. Physical Review Letters 23, 880–884.
[5] Aerts, D. and Sozzo, S. (2014). Quantum entanglement in concept combinations. International Journal of Theoretical Physics 53, 3587–3603.
[6] Aerts, D. (2009). Quantum structure in cognition. Journal of Mathematical Psychology 53, 314–348.
[7] Busemeyer, J. and Bruza, P. (2012). Quantum Models of Cognition and Decision. Cambridge: Cambridge University Press.
[8] Aerts, D., Broekaert, J., Gabora, L. and Sozzo, S. (2013). Quantum structure and human thought. Behavioral and Brain Sciences 36, 274-276.
[9] Haven, E. and Khrennikov A. Yu. (2013). Quantum Social Science. Cambridge: Cambridge University Press.
[10] Melucci, M. (2015). Introduction to Information Retrieval and Quantum Mechanics. Berlin Heidelberg: Springer.

**Silvia Crafa** (University of Padua) **Lucrezia Pelizzon** (University of Cagliari)
*Epistemological questions for a philosophical education in artificial intelligence*

Artificial Intelligence (AI) is one of the most powerful transformative forces of our time. Differently from other software systems, AI is able to act within a complex system, such as the physical or digital world, to interpret big data collected in a structured way or not, and to decide, based on the knowledge derived from them and according to predefined parameters, the best actions to be taken to achieve the given objective. AI systems based on Machine Learning are, therefore, designed to be flexible and to adapt their behaviour to the way in which the environment has been modified by users or by their previous actions. Therefore, similarly to human understanding, AI systems have the ability ``to learn''.

As it has happened for the previous technological revolutions, the scientific, economic and social consequences of AI are profound and still for the most part unpredictable. Their use raises concrete problems, both from a technical, ethical and legal point of view. On the one hand, software developers have the urgent need to implement solutions that optimize the obtained results to quickly solve practical design problems. On the other hand, analysis reports and ethical codes are rapidly spreading out recommendations (see [7]) and guidelines for the development of an AI which guarantees both respect for the fundamental human rights, principles and values, and which promotes, at the same time, a reliable technology and a sustainable development of the society (see EU's recent Draft Ethics Guidelines for Trustworthy AI [3]). Since AI-based tools and services will have an impact on many aspects of the human life, these reports are written not only by scientists and engineers, but also by people with different expertise and roles in the society. These experts are involved at design level to support the development of this raising technology. Therefore, scholars are called for a dialogue between different disciplines about social, cognitive, ethical and legal issues, so to let the cross-disciplinary research play a fundamental role in providing effective, admissible, and implementable technological answers.

Given the complexity of the AI phenomenon, it is necessary to have a specific knowledge of this subject to properly understand the real impact of AI technology and, at the same time, to provide a practical contribution to the new problems AI is posing. However, it is less evident what is the clear and precise

description of the subjects to be studied and of the knowledge to be achieved from the philosophical point of view. In a more explicit way: what kind of training must have the philosopher in order to acquire skills on this topic, considering that AI evolves with exponential rapidity? What disciplines must know and what research methodology must apply the philosopher to obtain, without losing depth, modern answers to the rising problems of the technological revolution and to those that software developers are facing?

The first step for a philosopher to do research in AI will be to acquire a technical competence in the subject and, therefore, to accept the challenge of facing a cross-disciplinary study path. To understand how to deal in practice with a cross-disciplinary training is not trivial, because the objective of a philosopher is not to overlap the programmer expertise, but to have a cross disciplinary preparation on AI with proper objectives of philosophy. Then we ask: which technical subjects must be tackled and which level of specialization must we reach without losing the proposed objectives? And which research methodology should be adopted?

In the case of applied disciplines, training should not be limited to theoretical studies, but it should be integrated with practical (field) experience, interacting with people involved in developing this type of technology, and attending specifically dedicated laboratories and research centers. The conversation with programmers would give the opportunity to understand the topic from a privileged point of view, and would allow, not only to be updated on its developments, but to be at the heart of the application problems of this technology.

If the research objective is to understand the effects of this technology and to make a concrete contribution to the problems it is posing, this conversation could offer interesting insights for both parties, turning into a two-way connection. The proposed training will transmit to philosophers the needed technical knowledge to robustly deal with the fallout of AI technology in the world, and to AI experts a wider awareness of their design choices and the possible consequences of their work in the real world.

It is also desirable that the philosophical contribution can provide proposals to solve semantic problems: how to deal with imprecise or ambiguous concepts typical of natural language, whereas, instead, the algorithms require explicit representations; or how to appropriately formalize implicit concepts in order to propose ethical solutions to the problem of discriminating algorithms (see [2]).

Obviously, the dialogue between such different disciplines is not easy to achieve, not only for a question of skills, but also for a question of different research methodologies and languages, that are difficult to properly combine without losing their specificity. However, the complexity of the phenomenon seems to encourage us to address the challenge to produce new insights, as already shown by the interdisciplinary collaboration between engineers and philosophers. On the other hand, the specificity of the philosophical field cannot be solved in a simple way, whether it involves investigating the effects of AI, or in identifying the training path to be undertaken to acquire the appropriate skills. If we started from the idea that the first step to do research in AI is to acquire a technical preparation appropriate for the philosophical purposes, the second step must be accomplished by maturing a solid philosophical preparation to achieve those goals: but which philosophical disciplines will have to be studied?

If the goal is to contribute to the clarification of semantic problems, then training will focus on logic (fuzzy logic, for example). If the goal is, instead, to clarify the ethical implications of AI, then training must again be cross-disciplinary, this time in the narrower sense of comparison between different philosophical sectors. Ethics is, in fact, a specific branch of philosophy, which, applied to AI, must be addressed with a knowledgeable approach. The questions raised by AI are inherently difficult to solve, even in the framework of a moral debate having as its object of study the development of science and technology. For example, one of the legal and moral problems posed by self-driving cars is the fact that the AI algorithms for self-driving cars work autonomously, taking decisions that can also result in the death of human beings (see [13]). This means that, for the first time in history, the reflection on the responsibility of an action or of a choice of such significance no longer has man at his center: can a ``Copernican revolution" for morality (and for right) of this greatness be approached with a traditional separation of knowledge? Can the philosophy of science and moral philosophy give a disjoint answer when questioning what it means to delegate individual responsibility to the action of a technology? and on the wider consequences that this will have on our way of thinking

about life in the world? It would be interesting to understand whether the answers we are providing are ethically adequate and well suited to our society or whether they are only expression of a Western culture that must face a globalized world with conceptions of ethics different from ours, as demonstrated by the recent genetic manipulations carried out in China in the biomedical field.

The ethical implications of AI show that the questions posed by this technology are so radical that the need for an interdisciplinary approach is not restricted to moral philosophy, but must be broader. To recall the words that Russo used about Information Technology, these algorithms ``bring about profound changes at the ontological and epistemological level'' ([10, p.2]), as they are transforming the world, more than other technological achievements have already done in the past.

Therefore, it is possible that also for AI ``the ethical discourse (…) must recover a unity, an explicit connection with the fundamental issues related to knowledge and metaphysics'' ([10, p.5]) and that we should find ``a conceptual framework that allows us to hold together science and technology on the one hand and ethics, epistemology, and metaphysics on the other'' ([10, p.5]): ``what the world is (i.e., ontology), how we can get knowledge of the world (i.e., epistemology), and the normative dimension at ethico-political level (...) are essentially related and interconnected. However, due to the hyper-specialisation of the sciences and of philosophy, they grew into distinct sub-disciplines that, by and large, talk past each other rather than to each other'' ([10, p.11]).

This enterprise is undoubtedly difficult to achieve, but it might be worth trying because it could ``provide with the opportunity to bring together ethics, ontology, and epistemology in a coherent approach'' ([10, p.12]). Otherwise the risk could be, once again, that of not reaching the goal of fully understanding this phenomenon. It remains problematic to define how we can create a virtuous circle of cross-disciplinary research, this time from the philosophical point of view that, beyond the best intentions, it is feasible and produces effective conceptual tools in a world that is changing too quickly.

## Conclusion

Our goal was to highlight some issues that deserve further examination and discussion, especially to tackle the difficulty faced by those who, at this time of great technological changes, want to acquire an effective philosophical preparation in the subject of AI, to understand the phenomenon and to make a useful contribution to the AI practical problems. The distinctive feature of AI is its cross-disciplinary nature, both because, in a broadest sense, the object of study falls into a completely different discipline such as Computer Science, and because the disruptive consequences of this technology have such repercussions to fall within the areas of expertise of other philosophical disciplines (such as ethics and ontology). In both cases it is desirable to better understand how to create a virtuous circle of cross-disciplinary research, in which the dialogue with different disciplines is an opportunity to acquire new and necessary skills and to provide a concrete contribution to the general progress of research. From the methodological and content points of view, the most significant difficulty is getting the preliminary technical preparation. Given the vastness and complexity of the knowledge necessary to master the subject, it is important to understand to what extent it is necessary to go deep, without losing the objectives that are proper to philosophy. An important help can be given by the direct interaction with people dealing with this technology. From the philosophical point of view there are two types of difficulties about the preparation necessary to understand the impact of AI. The effects of AI are so pervasive that they must be treated by different philosophical areas and so radicals to challenge the consolidated categories of interpretation. Given these premises, cross-disciplinary research can be characterized in this case as the challenge of achieving a dialogue among philosophical disciplines that, beyond the specificity of each, gives a consistent approach to the subject.

[1] Amoroso D., Tamburrini G. (2017), The Ethical and Legal Case Against Autonomy in Weapons Systems, Global Jurist, Vol. 18(1).
[2] Crafa S. (2018), Artificial Intelligence and Human Dialogue, submitted to journal publication.

[3] EU Commission's High-Level Expert Group on Artificial Intelligence (2018), Ethics guidelines for trustworthy AI (Draft). Available on https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai
[4] Floridi L. (2016), The 4th Revolution: How the Infosphere is reshaping Human Reality, Oxford University Press.
[5] Heidegger M. (1977), The question concerning technology and other essays, Garland Publishing.
[6] Heidegger M. (1966), Discourse on Thinking, Harper & Row Publisher, New York.
[7] Informatics Europe e ACM-Europa (2018), When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making, available on http://www.informatics-europe.org/working-groups/ethics.html
[8] Jonas H. (1985), The Imperative of Responsability: in Search of an Ethics for the Technological Age, University of Chicago Press.
[9] Moor J. (1985), What is Computer Ethics?, Metaphilosophy, Vol. 16(4), pp. 266-275.
[10] Russo F. (2018), Digital Technologies, Ethical Questions, and the Need of an Informational Framework, Philosophy & Technology, Springer-Netherlands, https://doi.org/10.1007/s13347-018-0326-2.
[11] Schiaffonati V. (2003), A Framework for the Foundation of the Philosophy of Artificial Intelligence, Minds and Machines, Vol.13(4), pp. 537-552.
[12] Shakib J., Layton D. (2014), Interaction between Ethics and Technology, IEEE International Symposium on Ethics in Science, Technology and Engineering, IEEE Publisher.
[13] Tamburrini G. (2017), Autonomia delle macchine e filosofia dell'intelligenza artificiale,
Rivista di Filosofia, Vol. 108(2), pp. 263-275.