Fourth Philosophy of Language and Mind Conference

September 21-23, 2017

Ruhr University Bochum

Program & Abstracts

DFG Deutsche Forschungsgemeinschaft

RUB

Preface

PLM is a European network of centers devoted to the philosophy of language and mind. It was founded in 2010 and organizes international conferences, workshop and master classes taught by leading experts in the field. PLM includes the following centers, departments and institutes (with representatives in brackets) and is open for further membership applications from leading European institutions with a focus on the philosophy of language and mind:

- Arché, St. Andrews (Derek Ball)
- Department of Philosophy, CEU, Budapest (Ferenc Huoranszki)
- CLLAM, Department of Philosophy, Stockholm University (Peter Pagin)
- ILCLI, University of the Basque Country, San Sebastian (Kepa Korta)
- ILLC, University of Amsterdam (Paul Dekker)
- Institut Jean Nicod, Paris (François Recanati)
- Institute of Philosophy, London (Corine Besson)
- LanCog, University of Lisbon (Pedros Santos)
- LOGOS, University of Barcelona (Michele Palmira)
- Department of Philosophy II, Ruhr University Bochum (Markus Werning, speaker)

The main purpose of PLM is to further the philosophy of language and mind in Europe and to provide a platform for cooperation between members in research and in research training and to enhance joint applications to funding activities and organizations. PLM organizes a major language and mind conference every two years, with invited and contributed talks. Selected papers of the PLM conferences are published in renowned international journals. Previously, special issues from PLM conferences appeared in the journals *Synthese* and *Review of Philosophy and Psychology*.

The Fourth Philosophy of Language and Mind Conference (PLM4) will take place at Ruhr University Bochum from 21 to 23 September 2017. The venue on September 21 is the GA-building, Ruhr University Bochum, Universitätstraße 150, Bochum, Germany. The venue on September 22 and 23 is the Center of Internet Security (ZITS), Lise-Meitner Allee 4, Bochum.

PLM4 will bring together internationally renowned philosophers of language and mind as well as young researchers in these fields with a focus on Europe. We received 171 submissions. Submissions were reviewed in a two stage process by the members of the program committee. In addition to 8 plenary talks, 56 full talks, 22 spotlight talks and 26 poster presentations have finally been included in the program.

We are grateful for a generous grant from the German National Science Foundation DFG in support of PLM4.

Bochum, September 2017 Markus Werning

Table of Contents

– Click on page numbers below to jump to pages! –
– Click on slot numbers on the top right corner of abstracts to jump to program tables! –

Preface i

Program Committee viii

Program ix

Program Table x

List of Poster Presentations xv

Maps and Directions xvi

Invited Papers 1

The folk concept of pain is paradoxical: defending a polyeidic approach 1
Emma Borg and Tim Salomons

Are racial biases in perception truly automatic, and why does this matter? 2
Ophelia Deroy

Concept Utility 3
Paul Egré and Cathal O'Madagain

Language and embodiment – an evolutionary perspective 4
Nikola Kompa

Perspectival Content and Expressing an Attitude vs Asserting that One Has It 5
Max Kölbel

Seeing Things You Don't See 6
Bence Nanay

Cognition and Perception: Does higher-order background information influence our perceptual experience? 7
Albert Newen

The Interaction of Bayesian Pragmatics and Lexical Semantics in Sentence Meaning Composition .. 8
Markus Werning and Erica Cosentino

Contributed Papers 9

Millikan's Teleological Theory of Content, Indeterminacy Problem and Temporal Externalism 11
Keyvan Alasti

A combination problem for Self-Representationalism? 13
Alberto Barbieri

Perceptual Experiences: Content, Situation, Mode 15
Gregory Bochner

Semantic Restrictivism 17
Mark Bowker

Pain, Multimodality, and Bodily Awareness 19
Adam Bradley

Perceptual constancy and the role of perception in action 22
Alessandra Buccella

Model-theoretic semantics and semantic competence 24
Fabrizio Calzavarini

The empirical status of the pictorial view of meaning ... 27
Fabrizio Calzavarini

Doing and Undoing. Some Strategies to Backpedal on Our Speech Acts ... 30
Laura Caponetto

Towards a Shared Frame for Imaginative Episodes ... 33
Joan Casas Roma, Antonia Huertas and M. Elena Rodriguez

The reclamation of slurs ... 34
Bianca Cepollaro

Compositional Conceptual Role Semantics ... 36
Tadeusz Ciecierski and Paweł Grabarczyk

A Classical Implicational Account of Semantic Presuppositions ... 38
Roberta Colonna Dahlman

Semantic externalism(s) and content transparency ... 40
Philippe De Brabanter and Bruno Leclercq

Gender presuppositions in conditionals ... 42
Fabio Del Prete and Alessandro Zucchi

Names are Polyreferential ... 44
Laura Delgado

The Logical Geometry of Russell's Theory of Definite Descriptions ... 46
Lorenz Demey

A direction effect on taste predicates ... 48
Alexander Dinges and Julia Zakkou

Open-textured and Dynamic Meaning as a Ground for a New Metasemantic Framework for Natural Languages ... 50
Matej Drobňák

Do we hear meanings (and should we care whether we do)? ... 53
Anna Drożdżowicz

The Nature of the Fundamental Doxastic State ... 55
Edward Elliott

It's a wonderful world. Our language about the attitudes ... 57
Giulia Felappi

An Action-Based Theory of Picture Perception ... 59
Gabriele Ferretti

Language and Conceptual Development: a Philosophical Review of the Empirical Research ... 61
Mara Floris

Making sense of self talk ... 63
Bart Geurts

Introspection without Judgment ... 65
Anna Giustina

On the location of pain ... 68
Simone Gozzano

QUDs and Context-Sensitivity ... 69
Jumbly Grindrod

The untenability of nativism in current research on mindreading ... 70
Arkadiusz Gut and Robert Mirski

Cause for Concern Searle's Background Theory and the Context Sensitivity of Thought 72
Neil Hamilton Fairley

Deferring to Future Speakers 73
Jussi Haukioja

Why computational explanations are not mechanistic: An argument from idealization 75
Shahar Hechtlinger

Beyond paradoxes and bifurcations: Proposal for a unified theory of metaphor based on speaker's intention 76
Andreas Heise

Generic Disjunctive Syllogism 77
Miguel Hoeltje

Forced Choices: Does accepting its consequences suffice for deeming a deed intentional? 79
Philipp Huebl

Ironic name reference as echoic mention – A pragmatic analysis based on empirical data 80
Holden Härtl

The semantic content of imperatives 82
Mark Jary and Mikhail Kissine

Confabulation and Rational Requirements for Self-Knowledge 84
Sophie Keeling

Polysemous names 86
Katarzyna Kijania-Placek

Too-Shifty Indexicals in Answering Machine Puzzle 88
Jiwon Kim

On the Availability of Presuppositions in Conversation 90
Dirk Kindermann

Bildung, Second Nature, and Conceptual Content in McDowell's "Mind and World" 92
Vitaly Kiryushchenko

Norms of Assertion: Empirical Data 94
Markus Kneer

Mechanisms and representations do not mix 97
Matej Kohar

Are the Mental States Underlying Implicit Biases Unconscious? – A Neo-Freudian Answer 99
Beate Krickel

Making Sense of the Emergent Property: Association or Inference? 101
Hsiulin Ku

A Unified, Kaplanian Metaphysics for Linguistic Types 102
Giorgio Lando

Multipropositionalism and Necessary A Posteriori Identity Statements 104
Armando Lavalle Terrón

Against Predicativism of Names 106
Jeonggyu Lee

On the Uniformity of Proper Names 108
Junhyo Lee

Towards a Unified Semantics for Extensional and Cognitive Verbs 110
Kristina Liefke

Abstract Concepts, Compositionality and The Contextualism-Invariantism Debate . . . 112
Guido Löhr

What about the colours? Shoemaker on appearance properties . . . 114
Giulia Martina

The Normative Role of Logic . . . 116
Joshua Matthews

A commitment to haecceitism in Robert Stalnaker's solution to indexical beliefs . . . 118
Maria Matuszkiewicz

A New Metapher Illustrating the Vector Based Information Processing in Systemtheoretical Cognitive Neuroarchitectures in Connectionistic Cognitive Science . . . 119
Harald Maurer

Challenging Evolutionary Psychology with the eternalist view of the world . . . 121
César Meurer

De Se Skepticism and Rational Action . . . 122
Maik Niemeck

How to account for Subjective Character? . . . 124
Maik Niemeck

Sub-sentential speech acts: a situated contextualist account . . . 126
Joanna Odrowaz Sypniewska

Handling Concurrency through Consciousness . . . 128
Erica Onnis

Cognitive complexity and mathematical problem solving . . . 130
Markus Pantsar

Semantic Antirealism and the Self-Ascription of Attitudes . . . 132
Fabrice Pataut

A Neurobiological Case for Animal Rationality . . . 134
David Evan Pence

Recalcitrance Poses No Threat to Cognitivism . . . 136
Kathryn Pendoley

A Puzzle for Semantics . . . 138
Mark Pinder

Degrees of Analyticity . . . 140
Salvatore Pistoia Reda and Uli Sauerland

Two critical remarks to classical identity theory . . . 142
Michal Polak

Context and communicative success . . . 143
Joanna Pollock

Is the debate about doxasticism merely verbal? . . . 145
Krzysztof Poslajko and Bartosz Janik

The Recognitive Function of Explicit Expressions of Belief . . . 147
Blaise Prentice-Davidson

How to Twin-Earth a Phenomenal Concept . . . 149
Gabriel Rabin

Predication, Meaning and Transparent Hyperintensional Logic . . . 151
Jiří Raclavský

How to Represent the Characters of Indexicals 153
Alexandru Radulescu

How to talk about subjective experience? 155
Thomas Raleigh

Attitude verbs as reference-shifting operators 157
Dolf Rami

Kripke, Functionalism, and the Formal Indeterminacy of the Physical 159
Antonio Ramos Díaz

Chalmers on Indexicals 161
Stefan Rinner

Naive Russellians and the Goldbach Puzzle 163
Stefan Rinner

Literal meaning, idiolects, and language learning – the impossible equation 165
Sami Rissanen

Hume's maxim with impossible worlds 167
Pierre Saint-Germier

Unpleasant pain = Δ 169
Abraham Sapien

A problem with armchair arguments against explaining consciousness 171
Astrid Schomäcker

Mentalised prototype frames 173
Annika Schuster and Leda Berio

Fiction, lies and assertions 175
Merel Semeijn

My Body is the Subject's Body 177
Carlota Serrahima

The Acceptability and The Probability of the Indicative Conditionals 179
Michał Sikorski

Does the sense of agency include a phenomenology of freedom? 182
Andrew Sims

Neurocognitive evidence in the semantics-pragmatics interface debate: the case of implicatures 184
Maria Spychalska

Processing of affirmation and negation in contexts with unique and multiple alternatives:
Evidence from event-related potentials 186
Maria Spychalska, Viviana Haase, Jarmo Kontinen and Markus Werning

Uncertain Progress 189
Alexander Stathopoulos

Modification: Pragmatics and Constraints 190
Corina Ströß ner and Gerhard Schurz

Social Externalism and Ordinary Speech 192
Maik Sühr

The Forms of Words 194
Nick Tasker

Beyond the What is Said Debate: Towards a Psychologistic Theory of Successful Communication .. 196
Samuel Taylor

Lexical files as encyclopedic files 198
Antonin Thuns

Self-Ascription and Agency of the First-Person 200
Chiu Yui Plato Tse

A Mechanistic Take on the Theory Theory of Concepts 202
Matthias Unterhuber

Conditionals as representative inferences 204
Robert van Rooij and Katrin Schulz

The Role of Perceiver in the Predictive Processing Framework 206
Elmarie Venter

Philosophy of Perception as Model-Building 207
Alfredo Vernazzani

Referential Intuitions are Still Problematic 209
Massimiliano Vignolo and Filippo Domaneschi

De jure Co-Reference: In Defence of Pointer Relations 211
Sara Vikesdal

The Development of Concepts of Mind and Mental States 213
Julia Wolf

Program Committee

Derek Ball	University of St Andrews
Hanoch Ben-Yami	Central European University
Corine Besson	University of Sussex
Paul Dekker	ILLC/University of Amsterdam
Kepa Korta	UPV-EHU
Beate Krickel	Ruhr-University Bochum
Albert Newen	Ruhr University Bochum
Peter Pagin	Stockholm University
Michele Palmira	University of Barcelona
Francois Recanati	Institut Jean Nicod, CNRS Paris
Pedro Santos	Universidade do Algarve
Matthias Unterhuber	Ruhr University Bochum
Heinrich Wansing	Ruhr University Bochum
Markus Werning	Ruhr University Bochum

Program

Thursday Morning (GA-Building, Ruhr University Bochum)

8:30	**Registration GA 04/42+44+45**								
9:15	**Welcome (HGA 10)**								
9:30	**1-0-0**	**Keynote: *Emma Borg & Tim Salomons* – The folk concept of pain is paradoxical: defending a polyeidic approach (HGA 10)**							
10:30	**Coffee Break/Registration GA 04/42+44+45**								
11:00	**2-1**	**Mind – Agency, Intentionality, and Mindreading (HGA 10)**	**2-2**	**Mind – Memory, Delusions, and Imagination (HGA 20)**	**2-3**	**Mind & Language – Concepts and Meaning (HGA30)**	**2-4**	**Language – Literal Meaning, Metaphor, and Irony (GAFO 03/252)**	**2-5**	**Language – Logical Form and Meaning (GA 04/187)**
11:00	2-1-1	*Alexander Stathopoulos:* Uncertain Progress	2-2-1	*Joan Casas Roma:* Towards a Shared Frame for Imaginative Episodes	2-3-1	C. *Strößner & G. Schurz:* Modification: Pragmatics and Constraints	2-4-1	*S. P. Reda & U. Sauerland:* Degrees of Analyticity	2-5-1	*Alexander Dinges:* A direction effect on taste predicates
11:40	2-1-2	*Andrew Sims:* Does the sense of agency include a phenomenology of freedom?	2-2-2	*Krzysztof Poslajko*: Is the debate about doxasticism merely verbal?	2-3-2	*T. Ciecierski & P. Grabarczyk:* Compositional Conceptual Role Semantics	2-4-2	*Sami Rissanen*: Literal meaning, idiolects, and language learning – the impossible equation	2-5-2	*Miguel Hoeltje:* Generic Disjunctive Syllogism
11:55	**Short break**									
12:00	2-1-3	*Philipp Huebl:* Forced Choices: Does accepting its consequences suffice for deeming a deed intentional?	2-2-3	*Fabrizio Calzavarini*: Language comprehension and visual imagery	2-3-3	*Matthias Unterhuber:* A Mechanistic Take on the Theory Theory of Concepts	2-4-3	*Andreas Heise:* Beyond paradoxes and bifurcations: Proposal for a unified theory of metaphor based on speaker's intention	2-5-3	*M. Spychalska, V. Haase, J. Kontinen & M. Werning:* Processing of affirmation and negation in contexts with unique and multiple alternatives: Evidence from event-related potentials
12:40	2-1-4	*R. Mirski & A. Gut:* The untenability of nativism in current research on mindreading	2-2-4	*Gregory Bochner:* Perceptual Experiences: Content, Situation, Mode	*2-3-4*	*Kristina Liefke*: Towards a Unified Semantics for Extensional and Cognitive Verbs	2-4-4	*Holden Härtl:* Ironic name reference as echoic mention – A pragmatic analysis based on empirical data	2-5-4	*Nick Tasker:* The Forms of Words
13:20	**Lunch break/Registration (GA 04/42)**									

Thursday Afternoon (GA-Building, Ruhr University Bochum)

14:30	**3-1**	**Mind –Introspection, Consciousness, and the Self (HGA 10)**	**3-2**	**Mind & Language – Concepts and Meaning II (HGA 20)**	**3-3**	**Language – Indexicals and Conditionals (HGA 30)**	**3-4**	**Language – Slurs and Lies (GA 04/187)**
14:30	3-1-1	*Alberto Barbieri:* A combination problem for Self-Representationalism?	3-2-1	*A. Schuster & L. Berio*: Mentalised prototype frames	3-3-1	*Jiwon Kim*: Too-Shifty Indexicals in Answering Machine Puzzle	3-4-1	*M. Jary & M. Kissine*: The semantic content of imperatives
15:10	…		3-2-2	*Mark Pinder*: A Puzzle for Semantics	3-3-2	*Stefan Rinner*: Chalmers on Indexicals	3-4-2	*Bianca Cepollaro*: The reclamation of slurs
15:25	**Coffee Break**							
15:55	3-1-3	*Anna Giustina*: Introspection without Judgment	3-2-3	*Joanna Pollock*: Context and communicative success	3-3-3	*Alexandru Radulescu*: How to Represent the Characters of Indexicals	3-4-3	*Merel Semeijn:* Fiction, lies and assertions
16:35	3-1-4	*Beate Krickel*: Are the Mental States Underlying Implicit Biases Unconscious? A Neo-Freudian Answer	3-2-4	*G. R. Löhr*: Abstract Concepts, Compositionality and The Contextualism-Invariantism Debate	3-3-4	*F. Del Prete & S. Zucchi*: Gender presuppositions in conditionals	…	
17:15	**4-0**	**Poster Session with Coffee (GAFO 03/252)**						
18:15	**5-0-0**	**Keynote: *Markus Werning & Erica Cosentino* – The Interaction of Bayesian Pragmatics and Lexical Semantics in Sentence Meaning Composition (HGA 10)**						
19:15	End							
20:15	**Conference Dinner at Strätlingshof**							

Friday Morning (Center for IT Security (ZITS), Bochum)

8:30	**Registration**						
9:30	**6-0-0**	**Keynote: *Albert Newen* – Cognition and Perception: Does higher-order background information influence our perceptual experience? (Görtz)**					
10:30	**Coffee Break/Registration**						
10:50	**7-1**	**Mind – Functionalism and Explanation (Görtz Ost)**	**7-2**	**Mind – Pain, Subjectivity, and the Self (Görtz West)**	**7-3**	**Language – Metaphysics (von Neumann Ost)**	**7-4**	**Language – Presuppositions and Speech Acts (von Neumann West)**
10:50	7-1-1	*Shahar Hechtlinger:* Why computational explanations are not mechanistic: An argument from idealization	7-2-1	*Thomas Raleigh:* How to talk about subjective experience?	7-3-1	*Mark Bowker:* Semantic Restrictivism	7-4-1	*Dirk Kindermann*: On the Availability of Presuppositions in Conversation
11:30	7-1-2	*Markus Pantsar*: Cognitive complexity and mathematical problem solving	7-2-1	*Abraham Sapien:* Unpleasant pain = Δ	7-3-2	*Maria Matuszkiewicz:* A commitment to haecceitism in Robert Stalnaker's solution to indexical beliefs	7-4-2	*Blaise Prentice-Davidson:* The Recognitive Function of Explicit Expressions of Belief
11:45	**Short break**						
11:50	7-1-3	*Antonio Ramos Diaz:* Kripke, Functionalism, and the Formal Indeterminacy of the Physical	7-2-3	*Carlota Serrahima*: My Body is the Subject's Body	7-3-3	*Giorgio Lando:* A Unified, Kaplanian Metaphysics for Linguistic Types	7-4-3	*Joanna Odrowaz Sypniewska*: Sub-sentential speech acts: a situated contextualist account
12:30	…		7-2-4	*Adam Bradley*: Pain, Multimodality, and Bodily Awareness	7-3-4	*Pierre Saint-Germier:* Hume's maxim with impossible worlds	7-4-4	*Laura Caponetto:* Doing and Undoing. Some Strategies to Backpedal on Our Speech Acts
13:10	**Lunch break/Registration**						

Friday Afternoon (Center for IT Security (ZITS), Bochum)

14:20	**8-1**	**Mind – Rationality (Görtz Ost)**	**8-2**	**Language & Mind – Content (Görtz West)**	**8-3**	**Language & Mind – Rationality and Beliefs (von Neumann Ost)**	**8-4**	**Language – Reference (von Neumann West)**
14:20	8-1-1	*Niemeck Maik:* De Se Skepticism and Rational Action	8-2-1	*Jussi Haukioja:* Deferring to Future Speakers	8-3-1	*Edward Elliott:* The Nature of the Fundamental Doxastic State	8-4-1	*Massimiliano Vignolo:* Referential Intuitions are Still Problematic
15:00	8-1-2	*Dacid Evan*: A Neurobiological Case for Animal Rationality	8-2-2	*P. De Brabanter & B. Leclercq:* Semantic externalism(s) and content transparency	8-3-2	*Joshua Matthews:* The Normative Role of Logic	8-4-2	*Katarzyna Kijania-Placek:* Polysemous names
15:15	**Coffee Break**							
15:35	…		8-2-3	*Gabriel Rabin:* How to Twin-Earth a Phenomenal Concept	8-3-3	*Markus Kneer*: Norms of Assertion: Empirical Data	8-4-3	*Sara Vikesdal:* De jure Co-Reference: In Defence of Pointer Relations
16:15	…		8-2-4	*Vitaly Kiryushchenko:* Bildung, Second Nature, and Conceptual Content in McDowell's "Mind and World"	8-3-4	*R. van Rooj & K. Schulz:* Conditionals as representative inferences	8-4-4	*Junhyo Lee:* On the Uniformity of Proper Names
16:55	**Coffee Break**							
17:25	**9-0-0**	**Keynote: *Nikola Kompa* – Language and embodiment – an evolutionary perspective (Görtz)**						
18:25	**10-0-0**	**Keynote: *Paul Egré & Cathal O'Madagain* – Concept Utility (Görtz)**						
19:25	End							

Saturday (Center for IT Security (ZITS), Bochum)

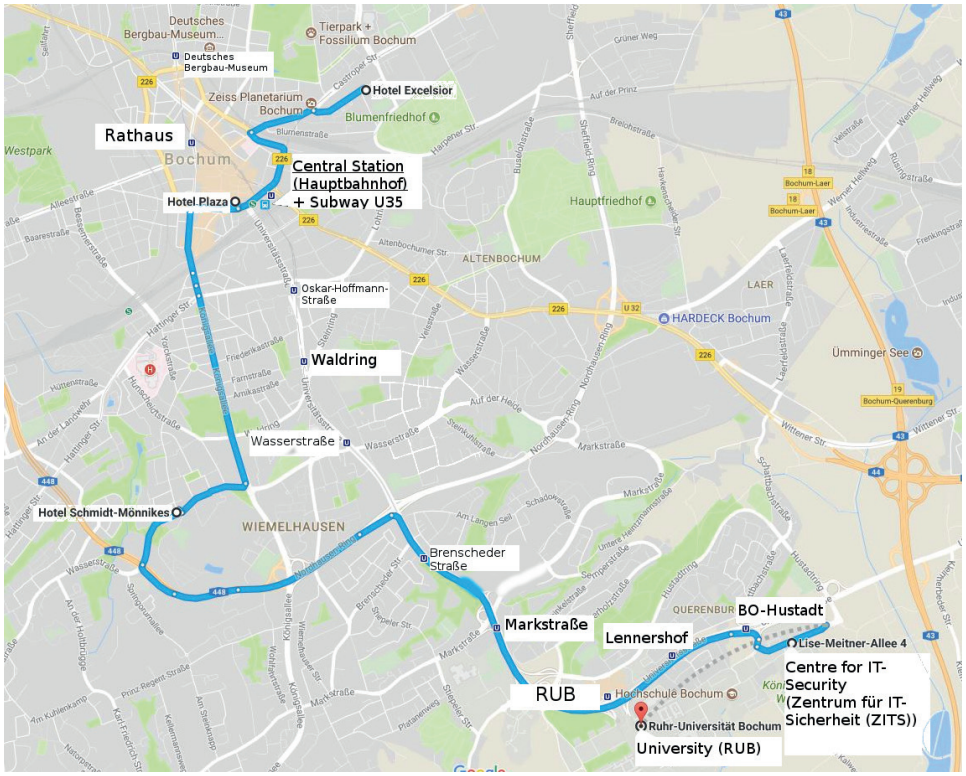
9:30	**11-0-0**	**Keynote: *Max Kölbel* – Perspectival Content and Expressing an Attitude vs Asserting that One Has It (Görtz)**					
10:30	**Coffee Break/Registration**						
11:00	**12-1**	**Mind – Rationality, Emotions, and the Self (Görtz Ost)**	**12-2**	**Mind – Perception (Görtz West)**	**12-3**	**Language & Mind – Semantics/Pragmatics Interface (v. Neumann Ost)**	**12-4**	**Language – Semantics (von Neumann West)**
11:00	12-1-1	*Kathryn Pendoly*: Recalcitrance Poses No Threat to Cognitivism	12-2-1	*Alfredo Vernazzani*: Philosophy of Perception as Model-Building	12-3-1	*Antonin Thuns:* Lexical files as encyclopedic files	12-4-1	*Giulia Felappi*: It's a wonderful world. Our language about the attitudes
11:40	12-1-2	*Chiu Yui Plato Tse:* Ascription and Agency of the First-Person	12-2-2	*A. Buccella:* Perceptual constancy and the role of perception in action	12-3-2	Jumbly Grindrod: QUDs and Context-Sensitivity	…	
11:55	**Coffee Break**							
12:25	12-1-3	*Sophie Keeling*: Confabulation and Rational Requirements for Self-Knowledge	12-2-3	*Giulia Martina*: What about the colours? Shoemaker on appearance properties	12-3-3	*Samuel Taylor:* Beyond the What is Said Debate: Towards a Psychologistic Theory of Successful Communication	12-4-3	*Dolf Rami:* Attitude verbs as reference-shifting operators
13:05	12-1-4	*Bart Geurts:* Making sense of self talk	12-2-4	*Gabriele Ferreti:* An Action-Based Theory of Picture Perception	12-3-4	*Maria Spychalska*: Neurocognitive evidence in the semantics-pragmatics interface debate: the case of implicatures	12-4-4	*Anna Drożdżowicz*: Do we hear meanings (and should we care whether we do)?
13:45	**Lunch Buffet with Coffee**				**13-0**	**PLM Board Meeting (Von Neumann West)**		
						D. Ball, F. Huoranszk, P. Pagin, M. Lipman, F. Recanati, M. Werning, C. Besson, P. Santos, M. Palmira: Reports on Trends in the Philosophy of Language and Mind from the Member Countries		
15:00	**14-0-0**	**Keynote: *Bence Nanay* – Seeing Things You Don't See (Görtz)**						
16:00	**15-0-0**	**Keynote: *Ophelia Deroy* – Are racial biases in perception truly automatic, and why does this matter? (Görtz)**						
17:00	End							

List of Poster Presentations

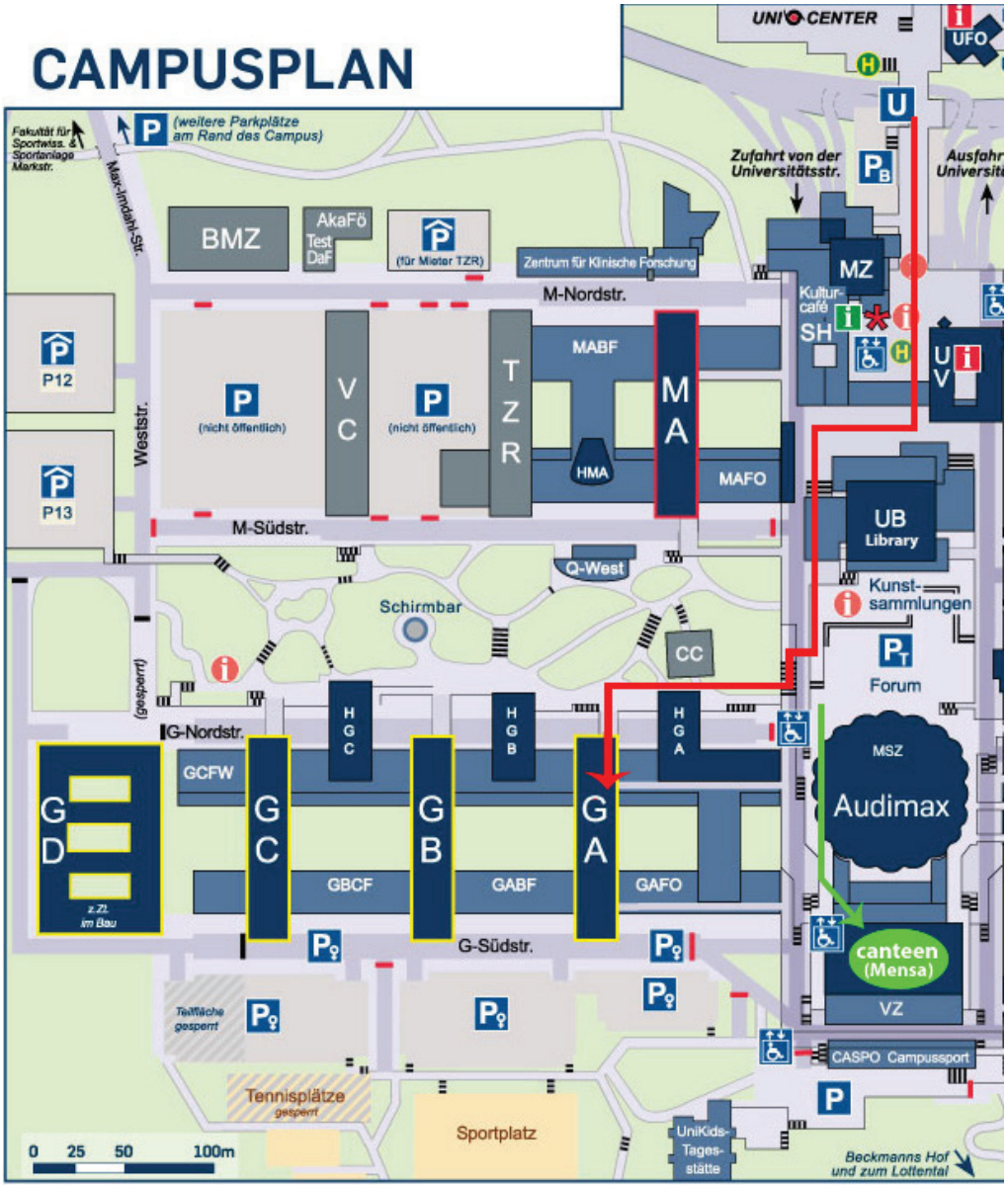
4-0-1	Keyvan	Alasti	Millikan's Teleological Theory of Content, Indeterminacy Problem and Temporal Externalism
4-0-2	Fabrizio	Calzavarini	Model-theoretic semantics and semantic competence
4-0-3	Roberta	Colonna Dahlman	A Classical Implicational Account of Semantic Presuppositions
4-0-4	Laura	Delgado	Names are Polyreferential
4-0-5	Lorenz	Demey	The Logical Geometry of Russell's Theory of Definite Descriptions
4-0-6	Matej	Drobňák	Open-textured and Dynamic Meaning as a Ground for a New Metasemantic Framework for Natural Languages
4-0-7	Mara	Floris	Language and Conceptual Development: a Philosophical Review of the Empirical Research
4-0-8	Simone	Gozzano	On the location of pain
4-0-9	Neil	Hamilton Fairley	Cause for Concern Searle's Background Theory and the Context Sensitivity of Thought
4-0-10	Matej	Kohar	Mechanisms and representations do not mix
4-0-11	Hsiulin	Ku	Making Sense of the Emergent Property: Association or Inference?
4-0-12	Armando	Lavalle Terrón	Multipropositionalism and Necessary A Posteriori Identity Statements
4-0-13	Jeonggyu	Lee	Against Predicativism of Names
4-0-14	Harald	Maurer	A New Metapher Illustrating the Vector Based Information Processing in Systemtheoretical Cognitive Neuroarchitectures in Connectionistic Cognitive Science
4-0-15	César	Meurer	Challenging Evolutionary Psychology with the eternalist view of the world
4-0-16	Maik	Niemeck	How to account for Subjective Character?
4-0-17	Erica	Onnis	Handling Concurrency through Consciousness
4-0-18	Fabrice	Pataut	Semantic Antirealism and the Self-Ascription of Attitudes
4-0-19	Michal	Polak	Two critical remarks to classical identity theory
4-0-20	Jiří	Raclavský	Predication, Meaning and Transparent Hyperintensional Logic
4-0-21	Stefan	Rinner	Naive Russellians and the Goldbach Puzzle
4-0-22	Astrid	Schomäcker	A problem with armchair arguments against explaining consciousness
4-0-23	Michał	Sikorski	The Acceptability and The Probability of the Indicative Conditionals
4-0-24	Maik	Sühr	Social Externalism and Ordinary Speech
4-0-25	Elmarie	Venter	The Role of Perceiver in the Predictive Processing Framework
4-0-26	Julia	Wolf	The Development of Concepts of Mind and Mental States

Maps and Directions

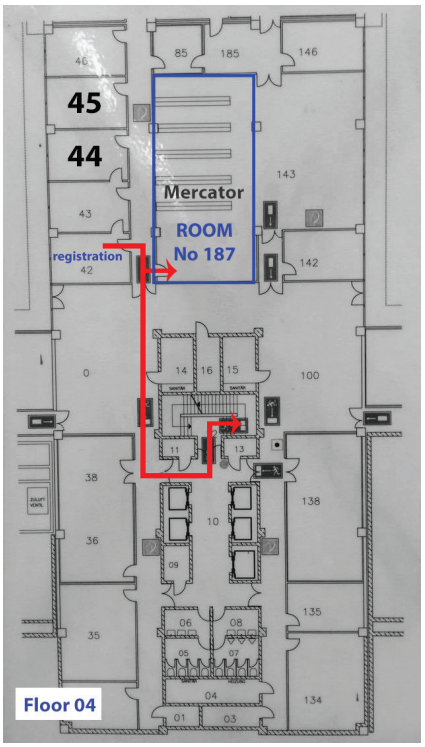
How to Get to Conference Venues: Ruhr University Bochum (RUB) and Center for IT Security, Bochum

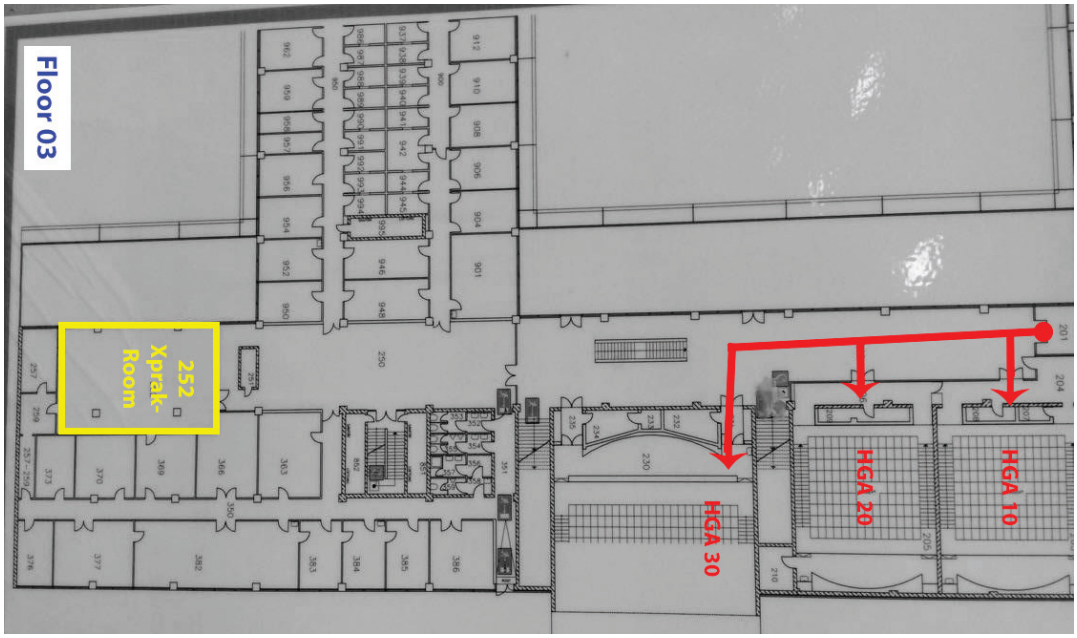


How to Get from the Subway Station to the Venue at the Ruhr University Bochum (RUB)



Floor Plan of the Conference Venue at the Ruhr University Bochum (RUB)

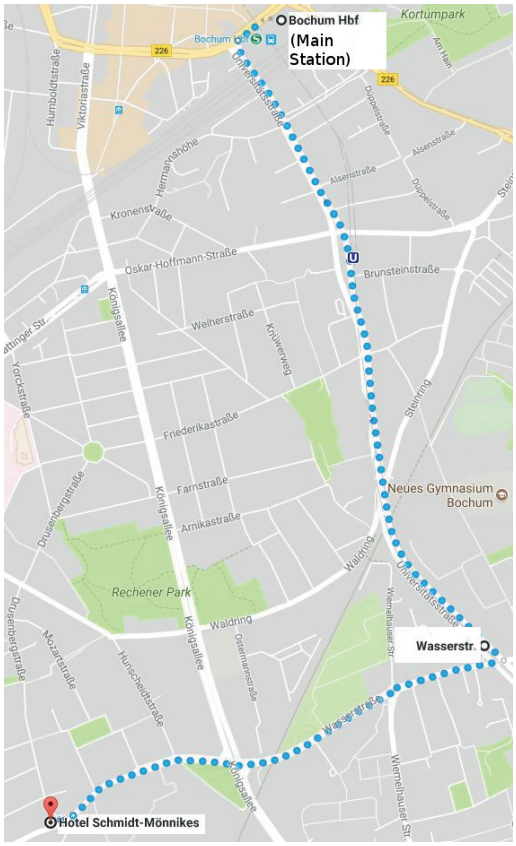




How to Get from the Subway Station to Center for IT Security (Venue)

Way from the Center for IT Security to the Canteen

From the Hotels to the Main Station (Change for Transfer to the Conference Venues)



Helpful information

Mobile phone numbers of
Christiane Dahl: +49 1575-6276389
Marius Markmann: +49 174 1663285
Nils Bremhorst: +49 1578 9009411

Police: 110
Fire emergency call: 112

Taxi companies
Taxi individuell: +49 234-61065900 or -901
Taxi Bednarz: 49+ 234-490024 or 26
Taxi soenmez & Toku: +49 234-7940824

Public transport
National Rail Company (Deutsche Bahn)
https://www.bahn.com/en/view/index.shtml

Regional public transport (also for Bochum)
https://www.vrr.de/en/fares-and-tickets/
info@vrr.de

Bochum public transport
http://www.bogestra.de/tickets-tarife.html
http://www.keine-ist-wie-mutti.de/
(Not available in English)

Invited Papers

The folk concept of pain is paradoxical: defending a polyeidic approach

Philosophers have commonly assumed that the ordinary view of pain sees it as a paradigm mental state: pains are held to be experiential states, exhibiting classic Cartesian characteristics (privileged first-person access, privacy, and incorrigibility). However the view of pain as a paradigm mental state is challenged by people's willingness to assign pains (non-brain-based) bodily locations: unlike most other mental states, pains are held to exist in arms, feet, etc. This recognition has led some as some (e.g. Hill 2005, Aydede 2005) to talk of the 'paradox of pain', whereby the folk notion of pain is inherently conflicted. Recently, Justin Sytsma and co-authors have argued against the paradox of pain view. Using the methods of experimental philosophy they argue that, contra the common philosophical assumption, the folk view of pain is of a bodily not a mental state (i.e. according to ordinary folk, pains exist in the world and not in the mind). This paper argues against Sytsma et al and in favour of the paradox: first, we re-examine the evidence Sytsma et al provide for their claims and raise some putative worries. Second, we note that even if folk do sometimes utilize a bodily notion of pain this is just what one would expect if, as the paradox view contends, the folk notion of pain is conflicted. Furthermore, we suggest that Sytsma et al's non-paradoxical, bodily view is unable to accommodate certain experimental findings concerning folk thinking about pain. Finally we conclude by outlining a direction for future research in developing an alternative, positive account – what we term the 'polyeidic approach' to the concept of pain – whereby 'pain' is viewed as expressing a complex concept, one which encompasses various divergent and potentially conflicting strands, and we suggest that certain surface phenomena surrounding pain and its communication could be usefully viewed through the lens of the polyeidic approach.

Are racial biases in perception truly automatic, and why does this matter?

The fact that people tend to perceive the same face as brighter or darker depending on its morphology is one of the most discussed cases of 'automatic categorical bias' in psychology and philosophy. While the manifestation of this effect has led to various interpretations, either as a canonical case of cognitive penetration or a mere bottom-up effect, everyone accepts that this process is fully involuntary and unconscious. Here, I will challenge that this diagnosis follows from the available evidence, and present new arguments to qualify the assumed automaticity of the effect. This has implications not only for the way we assign responsibility to people who act on biased percepts, but also to better understand the nature of perceptual biases.

Concept Utility

We often find ourselves revising our understanding of some concept, as witnessed recently in the case of the astronomical concept "planet". Such revisions are typically accompanied by the intuition that we have come to a better understanding of the concept we were working with. Standard theories of concepts, however, make this phenomenon difficult to account for. This paper considers a new way of thinking about concepts, derived from the literature on epistemic utility, allowing us to explain how such conceptual revision can come about. What makes our beliefs useful is not just their plausibility, but also their informativeness – how much they tell us about the world. Given that concepts play an integral role in the formation of beliefs, some conceptual schemes will be more likely to yield plausible and informative beliefs than others. We consider two analogous properties for categories: how inclusive a category is, and how homogeneous it is. Both measures vary in inverse direction from one another, hence their product gives us a notion of concept utility, allowing us to show that some taxonomies are measurably better than others. We explore as a case study the recent revision of the concept "planet" by the International Astronomical Union, and discuss the aspects in which an approach of concepts based on epistemic utility relates to natural kind externalist theories of conceptual meaning.

Language and embodiment – an evolutionary perspective

Within the paradigm of Situated Cognition, the claim that cognition is embodied has gained some prominence. My concern will be with (so-called) embodied approaches to language comprehension. In the first part of the talk, I will sketch the basic idea and examine some of the evidence that has been brought forth in favor of the claim that language comprehension is embodied. In the second part of the talk, some problems will be addressed; I will claim that one of the major problems embodied approaches face is that they greatly underestimate the cognitive functions language fulfills; as a result, they employ an impoverished notion of language comprehension. In the third part of the talk, I will adopt an evolutionary perspective; thereby, two of the main cognitive benefits of language will come into view. The reminder of the talk will be devoted to arguing for the cognitive indispensability of non-embodied representations and a more refined notion of language comprehension.

Perspectival Content and Expressing an Attitude vs Asserting that One Has It

For Ayer's emotivism to work, he needs to distinguish between expressing an attitude and asserting that one has it. I shall defend Ayer's distinction. However, an analogous distinction seems to get the relativist into trouble: there seems to be a clear difference between expressing one's preference for Vanilla over chocolate by saying "Vanilla is better than chocolate.", and asserting that one has that attitude, by saying, for example, "I prefer Vanilla to chocolate". However, which perspectival content should the relativist assign to utterances of these two sentences, and to the mental states thereby expressed? De se-type considerations suggest that the content of the latter must also be perspectival. But how does it differ from the other? I shall attempt to solve this puzzle.

Seeing Things You Don't See

This talk is about the importance of mental imagery in general and of multimodal mental imagery in particular in our everyday perception. I bring together empirical findings about multimodal perception and empirical findings about (visual, auditory, tactile) mental imagery I order to argue that multimodal mental imagery, understood as perceptual processing in one sense modality that is triggered by sensory stimulation in another sense modality is a crucial element of almost all instances of everyday perception, which has wider implications to philosophy of perception and beyond, to epistemological questions about whether we can trust our senses. Focusing on multimodal mental imagery can help us to understand a number of puzzling perceptual phenomena, like sensory substitution and synaesthesia. Further, manipulating mental imagery has recently become an important clinical procedure in various branches of psychiatry as well as in counteracting implicit bias – using multimodal mental imagery rather than voluntarily and consciously conjured up mental imagery can lead to real progress in these experimental paradigms.

Cognition and Perception: Does higher-order background information influence our perceptual experience?

To what extent is our perceptual experience influenced by higher cognitive phenomena like beliefs, desires, concepts, templates? Given recent arguments against the possibility of cognitive penetration, I present strong evidence against the impenetrability claims. The weak impenetrability claim cannot account for (1) extensive structural feedback organization of the brain, (2) temporally very early feedback loops and (3) functional top-down processes modulating early visual processes by category-specific information. The strong impenetrability claim could incorporate these data by widening the ‘‘perceptual module” such that it includes rich but still internal processing in a very large perceptual module. I argue that this latter view leads to an implausible version of a module. Therefore, we have to accept cognitive penetration of our perceptual experience as the best theoretical account so far given the available empirical evidence. Finally, consequences for the relation between perception and cognition are discussed.

The Interaction of Bayesian Pragmatics and Lexical Semantics in Sentence Meaning Composition

We contrast two views of how contextual influence on sentence meaning composition can be explained. The Semantic Similarity View maintains that discourse context affects the semantic composition of a sentence only by means of the semantic similarities between the words in the discourse context and in the sentence. Pragmatic factors of the discourse context do not effect sentence meaning composition. This view is thus consistent with a rigorous interpretation of the principle of compositionality as favored by semantic minimalists and according to which the truth-evaluable semantic content of a sentence is fully determined by its syntactic structure and lexical content where only a small number of lexical items (e.g., indexicals and anaphors) allow for a context-sensitive meaning contribution (Borg, 2004, 2012). The Free Pragmatic View, in contrast, challenges a rigorous notion of compositionality and maintains that pragmatic aspects of the discourse context can affect sentence meaning composition directly: Pragmatic discourse information can modulate meanings at every node in the semantic composition tree (Recanati, 2010, 2012). This phenomenon can be quantitatively modelled by Bayesian Pragmatics (Frank & Goodman, 2012). We introduce a Predictive Completion Task in which the hearer at every moment in a communicative situation has to generate a probabilistic prediction about how a discourse being uttered by the speaker is continued (Cosentino, Baggio, Kontinen, & Werning, 2017; Werning & Cosentino, 2017). We test the predictions of the two views in EEG using the well-established observation that the conditional probability of a word given a context is negatively correlated with the amplitude of its N400 component (Kuperberg & Jaeger, 2016). The results speak strongly in favor of the Free Pragmatic View and against the Semantic Similarity View.

References

Borg, E. (2004). *Minimal Semantics*. Oxford: Clarendon Press.

Borg, E. (2012). *Pursuing Meaning*. Oxford University Press.

Cosentino, E., Baggio, G., Kontinen, J., & Werning, M. (2017). The time-course of sentence meaning composition. N400 effects of the interaction between context-induced and lexically stored affordances. *Frontiers in Psychology*, *8*(818). http://doi.org/10.3389/fpsyg.2017.00813

Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, *336*(6084), 998–998. http://doi.org/10.1126/science.1218633

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language Cognition & Neuroscience*, *31*(1), 32–59. http://doi.org/10.1080/23273798.2015.1102299

Recanati, F. (2010). *Truth-Conditional Pragmatics*. Oxford University Press.

Recanati, F. (2012). Compositionality, Semantic Flexibility, and Context-Dependence. In M. Werning, W. Hinzen, & E. Machery (Eds.), *Oxford Handbook of Compositionality* (pp. 175–191). Oxford: Oxford University Press.

Werning, M., & Cosentino, E. (2017). The interaction of Bayesian pragmatics and lexical semantics in linguistic interpretation: Using event-related potentials to investigate hearers' probabilistic predictions. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 3504–3509). Austin, TX: Cognitive Science Society.

Contributed Papers

Millikan's Teleological Theory of Content, Indeterminacy Problem and Temporal Externalism

According to Millikan's theory, the content of a mental representation is the certain condition under which a consumer system had been selected. But the view is subject to a problem i.e. Indeterminacy problem; It is supposed that the system is selected for its unique biological function, but in some cases a trait is selected, not for one certain function but, for two (or more) different functions which all have taken part in selecting the consumer system. For example, it is because of (1) the ability of detecting black small objects, that frogs have (2) the ability of hunting foods. If the former was the very function that the digestive system selected for then the content of frogs' brain states would be flies and so ingesting little bullets should be counted as misrepresentation. But taking the latter one as main function would make all small moving objects as a part of the content and so hunting small bullets would not be misrepresentation. So contents described by teleosematics theory are indeterminate.

I suggest that by defending a version of semantic externalism called "Temporal Externalism" the indeterminacy problem can be rationally explained. Semantic externalism is the idea that mental contents do not supervene on mental intrinsic properties and some extrinsic properties related mental contents to social or environmental factors also take part in determining contents. Recently Henry Jackman (1999) and Tom Stoneham (2003) defended the idea of "Temporal Externalism" according to which not only social and environmental factors but also accidental future events contribute to fixing contents. Although Millikan's account is in accord with the general idea of semantic externalism, understanding participation of future events in determining contents is not as clear as participation of past events. Past events have made the conditions under which a system is selected, but it is not obvious how the future events can be involved in determination of contents. Thus I try to shed the light on the role of future events.

The paper consists of two arguments. Firstly I argue that regarding Millikan's account, future accidental events play role in fixing contents. Secondly I argue that if it is so then indeterminacy problem can be solved.

The First argument consists of two parts. The part one is dedicated to argue that there may be some samples which have not been involved in selecting a consumer system but intuitively we count them as a part of the content. For example, imagine a species of insects which has not been in frog's evolutionary environment so far and recently immigrates to frog's environment. If they are eatable by frogs then just because of not participating in selecting frog's digestive system, hunting them by frogs will be not counted as misrepresentation. Nevertheless, this case is not a problem for the theory since the issue can be explained by appealing to Millikan's difference between derived functions and relational functions. This difference originally has been coined for describing how a new mental representation that have not had any role in evolutionary history. According to Millikan, these cases also have biological functions and although they are not selected during evolutionary process, they are derived from other mechanism whose biological functions are producing other mechanisms having derived functions. For example, the mechanism for recognizing President Obama is not the very mechanism which is selected in an evolutionary process.

This mechanism has a function which is derived from the mechanism of recognizing faces. The mechanism of learning is also a relational function but the mechanism of learning how to find new foods is a derived

function. So every condition (including the migrant insects) which in principle increases fitness of the trait should be counted as a part of its content.

In part two, I try to show that the argument of part one is not enough for explaining indeterminacy problem, since it might be possible that something which is not actually frog's food, (not to mention selections) will be the frog's food in future because of accidental events. For clarifying this case, imagine a situation in which insects of another species that frogs cannot distinguish them from ordinary flies, immigrate to frogs' environment. The new migrant insects are not the best frog's food and so ingesting them produces negative effects on their surviving or proliferation. In this case, there will be two possibilities: If the insects have low population then ingesting them will be counted as misrepresentation, but if the population is not low then new insects will become the frog's main food source. In latter situation it is hard to take hunting the insects as misrepresentation.

The described scenario is similar to the Jackman's scenarios which encouraged him to introduce the concept of Temporal Externalism. According to him, not only the environmental and social factors but also future accidental events are constitutive of mental contents. By adding the idea of temporal externalism to Millikan's teleosemantics theory, we can say: not only the past conditions which have caused a consumer system naturally selected but also the future conditions which will help the system to survive are also constitutive of contents.

The second argument is devoted to show how this reading of externalism, can be used for explaining indeterminacy problem. If the system has been selected for more than one function then the sufficient condition for determining content will be made by future events. For clarification, imagine a possible condition which has not taken part in selecting a consumer systems so far, but in future it will play role for retaining the system. So, in addition to flies, which had been actually involved in selection, every other future condition under which the successfulness of former function(s), e.g. detecting small black objects, conduces to the successfulness of the latter function(s), e.g. hunting foods, will be a part of content. If it is so then many (but not all) of small black insects in future under certain accidental circumstances have the ability of being a part of the content. And other future conditions under which successfulness of the former function does not conduce to successfulness of the latter one is a sort of misrepresentation. So hunting the same insect under different future circumstance may be counted as misrepresentations.

A combination problem for Self-Representationalism?

In the context of the reductive theories of consciousness, self-representational (or neo-Brentanian) approaches have recently become greatly developed and popular. Self-representationalism (SR) is a one-level theory that promises to account for the subjective character of experience (the *for-me-ness*) in representational terms without encountering the problems of higher-order theories (HOT). According to SR, a mental state exhibits *for-me-ness* (and hence, is conscious) if and only if it represents itself in the right way. Kriegel is one of the most prominent defenders of SR. He spent a lot of energies attempting to show the phenomenological suitability of SR (Kriegel 2003, 2009a, 2009b) and its ability to naturalize phenomenal character (Kriegel 2005, 2006, 2009b, 2011). In this paper, I will focus mainly on Kriegel's model (for other versions of SR see Thomasson 2000; Hossack 2002; Kriegel&Williford 2006). As a metaphysical framework, SR can be defined, at least, as the conjunction of the following three claims:

i. *Subjectivism about phenomenal character:* at least part of the mystery of phenomenal consciousness relies on for-me-ness (subjective character).
ii. *(Self-)Representationalism about for-me-ness*: we can account for the subjective character in terms of suitable (self-)representation. [the definitional claim].
iii. *Priority of state consciousness*: our ordinary experience (creature consciousness) conceptually depends on being in conscious mental states (state consciousness).

My aim is to argue that SR, alone, is not able to account for the phenomenology of conscious experience (especially the *for-me-ness* of experience), and that this result can be achieved without rejecting any of the claims mentioned above. Indeed, in the last years, the critics of SR have tried to deny at least one of these claims. Some have rejected (i): this is the attitude of those, like strong representationalists (Drestke 1995, Tye 2000), who does not acknowledge the *for-me-ness*¸ a minimal form of self-awareness, as a fundamental feature of phenomenal consciousness. However, most of them have rejected (ii). This can be achieved in two ways: (a) claiming that any "self-representation" leads to infinitive regress (Zahavi 1999), or (b) claiming that the self-awareness at stake in the subjective character of experience is essentially a non-reflexive relation, a form of non-object-consciousness (Levine 2006, 2010; Zahavi 2004). More precisely, Levine (2006) develops what Kriegel (2011) calls the "just more representation" objection, according to which representations endowed with subjective character are of a categorically different kind from other representation.

Although objections to (ii) constitute a hard problem for the naturalizing strategy of SR, a self-representationalist might resist embracing a weakened form of SR, or denying that the gap between self-awareness and "self-representation" should be epistemically transparent (see Kriegel 2011). In this way, the objections to (ii) weaken the naturalizing strategy of SR, not SR in and of itself. However, we can question SR without rejecting any of the substantial claims but only based on the phenomenology that the model gives us back. I will argue that, if we start looking at SR with a bottom-up analysis (from mental states to phenomenal consciousness *via* "self-representation"), instead of a classical top-down analysis (from phenomenal consciousness to mental state representation *via* self-awareness), the model leads to a sort of combination problem (or, if you prefer, a sort of binding problem). This will lead us to a crossroads: or (1) we reject one of the three substantial claims to accommodate SR with phenomenology of our ordinary experience, or (2) we leave SR untouched, accepting its phenomenological consequences. In both case SR comes with important cost.

Now I will briefly describe my strategy. Firstly, I will extend the SR definition with two uncontroversial (especially for a reductive philosopher) sub-claims: (iv) Mind can have distinct mental states at the same time *t* (some of them conscious and other non-conscious); (v) Mental states (at least non-conscious mental states) can exist relatively unconnected to each other. Secondly, I will argue that, according to SR, and its definition of *for-me-ness*, it is possible to conceive a situation S in which, given two non-conscious mental states at time *t*, they become conscious representing themselves in me, without being *for me*. In this scenario, we shall have two experiences in me with two different perspectives: so, a mental state seems for itself (*for-itself*-ness), not for me (*for-me-ness*). From a phenomenological perspective, it sounds strange; it appears to us that the experience comes *phenomenally unified* (Bayne 2010; Bayne & Chalmers 2003), a fact which is revealed in the experience as much as the *for-me-ness*. But why, given SR, should we find *for-me-ness* rather than *for-*

itself-ness? SR tells us nothing of the fact that distinct mental states could represent themselves at the same time. I will call this issue the combination problem for SR.

References

Bayne, T. & Chalmers, J. D. (2003). What is the Unity of Consciousness. In A. Cleeremans (ed.), *The Unity of Consciousness*. Oxford: Oxford University Press.

Bayne, T. (2010). *The Unity of Consciousness*. New York: Oxford University Press.

Drestke, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Hossack, K. (2002). Self-knowledge and consciousness. *Proceeding of the Aristotelian Society*, 102(2), pp.163-181.

Kriegel, U. & Williford, K. (2006). *Self-Representational Approaches to Consciousness*. Cambridge MA: MIT Press.

Kriegel, U. (2003). Consciousness as Intransitive Self-Consciousness: Two Views and an Argument. *Canadian Journal of Philosophy*, 33, pp.103-132.

Kriegel, U. (2005). Naturalizing Subjective Character. *Philosophy and Phenomenological Research*, *71*(1). pp. 23-57.

Kriegel, U. (2006). The Same-Order Monitoring Theory of Consciousness. In U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness*. Cambridge MA: MIT Press, pp. 143-170.

Kriegel, U. (2009a). Self-Representationalism and Phenomenology. *Philosophical Studies*, 143, pp.375-381.

Kriegel, U. (2009b). *Subjective Consciousness*. New York: Oxford University Press.

Kriegel, U. (2011). Self-Representationalism and the Explanatory Gap. In J. Liu and J. Perry (eds.). *Consciousness and the Self: New Essays*. Cambridge and New York: Cambridge University Press, pp. 51-75.

Levine, J. (2006). Awareness and (Self-)Representation. In U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness*. Cambridge MA: MIT Press, pp. 173-197.

Levine, J. (2010). Phenomenal Experience: a Cartesian Theater Revival. *Philosophical Issues*, *20*(1), 209–225.

Thomasson, A. L. (2000). After Brentano: A one-level theory of consciousness. *European Journal of Philosophy*, 8(2), pp.190-209.

Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.

Zahavi, D. (1999). *Self-awareness and Alterity. A phenomenological investigation*. Evanston: Northwestern University Press.

Zahavi, D. (2004). Back to Brentano. *Journal of Consciousness Studies*, *11*(10–11), 66–87.

Perceptual Experiences: Content, Situation, Mode

I see that this flower there is red. What is the representational content of my perceptual experience? In this talk, I want to motivate an original answer to this question. This will draw on the analyses due to Searle (1983, *Intentionality*) and Recanati (2007, *Perspectival Thought*).

Searle holds the following. My experience has two components: a mode (perception) and a content (what is represented and must be the case if the experience is to be veridical). I am seeing a particular object, the flower o_1, yet the content of my experience does not directly feature o_1 itself. Instead, it involves a causal-reflexive, descriptive condition, imposed by the perceptual mode:

1. Searle's analysis: VIS EXP (that there is a red flower there and that there is a red flower there is causing this visual experience)

On this analysis, the conditions of satisfaction of a perceptual experience and the content of that experience are one and the same thing. For Searle, generally speaking, the conditions of satisfaction of a mental state or a speech act specify (together with a direction of fit) what must be the case for the state or the act to count as satisfied. In non-reflexive cases (e.g. belief, assertion), the conditions of satisfaction do not mention the state or the act itself. In reflexive cases (e.g. intention, perception), they do. Searle (in this, followed by Recanati or Bach 2007, *Searle against the World*) holds that, in order for my perceptual experience to be satisfied, it must have been caused by the fact it represents. Then, given his identification of content and conditions of satisfaction, Searle must conclude that the causal-reflexive condition of satisfaction of my perceptual experience is part of its content.

Like Bach and Burge (1991, *Vision and Intentional Content*), Recanati criticises this result. He argues that Searle failed to disentangle content and conditions of satisfaction. For Recanati, the (relative, narrow) "content" of a perceptual experience is the scene explicitly represented from the perspective of the subject. Although they feature in its conditions of satisfaction, the subject, the experience, and the causal relations to the object, are not part of its "content" in this narrow sense:

2. Recanati's analysis: VIS EXP (that there is a red flower there)

By itself, this content has incomplete truth-conditions: it is not capable of being true or false. To complete it, we need a situation of evaluation, specifying a time and the subject whose experience would be caused by the fact that there is a red flower there. The pair <content, situation> will yield complete (absolute, wide) conditions of satisfaction for the experience. On Recanati's view, the situation is determined by the mode. Perception is a reflexive mode imposing "the perceptual situation of the subject" (the context of the experience) as the relevant situation. On that analysis, inspired by Lewis (1979, *Attitudes* De Se *and* De Dicto), the situation of evaluation corresponds to a world centred on the subject and the time of her experience.

So whereas (i) Searle used *one* notion of ("Intentional") content, (ii) Recanati invokes *two* notions: the (relative) content of the experience (that there is a red flower there), and the (absolute) conditions of satisfaction of the experience, the latter being jointly determined by its content and the mode (that there is a red flower there and that there is a red flower there is causing *my current* experience). (iii) I will argue for a third view, according to which the (representational) content of a perceptual experience (a property) and its situation of evaluation (an object) jointly determine its conditions of satisfaction; the situation of evaluation is not determined (reflexively) by the mode (*pace* Bach and Recanati); and the reflexive condition is not part of the conditions of satisfaction of the experience, but only of its conditions of formation.

On the analysis I want to promote, the content of my perceptual experience (considered in isolation from my other attitudes at the same time) is just a property:

3. Alternative analysis: VIS EXP (being red)

This content must be evaluated at the world centred on the *object* of my experience, $<w_1, t_1, o_1>$. The property of being red and that centred world together determine the *absolute and referential* conditions of satisfaction of my visual experience. The experience gives a veridical representation of my environment iff this flower is red—*full stop*. More precisely, it is veridical iff $<w_1, t_1, o_1> \in \{<w, t, o>: o \text{ is red at } t \text{ in } w\}$. This condition is independent of the mode: a belief, a desire, a hope, etc. might have the same condition of satisfaction. The mode and the reflexive condition it imposes on the experience belong to a different level: I could not (be said to) *see* that this flower is red if the fact that it is red had not caused my experience. Compare: I could (be said to) *believe* (*de re*) that this flower is red even if the fact that it is red had not caused my belief (or even if it is not red). What has to be the case for what I see to be veridical is not identical to what has to be the case for me to see it. The former requires only that *this flower* be red, but the latter requires further that *my experience* be caused by the fulfilment of the former requirement. As perceptual reports with a *that*-clause are factive, the claim that I perceive that this flower is red presupposes that this flower is red.

The alternative analysis I advocate distinguishes the following levels. **(1) The conditions of satisfaction** of my perceptual experience, involving: (a) the incomplete-relative conditions of satisfaction of my experience (= *its narrow content*, the property of being red, which would remain the same even if a qualitatively indistinguishable and numerically distinct flower o_2 had caused it); (b) the situation of evaluation of my experience (= *its object*, the centred world $<w_1, t_1, o_1>$); (c) the complete-absolute conditions of satisfaction of my experience (= *its wide content*, the pair consisting of the property of being red and $<w_1, t_1, o_1>$, featuring the object but not the subject or the experience). **(2) The conditions of formation** of my perceptual experience, jointly determined by its complete-absolute conditions of satisfaction *and reflexive conditions*, which further demand that there be a causal relation between some facts about the context of my experience (= *its subject*, the centred world $<w_1, t_1, me>$) and the fact that its complete-absolute conditions of satisfaction are fulfilled.

Semantic Restrictivism

The theory of communication has traditionally centred around the *determinist* approach. Speakers are taken to express unique propositions, determined by the denotations of the terms they utter, their compositional structure and the contexts in which they speak. This paper presents an alternative: the *restrictivist* approach, according to which the role of utterance is not to express a unique propositional semantic content, but to restrict the various propositional interpretations of the uttered sentence to the point at which the speaker's meaning is made clear. This restrictivist view will be deployed to deal with two significant problems for the determinist approach: *lexical* underdetermination and *compositional* underdetermination.

The problem of lexical underdetermination is often associated with Charles Travis. The proposition expressed by an utterance may be underdetermined by the lexical meanings of the words uttered. Consider the case in which Pia has a Japanese maple with red leaves which she has painted green. Is the sentence 'Pia's maple is green' true or false? Intuitions vary depending on the context in which the sentence is used. If spoken to a photographer looking for a green subject, the sentence appears to be true. If spoken to a scientist interested in the chemistry of green leaves, it appears to be false. The most common response manifest in the literature is the *contextualist* approach, which takes the proposition expressed by the sentence to be sensitive to context. When speaking to the photographer, for example, the sentence is taken to express the proposition that Pia's maple is green *on the surface* but when speaking to the scientist the sentence is taken to express the proposition that Pia's maple is *naturally* green. This response fails to address the problem, however. If the sentence expresses a proposition in conversation with the scientist, the it should take a truth value at any possible world. Yet it is unclear whether the sentence is true or false if the colour of Pia's maple is the result of genetic engineering.

Compositional underdetermination occurs when the structure of an utterance is incomplete. This form of underdetermination is most clearly exhibited by cases of nonsentential assertion, as when a verb or noun phrase is uttered in isolation. Past accounts of nonsentential assertion generally fall into two categories: syntactic, semantic, and pragmatic accounts. Respectively, these views interpret nonsentential assertions by appealing to additional linguistic material drawn from the context, by shifting the semantic type of subsentential expressions when interpreted in isolation, or by supplementing the object or property linguistically represented with some non-linguistic content.

What these three approaches all have in common is the assumption the assumption that successful communication requires a nonsentential assertion to express - whether semantically or pragmatically - a unique proposition. This assumption is questioned by Ray Buchanan, who notes that there are often various propositions that can be derived from a nonsentential assertion. If it is common ground that the person in the door way is Barbara Partee, then an utterance of 'A famous linguist' may be taken to express the proposition that the person in the doorway is a famous linguist, or the proposition

that Barbara Partee is a famous linguist. Buchanan concludes that nonsentential utterances do not communicate propositions.

The restrictivist view offers an alternative to the syntactic, semantic, and pragmatic approaches to compositional underdetermination on the one hand and Buchanan's denial of propositional communication on the other. While Buchanan is correct that nonsentential assertions do not express unique propositions, they will make the speaker's meaning clear as long as all contextually viable interpretations make the same contribution to the common ground, which may be represented by a set of worlds. If it is common ground that the person in the door way is Barbara Partee, then an utterance of 'A famous linguist' cannot be taken to express either the proposition that the person in the doorway is a famous linguist, or the proposition that Barbara Partee is a famous linguist, but each of these propositions will have the same effect on the common ground: to eliminate all worlds at which it is not the case that Barbara Partee, the person in the doorway, is a famous linguist.

This approach to nonsentential assertion suggests a novel approach to generics such as 'Birds fly' or 'Women are prone to anxiety'. Rather than trying to explain how a unique proposition is expressed by a generic in context, we might characterise generics as incomplete tripartite structures consisting of a restrictor and scope without any quantifier. This approach promises to explain one of the characteristic features of generics: resistance to counterexample. As no quantifier is made explicit, the user of a generic can deny their commitment to any particular quantifier on an ad hoc basis.

On the restrictivist account of lexical underdetermination, predicates are not viewed as expressing unique properties. As noted by the radical contextualist, there are at least two different propositions consistent with the meaning of the sentence uttered, yet these propositions have different effects on an unstructured common ground, given that one is true and the other false at the actual world. The solution is to impose structure on the common ground. Following recent work on the semantics of questions, we can view the context as partitioned according to the question under discussion (QUD). If the QUD is whether Pia's maple will make a good subject for the photographer, we have two answers: YES and NO. The proposition that Pia's maple is green on the outside eliminates all worlds at which Pia's maple is not a good subject for the photographer, while the proposition that her maple is naturally green fails to eliminate either answer to the QUD, being consistent with worlds at which it is painted red. Given that the speaker intends to provide an answer to the QUD, only interpretations that provide an answer are contextually viable. As the interpretations that provides an answer to the QUD is true at the actual world, we explain the intuition that the utterance is true. Note that there may be other interpretations compatible with the meaning of the sentence uttered that eliminate the same worlds from the context. If we know that Pia's maple is painted, the interpretation on which Pia's maple is painted green and the interpretation on which Pia's maple is naturally green or painted green will eliminate the same worlds from the context. The QUD doesn't serve to determine a unique propositional content, but to further restrict the interpretations of the sentence uttered.

Pain, Multimodality, and Bodily Awareness

Word Count: 485, excluding references

1 Abstract

Bodily awareness is the conscious awareness we have of our own bodies 'from the inside.'[1] Much recent work on bodily awareness has emphasized its *multimodal* character, or the fact that our sense of bodily awareness derives from a heterogeneous set of sources of bodily information including proprioception, kinesthesia, and vision.[2] Often neglected in such discussions, however, is the role that pain and other bodily sensations play in providing for our bodily self-awareness.[3] In this talk, my goal is to correct this oversight by drawing upon empirical results which demonstrate that pain experience plays a constitutive role in bodily awareness.

The basic fact that pain experience is a form of bodily awareness is uncontroversial: we typically experience pains in regions of our body. But the exact nature of the relationship between pain experience and other forms of bodily awareness is far from clear. To clarify, I distinguish between the *Template Model* and the *Constitutive Model* of the relationship between pain experience and bodily awareness. On the Template Model, pains and other bodily sensations are a derivative form of bodily awareness, 'filling in' portions of an antecendently given 'body image.' By contrast, on the Constitutive Model the experience of bodily pain is a non-derivative form of bodily awareness which, while ordinarily integrated with other forms of bodily information, can function independently of these other sources to make us aware of parts of our body.

To settle the question I cite three lines of evidence which favor the Constitutive Model over the Template Model. First, I use the under-discussed case of *Complex Regional Pain Syndrome* (CRPS) to show that the experience of pain can alter the felt limits of one's own body.[4] Second, I use the case of *mechanoceptive deafferntation* to show that one can experience pain in body parts despite

[1] Vignemont (2015) Bermudez (2011)
[2] Vignemont (2014) O'Callaghan (2015) Eilan, Marcel, and Bermudez (1995)
[3] Though see Martin (1993) Martin (1995) and Brewer (1995)
[4] Lotze and Moseley (2007)

a complete lack of proprioception or tactile sense.[5] Third, I cite cases of *extra-somatic sensation* to show that one can experience bodily sensations outside of the felt boundaries of one's own body as established by other forms of bodily awareness.[6] Taken together, these findings show that feeling pain in a body part is an intrinsic form of bodily self-awareness which can function independently of other forms of bodily awareness.

I end by examining the implications of these findings for understanding the nature of pain awareness. I argue that acceptance of the Constitutive View motivates rejecting the commonly-made claim that awareness of bodily pains is introspective in character.[7] Since introspective awareness is awareness of one's mental states, while pain awareness is a form of bodily awarness, it follows that pain awareness is no more a form of introspection than is proprioception or kinesthesia. Thus adoption of the Constitutive Model tends to support views which treat pain as a bodily condition, rather than a mental state of the subject.

[5]Cole (1991) p. 132 Gallagher (2005) Ch. 2
[6]Bekesy (1967) p. 220-6
[7]Aydede (n.d.)

References

Aydede, Murat. n.d. "Pain: Perception or Introspection?" In *The Routledge Handbook of Philosophy of Pain*, edited by Jennifer Corns. Routledge.

Bekesy, Georg von. 1967. *Sensory Inhibition*. Princeton University Press.

Bermudez, Jose Luis. 2011. "Bodily Awareness and Self-Consciousness." In *The Oxford Handbook of the Self*, edited by Shaun Gallagher, 157–79. OUP.

Brewer, Bill. 1995. "Bodily Awareness and the Self." In *The Body and the Self*, edited by Jose Luis Bermudez, Naomi Eilan, and Anthony J. Marcel, 291–303. MIT Press.

Cole, Jonathan. 1991. *Pride and a Daily Marathon*. 1st MIT Pr. MIT Press.

Eilan, Naomi, Anthony Marcel, and Jose Luis Bermudez. 1995. "Self-Consciousness and the Body." In *The Body and the Self*, edited by Jose Luis Bermudez, Anthony Marcel, and Naomi Eilan, 1–28.

Gallagher, Shaun. 2005. *How the Body Shapes the Mind*. OUP.

Lotze, Martin, and G Lorimer Moseley. 2007. "Role of Distorted Body Image in Pain." *Current Rheumatology Reports* 9: 488–96.

Martin, M.G.F. 1993. "Sense Modalities and Spatial Properties." In *Spatial Representation*, edited by Naomi Eilan, Rosaleen McCarthy, and Bill Brewer, 206–18. OUP. doi:10.1111/j.1468-0149.1963.tb00767.x.

———. 1995. "Bodily Awareness: A Sense of Ownership." In *The Body and the Self*, edited by Jose Luis Bermudez, Anthony Marcel, and Naomi Eilan, 267–89. MIT PRess.

O'Callaghan, Casey. 2015. "The Multisensory Character of Perception." *The Journal of Philosophy* 112 (10): 551–69.

Vignemont, Frédérique de. 2014. "A multimodal conception of bodily awareness." *Mind* 123 (492): 989–1020. doi:10.1093/mind/fzu089.

———. 2015. "Bodily Awareness." http://plato.stanford.edu/archives/sum2016/entries/bodily-awareness/.

Perceptual constancy and the role of perception in action

Perception is a necessary part of action: it structures the environment so that a subject can act in it (Wu 2008, 2011). In particular, action is made possible by perception because the Many-Many Problem, or the Problem of Selection (Wu 2011, 2014, and forthcoming) is solved through the employment of perceptual capacities. In a nutshell, the problem of Selection is the following: among many possible perceivable objects in a scene, and many possible ways to act on them, an agent, before she can act at all, has to select a unique perceptual input-behavioral output couple. This selection is achieved through the deployment of visual attention, which in turn is explained at the neural level by mechanisms based on competition among neural populations for limited resources (Duncan and Desimone 1995, Reynolds, Chelazzi, and Desimone 1999, Wu 2014). In Wu's picture, attention re-structures the perceptual scene by influencing perceptual mechanisms themselves, and this is how perception plays a decisive role both in forming of an intention to act (i.e. selecting what to do among the many possibilities) and in guiding the execution of the action intended (i.e. selecting the appropriate target for the action intended). By means of attention, perception highlights what in the environment is relevant for effective behavior.

In the light of the role of perception in action as I have just described it, the following thought seems innocent enough: since the targets of bodily actions are objects (i.e. kicking a soccer ball), perception must "give" us those objects. In what follows, I propose a way to interpret the idea that perception gives us objects. Sometimes this idea is associated with the notion of perceptual presence (Noë 2004, Leddington 2009). Objects are perceptually present in our experience, that is, they are “right there”, presented directly and immediately to our consciousness, available for us to act on them. This is just an alternative way of saying that perception gives us objects, according to how I am using the term “presence”.

In turn, perceptual presence of objects to consciousness is often spelled out with the help of another notion: perceptual constancy. Different views have different takes on perceptual constancy and its relation to, on one hand, the role of perception and, on the other hand, what perception “gives” us as its content (broadly understood).

How exactly we should think about perceptual presence itself is up for debate. Alva Noë (2004, 2012), for example, argues that perceptual presence is virtual, that is, it is an effect of certain cognitive operations on something more “basic” that our senses pick out directly, similar to a-modal completion. Perception, according to this picture, has a purely sensory component as well as a cognitive, thought-like component. The sensory component is constituted by items which are, in a sense, similar to sense-data, and capture the way objects look from a certain perspective: Noë calls them perspectival aspects, or perspectival properties (p-properties). Objects in their “mind-independent” nature – i.e. objects presented in a way that allows them to be targets of actions and thus part of the subject’s external environment – are then present to consciousness because they are virtually constructed out of a specific perspectival aspect plus some implicit, practical knowledge of how that aspect would vary according to the subject’s movements. Employing such implicit knowledge of patterns of sensorimotor dependence (O’Reagan and Noë 2001) allows the not currently sensed parts of objects to be “present in absence”, thus integrating the awareness of the strictly speaking sensed perspectival aspect.

Perceptual constancy, in this picture, is the phenomenon which allows the mind-independent, objective nature of objects to show up while we deploy out sensorimotor knowledge: the rectangular shape of a table (i.e. a shape constancy) “shows up” as part of the content of a perceptual experience in which there is awareness of a perspectival aspect (i.e. a “non-constancy”) and deployment of the relevant sensorimotor knowledge.

Alternatively, according to some other authors (Gibson 1979, Burge 2010, Smith 2002), perception is by its nature exclusively an objectifying capacity, that is, its only purpose (and its very essence) is to put us in touch with what's constant, or invariant, or essentially perspective-independent about objects. This is the job of perceptual constancies. Despite some even very substantial differences, all these authors argue that objects – in the sense relevant for perception and action – are in fact *nothing but* collections of constancies, or invariants.

In both frameworks sketched above, the same assumption seems to be at work. In particular, the assumption is that the constancies, or invariants, are what captures the "objectivity" of perceptual experiences and, moreover, reveals the "truth" about how things are.

Although the views mentioned above are right in focusing on perceptual presence and perceptual constancy phenomena in order to get clarity on how we should understand the claim that perception gives us objects, I think they are both misguided. They share the same sort of misconception regarding what perceiving objects is like and, in particular, they understand the role and the significance of perceptual constancy in a phenomenologically incorrect way.

Re-elaborating some ideas originally found in Merleau-Ponty (1945), I argue that both the constancies and the non-constancies are *equally* limited in terms of the way they grant the perceptual presence of objects; in other words, neither of them is in a privileged position to give us "the things". Moreover, it seems that the constancy experience doesn't always consist in an awareness of particular "constancies", that is, properties attributed to objects independently of how they look from our current point of view.

As a consequence, I suggest the alternative idea that objects are perceptually present, and thus can be "given" in perception, only as part of a positively indeterminate *normative* background sustaining experiences of constancy and of non-constancy alike. Objects, on one hand, *transcend* every specification in terms of their experienced properties, real or apparent, constant or non-constant. Objects inhabit a dimension of experience which sustains the deployment of attention and the structuring work of perceptual capacities but, on the other hand, they manifest their presence to consciousness by influencing how good or bad we take our current view of them to be, as well as by *prescribing* what we'd have to do in order to improve such a view (see also Kelly 2005, 2007, 2010).

With its focus on the prescriptive role of perception instead of its descriptive one, the normative background view offers an intriguing way of understanding what it is to perceive objects, given perception's role in solving the Problem of Selection for action. Despite its unconventionality, this view provides conceptual tools useful to those in search of a theory of experience which differs from both representationalism and traditional forms of naïve realism/relationalism.

Model-theoretic semantics and semantic competence

As is well known, the first analyses in **model-theoretic semantics for natural languages** [MTS-NL] were conducted from the point of view of radical anticognitivism. Such approach represented the natural heritage of Frege's antipsychologism about logic, and was endorsed also by Montague and Davidson (among others). Early in the development of MTS-NL, researchers with a background in linguistics, such as Partee, Bach, Bartsch among the others, started to work on MTS-NL pursuing a different, more **cognitively oriented approach**. According to such view, which is the natural extension of the Chomskian approach to syntax (Chierchia and McConnell-Ginet 2000), MTS-NL is a branch of psychology. Its aim is to provide a descriptively adequate, explanatory account of meaning as a cognitive phenomenon (e.g., some body of semantic knowledge or information that speakers possess, or some set of psychological processes underlying human semantic abilities). That is, operations represented by the recursive rules of the theory mirror the actual computations performed by the semantic system of the brain, and the knowledge represented by lexical axioms of the theory is tacitly stored somewhere in the speaker's cognitive architecture (e.g., Borg 2004; Larson and Segal 1995). Nowadays such cognitivism is the dominant approach in linguistics and in philosophy, though it is often only tacitly assumed. It is explicitly endorsed in many introductory text-books for MTS-NL, both in Montagovian (e.g., Chierchia and McConnell-Ginet 2000; Heim and Kratzer 1998) and Davidsonian traditions (e.g., Larson and Segal 1995; Ludwig and Lepore 2008). Furthermore, it has often been defended in contemporary discussions in philosophy of language (e.g., Borg 2004; Recanati 2004).

In this talk, I will argue that1 **cognitivism about MTS-NL is problematic**. Some authors have argued that the recursive components of MTS_NL could have some empirical import in the study of the compositional aspects of semantic competence (e.g., Bonomi 1976; Marconi 1997; Partee 1997). However, with **lexical competence**, that is, the ability to understand and use words of a natural language, things are clearly different. Genuine competence cannot be model by attributing implicit knowledge of the axioms of (Davidsonian or Montagovian) MTS-NL such as:

[**A1**] Val (x, *cat*) iff x is a cat
[**A2**] $\mathrm{Val}_{\mathrm{M}}$ (*cat*) $\subseteq \mathrm{D}_{\mathrm{M}}$

How could we explain the speaker's ability to use a word such as *cat* by attributing to him/her knowledge that cat applies to a set of individuals in a model [**A2**], or that cat applies to cats [**A1**]? Such information is clearly incomplete. Granted, it is possible to increment the informational import of MTS-NL by adding a set of **meaning postulates** (i.e. conditional or biconditional statements that constrain the extensions of the constant that appear in the antecedent). Meaning postulates can be used to model the speaker's ability to recognize and draw semantic or material inferences (e.g. the inference from *Tom is a cat* to *Tom is an animal*), as opposed to logical or formal inference. Meaning postulates can also be used to model a type of explicit (and expressible) knowledge, i.e. knowledge about the network of connections between that word and other words and linguistic expressions. Finally, meaning postulates a used to model a variety of other "intralinguistic" lexical semantic abilities, such as the ability to make judgments about semantic relations, to define a word, to recover a word from a definition, and so on. These are the aspects of lexical competence that, following Marconi, we may term «**lexical inferential competence**» (Marconi 1997).

Meaning postulates might be enough to model lexical competence with words such as *democracy*, *justice*, *likely*, and so on, i.e. words expressing abstract concepts. Being competent in the use of such words is just being able to use them in inferences, and to know the intra-linguistic connections they bear with other words. Nevertheless, for words such as *cat*, *yellow*, *chair*, and so on, a competent speaker has further lexical abilities, which meaning postulates can hardly account for: she knows how to apply them in the real world. This is the aspect of lexical competence that, following Marconi, we may term «**lexical referential competence**» (Marconi 1997). It is plausible to think that the content of referential competence for concrete words consists in a set of cognitive procedures that systematically relate words to the output of perceptual processes; it is thanks to our perceptual recognition systems, particularly to the visual recognition system, that we can apply words to objects and circumstances in the world. Referential competence could be described as the ability to determine the semantic values (i.e., reference) for primitive, extra logical symbols of a language. So, the referential component of lexical semantics has to do with the determination of the real truth conditions of sentences.

Critically, the referential aspect of lexical competence is **problematic for MTS-NL**. No amount of meaning postulates can explicate this ability: to know that cats are mammals, that have four legs and

a tail, is not going to enable us to recognize a cat in the environment – not unless we know how to apply *legs*, *tail*, and many other words (Marconi 1997; Partee 1980). As Marconi notes, «no matter how many meaning postulates are associated to the entries of a lexicon, no matter how big and comprehensive such lexicon is, these intra-linguistic connections will never "reach out" to the world: the application of language needs referential competence» (1997, p.54). Thus, MTS-NL can be considered only a "modest" (Dummett 1976) theory of competence, for it does not provide a complete account of lexical competence.

This having been settled, it would seem that MTS-NL **needs not be thus incomplete**. It is possible to think that the problem could be simply solved by integrating MTS-NL with an adequate and complete account of an ordinary speaker's referential competence. For instance, we can imagine to integrate a MTS-NL with (a suitable representation of) the cognitive visual procedures that human speakers use to visually perceive and recognize objects in the real word. An interesting recent attempt to deal with the issue is the so-called *Model-theoretic approach* proposed by Ursini (2011), which tries to combine a particular form of MTS-NL (*Discourse Representation Theory*) with a representation of visual information given in terms of Marr's computational theory of vision. Simple as it is, the MTS-NL so completed might be an adequate and complete model of human competence. In the theory, the semantic values (reference) for extra logical words would be determined by implementing the cognitive procedures that ordinary speakers use to (visually) recognize entities in the world.

However, I believe that a **deep philosophical problem** affects any attempts to provide a completion of this sort for MTS-NL. In a nutshell, the problem is the following. MTS-NL has its source in the work of those logicians and philosophers of language who viewed semantics as a non-psychological enterprise, generally taking such notions as truth-conditions and entailment relations as not epistemically constrained. According to such tradition, semantic analysis aims to characterize the «**objective truth-conditions**» of a sentence (Dummett 1976; Jackendoff 1990; Langacker 1986), i.e. the state of affairs that would have to be in place for the sentence to be true, independently of the actual (or future) state of knowledge of the individual, or of the community as a whole. Let us call such view «**truth-conditions objectivism**». Is the **contemporary defender** of MTS-NL still committed to truth-conditional objectivism? Metaphysical questions are rarely discussed in the contemporary papers in MTS-NL, primary because they do not immediately affect the semanticist's work. However, when the theoretical framework of MTS-NL is presented and discussed -as in the introductory texts of the discipline- such view is often explicitly endorsed (e.g., de Swart 1998, Cann 1990).

This seems to have important philosophical consequences. Truth-conditions objectivism requires what we can call an «**objectivistic view of reference**» (Marconi 1997), according to which words have the reference they have in virtue of stable and determinate connections with objects in the world, independently of whatever knowledge or ability is available to the ordinary speaker (or to the linguistic community as a whole). For instance, the word *cat* has the reference it has in virtue of a direct connection between the word and the cats, or the property of BEING A CAT, or the species of cats, independently of how we perceive or recognize cats in the world. As is well known, such objectivistic view of reference (determination) has been proposed and defended by **Kripke and Putnam** among others. Kripke and Putnam have offered many arguments against the view that semantic values (intension, reference) are cognitively determined (Putnam 1975; Kripke 1980). According to the mainstream, these arguments have shown that the semantic values (intention, reference) of primitive expressions, and of sentences containing them, are not (wholly) determined by psychological facts. Thus, specification of word reference, including specification of the truth-conditions of sentences, cannot be provided by a theory of visual recognition, or by any other psychological theory.

If this is correct, then it seems that MTS-NL is committed to an objective or metaphysical notion of reference that can **hardly be implementable** by a psychological theory of human lexical semantic competence. Therefore, the idea that MTS-NL provide an adequate account of semantic competence is seriously undermined. In the period from the late 1960's into the early 1980's, this foundational problem was generally explicitly recognized by scholars working in MTS-NL (e.g., Dowty et al. 1981; Partee 1979). However, gradually the landscape changed, and foundational problems of MTS-NL disappeared from the scene (see Partee 2006, 2011). But although this question or family of questions has been relatively silent throughout the history of MTS-NL, such issue has remained a deep and unsolved problem. In this paper, I do not intend to offer a solution of this problem. Rather, at the end of our talk, I will just draft **two possible but problematic reactions** to this problem, namely the externalist solutions (e.g., Fodor 1981) and the separatist position (e.g., Marconi 2002; Mazzone 2000).

References

Bergmann, G. (1978). The Metaphysics of Logical Positivism. Greenwood Press.
Block, N. (1986). Advertisement for a Semantics for Psychology. Midwest Studies in Philosophy, 10(1), 615–78.
Bonomi, A. (2015, November 3). Linguistica e logica / Andrea Bonomi.
Borg, E. (2004). Minimal Semantics. Oxford University Press.
Burge, T. (1979). Individualism and the Mental. Midwest Studies in Philosophy, 4(1), 73–122.
Burge, T. (1982a). Two Thought Experiments Reviewed: Comments on J. A. Fodor's Paper: "Cognitive Science and the Twin-Earth Problem." Notre Dame Journal of Formal Logic, 23(3), 284–293.
Burge, T. (1982b). Two Thought Experiments Reviewed. Notre Dame Journal of Formal Logic, 23(July), 284–94.
Chierchia, G., & McConnell-Ginet, S. (2000). Meaning and Grammar: An Introduction to Semantics. MIT Press.
Cann, R. (1993). Formal Semantics: An Introduction. Cambridge University Press.
Dowty, D. R., Wall, R. E., & Peters, S. (1981). Introduction to Montague Semantics. Springer.
Dummett, M. (1976). What is a Theory of Meaning? (II). In G. Evans & J. McDowell (Eds.), Truth and Meaning: Essays in Semantics. Oxford: Clarendon Press.
Fodor, J. A. (1981). Representations : philosophical essays on the foundations of cognitive science (First MIT Press paperback edition.). Brighton: Harvester.
Harman, G. (1986). Change in View. MIT Press.
Heim, I., & Kratzer, A. (1998). Semantics in Generative Grammar. Blackwell.
Jackendoff, R. (1990). Semantic structures. Cambridge, Mass: MIT.
Kripke, S. A. (1980). Naming and Necessity. Harvard University Press.
Lakoff, G. (1980). Metaphors We Live By. University of Chicago Press.
Langacker, R. W. (1986). An Introduction to Cognitive Grammar. Cognitive Science, 10(1), 1–40.
Larson, R. K., & Segal, G. (1995). Knowledge of Meaning an Introduction to Semantic Theory. Monograph Collection (Matt - Pseudo).
LePore, E., & Ludwig, K. (2007). Donald Davidson's Truth-Theoretic Semantics. Clarendon Press.
Marconi, D. (1997). Lexical competence. Cambridge, Mass.: London : MIT Press.
Marconi, D. (2002). The normative ingredient in semantic theory. Hänsel-Hohenhausen
Mazzone, M. (2000). Tre "puzzles" sul riferimento. Lingua e Stile, (1/2000).
McGinn, C. (1989). Mental Content. Blackwell.
Partee, B. H. (1980). Montague Grammar, Mental Representations, and Reality. In S. Kanger & S. Öhman (Eds.), Philosophy and Grammar (pp. 59–78). Springer Netherlands. http://link.springer.com/chapter/10.1007/978-94-009-9012-8_5. Accessed 21 November 2015
Partee, B. (1979). Semantics — Mathematics or Psychology? In D. R. Bäuerle, P. D. U. Egli, & P. D. A. von Stechow (Eds.), Semantics from Different Points of View
Partee, B. H. (1997). Montague Semantics. In Benthem & Meulen (Eds.), Handbook of Logic and Language (pp. 5–91). MIT Press.
Partee, B. H. (2006). Richard Montague (1930-1971). In K. Brown (Ed.), Encyclopedia of Language and Linguistics (pp. 8–255). Elsevier.
Partee, B. H. (2011). Formal Semantics: Origins, Issues, Early Impact. The Baltic International Yearbook of Cognition, Logic and Communication, 6(1), 13.
Putnam, H. (1975). The Meaning of "Meaning." Minnesota Studies in the Philosophy of Science, 7, 131–193.
Recanati, F. (2004). Literal Meaning. Cambridge University Press.
Recanati, F. (2010). Truth-Conditional Pragmatics. Oxford University Press.
Russell, B. (1985). The Philosophy of Logical Atomism. Open Court.
Ursini, F.-A. (2011). Space and the Vision–Language Interface: A Model-Theoretic Approach, 5(3).
Whitehead, A. N. (1929). Process and Reality: An Essay in Cosmology. Free Press.
Wittgenstein, L. (1994). Tractatus Logico-Philosophicus: English Translation. Routledge.

The empirical status of the pictorial view of meaning

Advocates of the **pictorial theory of meaning** (e.g., Locke, Hobbes) claimed that meaning of a word is a mental picture, and lexical semantic competence – i.e. knowledge of word meaning - is closely connected to visual imagery. As a semantic theory, the pictorial theory was discredited in the 20th century. Nevertheless, there is evidence that visual imagery does play a role in semantic processing and is not just a possible side effect of it. **High imageable [IMG] words** (*banana*, *smile*, *chair*), as opposed to low IMG words (deduction, event, democracy) appear to facilitate several tasks related to semantic processing. For example, in long-term memory tests, normal subjects tend to have a better memory for nouns associated with high IMG ratings compared to low IMG words (Paivio 1966, 1967), and high IMG terms are more easily recognized by subjects with deep dyslexia (Jones 1985). Moreover, it seems that semantic processing of high IMG words is correlated with greater activation of visual-related areas such as left inferior temporal gyrus and fusiform gyrus (e.g., Mellet et al. 1998) and sometimes even occipital primary visual cortex (e.g., Goldenberg et al. 1991, 1992; Palmiero et al. 2009). Activation of visual related areas in connection with processing of high IMG words strongly reinforces the conjecture that such processing is, minimally, accompanied by visual imagery. If, as it appears to be the case, such imagery facilitates linguistic performance, this is a good reason to further conjecture that visual imagery is not merely a side effect of processing high IMG words but an active factor of it.

At least prima facie, such data could be seen as decisively vindicating the pictorial view of meaning, as far as high IMG words are concerned. Indeed, in the last few decades the theory's empirical success has much contributed to its partial **redemption as a philosophical theory of meaning** (e.g., Prinz 2002). Particularly, researchers working within the so-called Embodied or Simulation Framework have often claimed that mental imagery, and specifically *visual* imagery, is critical to semantic competence (e.g., Kemmerer 2010; Martin 2007). In recent years, advocates of the imagistic/pictorial view of meaning have devoted much philosophical and neuroscientific research effort to the issue of abstract, low IMG words. Some philosophers and neuroscientists have tried to account for abstract words in imagistic terms (Prinz 2002; Barsalou 2008; Lakoff 2014). Other scholars have postulated entirely different systems for the understanding of abstract words and the under-standing of visually (more generally, sensorimotorially) loaded words (e.g., Dove 2010; Louwerse 2011; Paivio 1973). However, it has been generally taken for granted that the pictorial theory is not problematic for words such as *dog*, *chair*, or *run*. In contrast, I believe that there are **important reasons to be cautious** about the empirical status of the pictorial view of meaning even when restricted to concrete, high IMG terms.

My specific aim in this talk is to critically assess the neuroscience data supporting the hypothesis that visual imagery is a critical component of semantic processing with high IMG words. Indeed, I do not aim to provide a completely exhaustive analysis of the relevant data. I shall limit myself to providing grounds for some caution concerning the cognitive reality of the pictorial view of meaning.

First, I am going to claim that the existing **neuroimaging studies** supporting the pictorial view of meaning (as restricted to high IMG words) are affected by some potential confounds and inconsistencies. For the purposes of the present talk, I performed a **qualitative review of 28 studies** that manipulated the IMG factor in neuroimaging experiments. In my review, we found 17 studies that reported selective visual related activation during the processing of high IMG words (as contrasted to low IMG words), as predicted by the pictorial view of meaning (*ref.*in the talk). However, caution should be taken for several reasons. The first, obvious reason is that not all the neuroimaging studies investigating the role of IMG in language processing did report activation of visual related areas during the processing of high IMG words (e.g., Fiebach and Friederici 2004; Grossman et al. 2002; Kiehl et al. 1999; Krause et al. 1999). The second reason to be cautious is that the specific visual areas that have been associated with IMG varied across the reviewed studies: in some studies the portion of visual cortex correlated with an increase in IMG was circumscribed to areas in the ventral temporal cortex, particularly in the middle portions of fusiform gyrus, (e.g., Bedny and Thompson-Schill 2006; D'Esposito et al. 1997), whilst in others it involved occipital cortex (Goldenberg et al. 1991; Palmiero et al. 2009). A third reason to be cautious is that in most of the studies included in our review, subjects were explicitly requested to deliberately produce a mental picture representing the semantic content of the linguistic stimuli used (words/sentences) (e.g., Mellet 1998) Arguably, the selective visual activation observed in these studies could be the effect of the additional task involved in the first condition (*image*!), not a component of semantic processing per se. The fourth and perhaps most important reason to be cautious is that in most of the visual studies included in our review linguistic stimuli were presented visually (e.g., Sabsevitz et

al. 2005; Binder et al. 2005; Hoffman et al. 2010). Clearly, from the standpoint of the present talk's issue-such a procedure is affected by potential visual presentation confounds.

Secondly, I will claim that **neuropsychological data**, i.e. data from patients with brain insults, do not conclusively demonstrate that visual imagery is a *necessary* component of semantic processing of high IMG words. Clearly, the pictorial view of meaning would be strengthened by the observation that, in the vast majority of cases, patients with lesions involving visual (imagery) cortex have deep semantic deficits with high IMG words, even when no visual stimuli are involved. As far as I know, some studies involving patients with left posterior vascular lesions seem to support this lesional correlation (Manning 2000; Forde et al. 1997; Humphreys et al. 1997). However, language impairment with high IMG words in these patients was evident only when they were formally tested in experimental conditions, but not in everyday conversational exchanges. This might suggest that, at least in some cases, language impairment following damage to visual (imagery) cortex can be subtle rather than catastrophic, and can be limited to relatively complex verbal tasks (see also Binder and Desai 2011). More importantly, several patient studies provide evidence against the hypothesis that preservation of visual (imagery) cortex is necessary for semantic competence with high-IMG words. For instance, HJA, a patient described by Riddoch and Humphreys (1987), suffered from a posterior cerebral artery stroke with consequent bilateral infarcts of the inferior occipitotemporal gyri, involving the fusiform gyri and temporo-occipital junction [BA 37]. These lesions caused prosopagnosia and deep visual agnosia (i.e. inability to recognize visually presented objects in spite of intact early vision). Despite this, HJA had preserved abilities with both high and low IMG words when no pictures were involved, as in spontaneous conversation. Furthermore, in experimental contexts he was perfectly able to provide accurate definitions of concrete words referring to common objects or animals, including details about their referents' visual characteristics. Other patients with left posterior vascular lesions were still perfectly able to use concrete, high IMG words not only in spontaneous conversation, but in more complex verbal semantic tasks as well (Behrmann et al. 1994; Carlesimo et al. 1998; Fery and Morais 2003).

Thirdly, I will claim that the case of **congenitally blind speakers** is crucial in deciding the latter issue. The pictorial view of meaning predicts that semantic competence with high IMG words in congenitally blind subjects should be deeply impaired, as these subjects are supposedly incapable of specifically visual imagery (e.g., Descartes, Locke). As regards adult competence, some studies have shown that the blind adults' linguistic ability with high IMG words (e.g., defining a word, recovering a word from a verbal definition) is not significantly different from that of sighted subjects (DeMott 1972; Rosel et al. 2005; Vinter et al. 2012). Interestingly, in memory tests blind speakers, like sighted subjects, tend to better perform with concrete, high IMG words than with abstract, low IMG words (Cornoldi et al. 1979). Even more surprisingly, it has been found that (congenitally) blind subjects activate a left-lateralized network of "visual" cortices during verbal semantic tasks (review in Bedny et al. 2011). It has been conjectured that such "visual" activation reflects general language processes that occur in left occipitotemporal regions as a result of plastic changes in blind subjects' early infancy (e.g., Bedny et al. 2014). If this were the case, "visual" activation should be observed during the processing of both high and low IMG stimuli; i.e., contrary to the sighted individuals, "visual" activity in the congenitally blind should be insensitive to manipulation of the "IMG" parameter. This cannot be established from extant neuroimaging studies, as in the vast majority of them the "IMG" parameter was not experimentally controlled. Only in one study (Lambert & colleagues 2004) six blind subjects were assessed during the production of mental images from high IMG words vs passive listening to low IMG, "abstract" words. Interestingly, in such study a strong IMG effect was found in various visually related areas such as the primary visual area, the lingual gyrus, and the posterior left temporal gyrus, suggesting that such areas were more (or exclusively) active when blind subjects were creating a mental image from a concrete word. Interestingly, the finding that the sighted and the blind brain do not respond differently to IMG manipulations in language tasks might suggest that, in both cases, what is involved is a kind of **structural mental imagery,** i.e. imagery relative to structural properties of objects (object shape, motion, spatial localization, size), that is not inherently visual. In this sense, these studies might provide some support for the «supramodal theory» concerning the functional organization of ventral visual cortex in both the sighted and the blind (Ricciardi et al. 2014; Ricciardi and Pietrini 2011). Indeed, the supramodal hypothesis might have strong impact for the empirical status of the pictorial view of meaning. As we argued, the intuition behind the pictorial view -at least in its original versions- is the idea that the meaning of a word is a mental picture, or at any rate, the phenomenological content of some visual imagery experience. The hypothesis that, in both sighted and blind individuals, the imagery factor in language understating is not strictly visual would undermine such a fundamental assumption.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577-609-660.
Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. doi:10.1146/annurev.psych.59.103006.093639
Bedny, M., & Saxe, R. (2012). Insights into the origins of knowledge from the cognitive neuroscience of blindness. *Cognitive Neuropsychology*, *29*(1–2), 56–84. doi:10.1080/02643294.2012.713342
Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4429–4434. doi:10.1073/pnas.1014818108
Behrmann, M., Moscovitch, M., & Winocur, G. (1994). Intact visual imagery and impaired visual perception in a patient with visual agnosia. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(5), 1068–1087.
Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex (New York, N.Y.: 1991)*, *19*(12), 2767–2796. doi:10.1093/cercor/bhp055
Binder, J. R., Medler, D. A., Desai, R., Conant, L. L., & Liebenthal, E. (2005). Some neurophysiological constraints on models of word naming. *NeuroImage*, *27*(3), 677–693. doi:10.1016/j.neuroimage.2005.04.029
Carlesimo, G. A., Casadio, P., Sabbadini, M., & Caltagirone, C. (1998). Associative visual agnosia resulting from a disconnection between intact visual memory and semantic systems. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *34*(4), 563–576.
Cornoldi, C., Beni, R. D., Roncari, S., & Romano, S. (1989). The effects of imagery instructions on total congenital blind recall. *ResearchGate*, *1*(4), 321–331. doi:10.1080/09541448908403092
Cornoldi, C., Calore, D., & Pra-Baldi, A. (1979). Imagery rating and recall in congenitally blind subjects. *ResearchGate*, *48*(2), 627–39.
D'Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., & Farah, M. J. (1997). A functional MRI study of mental image generation. *Neuropsychologia*, *35*(5), 725–730.
Fery, P., & Morais, J. (2003). A Case Study of Visual Agnosia without Perceptual Processing or Structural Descriptions Impairment. *Cognitive Neuropsychology*, *20*(7), 595–618. doi:10.1080/02643290242000880
Fiebach, C. J., & Friederici, A. D. (2004). Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. *Neuropsychologia*, *42*(1), 62–70.
Goldenberg, G., Steiner, M., Podreka, I., & Deecke, L. (1992). Regional cerebral blood flow patterns related to verification of low- and high-imagery sentences. *Neuropsychologia*, *30*(6), 581–586.
Hoffman, P., Binney, R. J., & Lambon Ralph, M. A. (2015). Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *63*, 250–266. doi:10.1016/j.cortex.2014.09.001
Humphreys, G. W., Riddoch, M. J., & Price, C. J. (1997). Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*(1358), 1275–1282. doi:10.1098/rstb.1997.0110
Jesse J. Prinz. (2002). *Furnishing the mind : concepts and their perceptual basis / Jesse J. Prinz.* Cambridge, Mass; London: MIT Press.
Kemmerer, D. (n.d.). How Words Capture Visual Experience: The Perspective From Cognitive Neuroscience. *cla.purdue.edu*. https://www.academia.edu/20527965/How_Words_Capture_Visual_Experience_The_Perspective_From_Cognitive_Neuroscience. Accessed 28 November 2016
Krause, B. J., Schmidt, D., Mottaghy, F. M., Taylor, J., Halsband, U., Herzog, H., et al. (1999). Episodic retrieval activates the precuneus irrespective of the imagery content of word pair associates. A PET study. *Brain: A Journal of Neurology*, *122 (Pt 2)*, 255–263.
Lakoff, G. (2014). Mapping the brain's metaphor circuitry: metaphorical thought in everyday reason. *Frontiers in Human Neuroscience*, *8*, 958. doi:10.3389/fnhum.2014.00958
Martin, A. (2007a). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. doi:10.1146/annurev.psych.57.102904.190143
Mellet, E., Tzourio, N., Denis, M., & Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *Neuroreport*, *9*(5), 803–808.
Goldenberg, G., Podreka, I., Steiner, M., Franzen, P., & Deecke, L. (1991). Contributions of occipital and temporal brain regions to visual and acoustic imagery--a spect study. *Neuropsychologia*, *29*(7), 695–702.
Jonides, J., Schumacher, E. H., Smith, E. E., et al. (1998). The role of parietal cortex in verbal working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *18*(13), 5026–5034.
Forde, E. M. E., Francis, D., Riddoch, M. J., Rumiati, R. I., & Humphreys, G. W. (1997). On the Links between Visual Knowledge and Naming: a Single Case Study of a Patient with a Category-specific Impairment for Living Things. *Cognitive Neuropsychology*, *14*(3), 403–458.
Paivio, A. (1973). *Imagery and verbal processes.* New York, London: Holt, Rinehart and Winston, 1973.
Paivio, A., Khan, M., & Begg, I. (2000). Concreteness and relational effects on recall of adjective-noun pairs. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Expérimentale*, *54*(3), 149–160.
Palmiero, M., Belardinelli, M. O., Nardo, D., Sestieri, C., Di Matteo, R., D'Ausilio, A., & Romani, G. L. (2009). Mental imagery generation in different modalities activates sensory-motor areas. *Cognitive Processing*, *10 Suppl*
Ricciardi, E., & Pietrini, P. (2011). New light from the dark: what blindness can teach us about brain function. *Current Opinion in Neurology*, *24*(4), 357–363. doi:10.1097/WCO.0b013e328348bdbf
Riddoch, M. J., & Humphreys, G. W. (1987). Visual object processing in optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology*, *4*(2), 131–185. doi:10.1080/02643298708252038
Roxbury, T., McMahon, K., & Copland, D. A. (2014). An fMRI study of concreteness effects in spoken word recognition. *Behavioral and brain functions: BBF*, *10*(1), 34. doi:10.1186/1744-9081-10-34
Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage*, *27*(1), 188–200. doi:10.1016/j.neuroimage.2005.04.012
Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring "how" from "where." *Neuropsychologia*, *41*(3), 280–292.
Wise, R. J., Howard, D., Mummery, C. J., Fletcher, P., Leff, A., Büchel, C., & Scott, S. K. (2000). Noun imageability and the temporal lobes. *Neuropsychologia*, *38*(7), 985–994.

Doing and Undoing. Some Strategies to Backpedal on Our Speech Acts

Starting with the pioneering work of J.L. Austin (1962), several philosophers of language have turned their attention to speech acts and their conditions for success[1]. In spite of providing an articulate framework for how to do things with words, the literature lacks discussion on how to *undo* what we have done in uttering certain expressions[2]. This paper sets out to identify and examine the most common strategies to backpedal on one's speech acts. In particular, it is an examination of three "backpedalling strategies" – namely, the Annulment Strategy, the Retraction Strategy, and the Amendment Strategy.

Austin himself pointed out that speech acts may be "annulled" if certain conditions are found not to be fulfilled[3]. To understand what he had in mind, imagine that, after a two-step online interview, Bob is told that he got the job. Very happy with that, he attends in person at the HR Department the next day, but they regretfully inform him that, unbeknownst to the recruiter, that job position had already been filled. Since the act of hiring someone is dependent on a specific position being actually vacant, the recruiter's act – at first regarded as felicitous both by the speaker and the hearer – is now to be ratified as null and void. As one can see, the Annulment Strategy is a mechanism for undoing illocutionary things only *loosely speaking*, in that an annulled act is an illocution that never really came into effect. Annulling an act amounts to recognizing that its validity was only purported. However, speakers have at their disposal other strategies to undo things *strictly speaking – i.e.*, to cancel acts they have carried out with success. Amongst them are what I call the Retraction Strategy and the Amendment Strategy.

Nigel Farage's decision to "unresign" in May 2015 offers a good example of how the Retraction Strategy works. As promised in his book (*The Purple Revolution*), Farage stepped back as UKIP leader on 8 May 2015, following his electoral defeat in South Thanet. However, his resignation was "unanimously rejected by the NEC members"[4] – and, on that basis, Farage withdrew it. It should be noted that the party's NEC was not in a position to *reject* Farage's resignation. To put it differently: no official acceptance from NEC members was needed for Farage's resignation to take effect[5]. Instead, NEC members were in a position to *recommend* that Farage retraced his steps. This remark is significant for our discussion, since Farage's unresignation is a proper instance of Retraction only if the resignation was successful in the first place. (It would have been a misfire, had its performance been conditional on NEC members' acceptance.) While we annul unsuccessful acts whose effects were deemed to be binding because of the ignorance of the conversational participants, we retract successfully performed acts whose conventional effects truly entered into force. The very difference between *annulling* and *retracting* is that in retraction cases the act is *made* null, whereas in annulment cases it is *recognized* as null. To retract an illocution means to take it back, to scratch it from the conversational context.

Let us move on to the Amendment Strategy. It is rather obvious that speakers sometimes alter the *content* of their utterances. We are pretty used to people who change their stories for all sorts of reasons. In my view, in the same way as one can amend what one has said ('propositional amendment'), one can, under certain circumstances, amend what one has done ('illocutionary amendment'). Even though it is less obvious, it is still pretty common that speakers alter the *force* of their acts. The most easily recognizable instances of Amendment are utterances such as

---

[1] Cf., among many others, Strawson (1964); Searle (1969, 1975a, 1975b); Bach & Harnish (1979); Searle & Vanderveken (1985); Alston (2000).

[2] A notable exception is the strand of literature on retracting assertives. Cf., esp., Bach & Harnish (1979); Walton & Krabbe (1995); Krabbe (2001); MacFarlane (2011, 2014).

[3] Some considerations on Annulment can be found in Austin's preparatory notes for *How to Do Things with Words*. Cf. Sbisà (2007) for a detailed discussion of those notes.

[4] See http://www.independent.co.uk/news/uk/politics/ukip-rejects-nigel-farages-resignation-statement-from-the-party-in-full-10242465.html (accessed June 12th, 2017).

[5] The party's constitution does not foresee the possibility for the NEC to refuse an official resignation.

(1) *P* … at least, I guess that *P*
(2) I promise to do *P* … well, I'll try to
(3) Do it! … Sorry, I mean, can you do it?

In each of such cases, the speaker seeks to reshape the illocutionary force of her utterance: (1) is an assertion amended into a conjecture, (2) is a promise amended into the expression of an intention, and (3) is an order amended into a request. An important point brought out by these examples is that an act cannot be amended into *any* other act. Indeed, when we amend an illocution, the original one leaves traces in the normative context, as its force imposes some constraints on the act we can replace it with. One can adjust, for instance, an assertion into a conjecture, but one cannot sensibly adjust it into a command, a refusal, or a promise[6]. To account for this feature, we may construe the Amendment Strategy as a mechanism to tamper with the degree of strength of one's speech acts, *i.e.*, to weaken or strengthen their normative burden. For example, in amending an order into a request (say, for politeness reasons), one is reducing the strength of the normative bond imputed on the hearer, who will now be petitioned to grant the request but will have no duty to do so.
The Amendment Strategy may also take the form of what Saul (forth.) calls 'figleaf': an utterance made in addition to an otherwise overtly racist one that provides cover for the racism of the speaker and the utterance[7]. Consider the following utterance of Glenn Beck's. Beck is interviewing Muslim congressman Keith Ellison.

> OK. No offense, and I know Muslims. I like Muslims. I've been to mosques. I really don't believe that Islam is a religion of evil [...]. With that being said, [...] I have to tell you, I have been nervous about this interview with you, because what I feel like saying is, "Sir, prove to me that you are not working with our enemies". And I know you're not. I'm not accusing you of being an enemy, but that's the way I feel[8].

If Beck had uttered

(4) Sir, prove to me that you are not working with our enemies

on its own, this would have amounted to an indirect (but very clear) accusation against Ellison. Crucially, Beck finishes with an explicit denial that his utterance is an accusation,

(5) I'm not accusing you of being an enemy, but that's the way I feel,

by which he amends what he is *prima facie* doing with his words into the mere expression of a feeling[9]. As Saul notices, there are at least three kinds of figleaves at work in Beck's speech: (i) Friendship Assertion ("I know Muslims. I like Muslims. I've been to mosques"), (ii) Mention (since Beck *feels like* saying (4) – an overtly Islamophobic utterance – rather than straightforwardly saying it, (4) is an instance of *mention* rather than *use*), and (iii) «a more complex further move», which I interpret as an illocutionary amendment ("I'm not accusing you of being an enemy, but that's the way I feel").
In conclusion, let me state clearly something that has been implicit throughout: when a speaker successfully performs a speech act, the conversational context adjusts in such a way as to include new

[6] I am concerned with direct speech acts. If an assertive is used to indirectly perform a request, a promise, or a speech act of another kind, then its amendment conditions will be completely different.
[7] Saul uses the term 'racism' in a broad sense, encompassing prejudices of race, nationality, religion, gender, etc.
[8] Quoted in Fatima (2013: 341) and Saul (forth.).
[9] An accusation is an assertion whose content ascribes responsibility to a certain individual for some bad state of affairs. While assertions are statements that the speaker commits to defend (if asked to), the expression of a feeling does not entail such a burden of proof. *Asserting* and *expressing a feeling* form the opposite ends of a single illocutionary strength scale: the first is a (strong) truth claim about how the world is, whereas the second is a (weak) "wholly personal" (and thus non-truth-bearing) claim about how the speaker feels. Cf. Kukla (2014).

normative facts (*e.g.*, obligations, commitments, rights) illocutionarily engendered by the act. In the course of a conversation, the deontic statuses of the participants evolve as new speech acts are accomplished. *Annulling*, *retracting*, and *amending* are mechanisms to rub out a certain deontic update – to disavow or undo the normative constraints generated by speakers' illocutions.

References

Alston, W.P. (2000), *Illocutionary Acts and Sentence Meaning*, Ithaca (NY): Cornell University Press;
Austin, J.L. (1962), *How to Do Things with Words*, London: Oxford University Press;
Bach, K., Harnish, R.M. (1979), *Linguistic Communication and Speech Acts*, Cambridge (MA): MIT Press;
Fatima, S. (2013), "Muslim-American Scripts", *Hypatia*, 28 (2), 341-359;
Krabbe, E.C. (2001), "The Problem of Retraction in Critical Discussion", *Synthese*, 127, 141-159;
Kukla, R. (2014), "Performative Force, Convention, and Discursive Injustice", *Hypatia*, 29(2): 440-457;
MacFarlane, J. (2011), "What Is Assertion?", in J. Brown & H. Cappelen (eds.), *Assertion. New Philosophical Essays*, New York: Oxford University Press, 79-96;
MacFarlane, J. (2014), *Assessment Sensitivity: Relative Truth and its Applications*, Oxford: Clarendon Press;
Saul, J.M. (forthcoming), "Racial Figleaves, the Shifting Boundaries of the Permissible and the Rise of Donald Trump", *Philosophical Topics*;
Sbisà, M. (2007), "How to Read Austin", *Pragmatics*, 17, 461-473;
Searle, J.R. (1969), *Speech Acts. An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press;
Searle, J.R. (1975a), "Indirect Speech Acts", in P. Cole, J.L. Morgan (eds.), *Syntax and Semantics - Speech Acts*, New York-London: Academic Press, 59-82;
Searle, J.R. (1975b), "A Taxonomy of Illocutionary Acts", in K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, VII, Minneapolis: University of Minnesota Press, 344-369;
Searle, J.R., Vanderveken, D. (1985), *Foundations of Illocutionary Logic*, New York: Cambridge University Press;
Strawson, P.F. (1964), "Intention and Convention in Speech Acts", *Philosophical Review*, 73(4), 439-460;
Walton, D.N., Krabbe, E.C. (1995), *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, Albany (NY): State University of New York Press.

Towards a Shared Frame for Imaginative Episodes

Imagination is, without a doubt, one of the most impressive skills of our minds. Our capacity to entertain scenarios that are not actual, to engage in counterfactual reasoning, or to engage in episodes of pretense (such as games of make-believe, theatre plays or role-playing games), suggests that the architecture and mechanisms governing imagination should be rather complex but, at the same time, quite "natural" (as it can be seen in studies showing the capacity of young children to recognize and engage in pretense play). Moreover, one of the main topics in the debate regarding the "place" of imagination within the mind concerns how it is related to other mental attitudes, and how similar and different it is from them. Relatively recently, many authors have given an account on how imagination could work. What happens, within our mind, when we decide to imagine a certain hypothetical scenario, or when we engage in an episode of pretense? How and where are these imaginary situations created, with respect to other mental attitudes? How do they interact with, for instance, our beliefs, or the reasoning mechanisms that work over them?

As the starting point of this work, we review three influential works on imagination: firstly, the distinction Timothy Williamson draws between two modes in which imagination can run (namely, voluntary and involuntary); then, Shaun Nichols and Stephen Stich's cognitive theory of pretense, in which they propose an architecture of the cognitive mind showing how pretense episodes relate to our beliefs; finally, Peter Langland-Hassan's theory on guiding chosen imaginings, in which the imaginative states resulting from an initial and voluntary act of imagination are then treated via certain sets of rules or algorithms that determine how the imaginative episode could evolve. When reviewing these works, we notice something that we think it is quite relevant: although these approaches do have their differences, they also have strong coincidences, specially in the underlying frame upon which they draw.

Through the analysis and comparison of the previously mentioned works, we identify and propose an underlying, theory-independent frame that we claim every theory about voluntary imagination should account for. This frame is characterized by identifying four main steps involved in every *agentive* act of imagination (that is: every act of imagination that is initiated by a conscious intention of the agent to imagine such-and-such): a first, imagining-initiation voluntary mental action in which the agent sets the initial content of the imaginative episode; an involuntary, rule-driven "concretization" in which the initial imagined scenario is filled up with details; a default advancement or evolvement of the imaginative episode, driven by rule-like scripts determining how the depicted situation would usually unfold; and a voluntary, desire-driven advancement of the imaginative episode using "off-script" new inputs that, in turn, involve a cyclical recursion of the whole process. Once we have set and explained what characterizes each of these four steps, we show how our proposed frame indeed identifies a minimum number of stages every imaginative episode should have and which, one way or another, any of the previously reviewed theories fulfill; furthermore, we argue how any other theory aiming to capture these kind of acts of imagination must account, one way or another, for the steps identified in our proposed frame.

The reclamation of slurs

Keywords: slurs, appropriation, affirmative action

Introduction: slurs and appropriation. In the last few decades, philosophers of language and linguists have turned their attention to slurs, expressions that derogate groups and individuals on the basis of their belonging to a certain category. Slurs are typically taken to be toxic words: not only to express discriminatory beliefs, but also to spread them and are therefore banned from public debate in liberal democracies.

In this paper I focus on particular uses of slurs, namely appropriative or reclaimed uses. Appropriation is the phenomenon for which the members of a group can use among themselves the slur targeting their own group, in such a way that the slur is not offensive anymore in those contexts; on the contrary, appropriated slurs are used to express solidarity and underline intimacy. Typical instances of appropriation are positive uses of 'nigga' in afro-american communities in the U.S., as well as positive uses of 'queer' in academic contexts, in acronyms such as LGBTQ+, expressions such as 'Queer Studies' or – outside academia – 'Queer Tango', etc.

Scholars got interested in appropriation because it constitutes an instance of meaning change and therefore it raises problems for most existing accounts of slurs, according to which slurs systematically encode derogatory content. Another reason for which appropriation became important in the debate has to do with the question as to whether reclaimed slurs should be allowed or not in liberal democracies and what kind of evaluative content they convey. In particular, such reclaimed positive uses are taken by Bianchi (2014) to challenge silentism, the thesis defended by Anderson and Lepore (2013) according to which *any* occurrence of slurs (including mentioned ones) should be simply banned. For Bianchi, reclaimed slurs should not be banned because reclamation is a process at the end of which terms can lose their derogatory meaning for good.

An argument allegedly vs. appropriation: the WA. Reclaimed uses of slurs (at least some of them) are analyzed as conveying a positive evaluative content. In this paper, I will stay neutral with respect to how the derogatory content of slurs should be analyzed in the first place, but supposing for the sake of the argument that it amounts to something like '*bad* for being P' (where 'P' is a descriptive property such as 'being gay', 'being black' etc.), reclaimed slurs are typically taken to convey something along the line of '*good* for being an P' (see *i.a.* Jeshion 2017, Ritchie (ms)). Reclaimed uses that convey positive content about the target class give rise to an argument that *prima facie* speaks in favor of silentism, namely the Warrant Argument (WA), according to which appropriative uses of slurs are no less problematic than offensive uses. The idea is the

following: what slurs do in general is to express an evaluation (typically negative) based on some descriptive properties (e.g. 'being gay', 'being black', etc.) that do not *per se* warrant any evaluation (neither positive or negative). As we said, while the standard uses of slurs convey a negative evaluation, appropriative uses (at least some of them) convey a positive one. The Warrant Argument holds that if one claims that what is wrong with a term like 'wop' is that being Italian *per se* never justifies nor warrants *negative* evaluation, then it is very problematic to use 'wop' as conveying that being Italian *per se* warrants *positive* evaluation. After discussing and articulating the WA, I consider some empirical data about stereotype threat that seem to indirectly support such an argument (Spencer et al. 1999).

Saving appropriation: the parallel to affirmative action. While claiming that the WA is both valid and sound, I show that despite appeareances it does not constitute an argument in favor of silentism *à la* Anderson and Lepore (2013). I discuss a way to endorse the Warrant Argument, according to which reclaimed uses of slurs are just as problematic as standard derogatory uses of slurs, while still rejecting silentism. In particular, I discuss a parallel between appropriation and affirmative action. I will show how reclaimed uses of slurs can be accepted and allowed only if we understand them as remedies to power imbalances. Just as in the case of affirmative action, in order to balance an unjust mechanism, we introduce another form of countervailing injustice that is supposed, over time, to balance the initial one. In this framework, I shall interpret the empirical data provided by Galinsky et al. (2013) that speak in favor of the positive effect of appropriation.

References

Anderson, Luvell and Lepore, Ernest (2013), Slurring words, *Nous*, 47 (1), pp. 25–48.

Bianchi, Claudia (2014), Slurs and appropriation: An echoic account, *Journal of Pragmatics*, 66, pp. 35–44.

Galinsky, Adam D. et al. "The Reappropriation of Stigmatizing Labels: The Reciprocal Relationship between Power and Self-Labeling", *Psychological Science*, 24 (10), pp. 2020–2029.

Jeshion, Robin (2017), Pride and Prejudiced: On the Appropriation of Slurs, Handout.

Ritchie, Kate (ms), Social Identity, Indexicality, and the Reclamation of Slurs.

Spencer, Steven J., Steele, Claude M. and Quinn, Diane M. (1999), Stereotype threat and women's math performance, *Journal of experimental social psychology*, 35 (1), pp. 4–28.

Compositional Conceptual Role Semantics

Conceptual role semantics (CRS henceforth) never gained as much popularity as their truth conditional competition. One of the important reasons why they seem unattractive to philosophers can be summarized by a following quote from Lepore:

> "(. . .) you can't identify meanings with conceptual roles tout court, since unlike meanings conceptual roles tout court aren't compositional." (Lepore (1994) p. 198)

In our paper we plan to response to this challenge and propose a version of CRS that is compositional (CCRS, henceforth).

A version of CCRS we are going to present is based on the Directival Theory of Meaning, a theory originally proposed by the inventor of categorial grammars - Kazimierz Ajdukiewicz - in the 1930s (cf. Ajdukiewicz (1978) pp. 35-89). Although it could practically be considered to be the first full-blown conceptual role semantics, it was never recognized as such and remains almost completely forgotten. We believe that the theory is well worth revisiting as it contains some original ideas that might have an impact on modern semantical and philosophical discussions.

In the first part of our paper we are going to describe basic ideas of the new version of Directival Theory of Meaning (nDTM henceforth). Roughly speaking, one of its most important features is that it derives linguistic meanings from a unique combination of syntax and pragmatics. It starts with the titular "meaning directives" which should be understood as specific sentences the language users are obliged to accept in particular types of situations. nDTM distinguishes four types of such directives: axiomatic, inferential, empirical and promotive. The meaning of an expression is, then, defined as the set of all places it occupies within the system of such directives. We will describe this definition of meaning in detail, illustrate it with examples and discuss briefly some problems it faces.

In the second part of our paper we provide a precise formulation of the compositionality problem for the Directival Theory of Meaning. Firstly, we start by noting that the problem of compositionality for nDTM can be formulated in an exact way only if we work with the so-called *Husserlian languages* (cf. Hodges (2001), i.e. languages where synonymy applies only to expressions of the same semantic category). Secondly, following the original idea of Ajdukiewicz who distinguished connected and unconnected languages, we single out the class of languages that meet the following principle:

(COMP) For each expression E of L there exists a meaning directive that contains (non-trivially) E within its scope.

Such languages contrast with the ones in which some expressions do not figure in the scope of meaning directives. The problem of compositionality for the Directival Theory of Meaning, therefore, has two versions that correspond to the

two distinguished kinds of languages.

In the third part of the paper we are going to show that the problem of compositionality has a positive solution both for Husserlian languages that satisfy (COMP) and Husserlian languages that do not satisfy (COMP). In the first case, compositionality is a direct consequence of the definition of meaning proposed by the Directival Theory of Meaning. This consequence, unfortunately, is followed by an unwelcome fact: we have to assume that the set of meaning directives is infinite. In the second case, we have to define a new notion of meaning (strictly speaking: an extension of the original notion of meaning given by the nDTM) for expressions that do not figure in the meaning directives of a language (we call the problem of meaning of expressions that do not figure in any meaning directives the *problem of semantic invisibility*). This new extended notion of meaning may appeal to syntactic composition of complex expressions (it is reasonable to assume that all syntactically simple expressions figure in at least one meaning directive). It may also appeal to the idea of meta-directive: a rule that tells how one can generate new meaning directives if certain other directives are given. In both cases, as we shall argue, the discussed version of CCRS is compositional.

References

1. Ajdukiewicz, K. (1978) The Scientific World-Perspective and Other Es-says", Synthese Library vol. 108, Reidel

2. Hodges W. (2001) Formal Features of Compositionality, Journal of Logic, Language, and Information 10: 7{28

3. Lepore, E. (1994) Conceptual role semantics. In A Companion to the Philosophy of Mind, ed. S. Guttenplan. Oxford: Blackwell, 193-200.

A Classical Implicational Account of Semantic Presuppositions

Abstract *Presupposition* is surely one of the most debated notions in the linguistic and philosophical literature. Historically, there are two main theoretical approaches on presuppositions. According to the first one, the semantic view, presuppositions are semantic implications, that is, truth-conditional relations between propositions and statements. In this sense, presuppositions are considered as properties of sentences and a presupposed proposition is a necessary condition for the truth of the presupposing statement: if a sentence S presupposes a sentence P, then S *entails* P, and if P is false, then S is neither true nor false (cf. Strawson 1950, 1952, van Fraassen 1968, Keenan 1971). On the second approach, the pragmatic view, presuppositions are not properties of sentences but rather properties of speakers or of linguistic performances given a certain context of utterance (cf. Stalnaker 1972, 1973, 1974, 1999, 2002). On this view, a presupposed proposition is a condition for the felicitous utterance of the presupposing statement in a given context.

Traditionally, it is commonly assumed that presuppositions differ from classical entailments, as presuppositions, unlike classical entailments, project under negation. If we compare a context of entailment to a context of presupposition, we should see that entailments, but not presuppositions, disappear under negation, as shown in (1)-(2):

(1) a. Beatrix killed Bill.
entails: 'Bill is dead'
b. Beatrix did not kill Bill.
does not entail: 'Bill is dead'
(2) a. The King of France is bald.
presupposes: 'There is a King of France'
b. The King of France is not bald.
presupposes: 'There is a King of France'

The sentence in (1a) entails the proposition 'Bill is dead', whereas its negative counterpart in (1b) does not. On the other hand, it is assumed that the sentence in (2a) implies an existential presupposition ('There is a King of France'), and that its negative counterpart in (2b) also does. The assumption that presuppositions are constant under negation stems from the Strawsonian account of semantic presupposition (Strawson 1950; 1952: 175 ff.), as illustrated in (3):

(3) *The Strawsonian notion of semantic presupposition*
S *presupposes* P iff:
(i) if S then it *must* be the case that P (= P is a truth condition of S);
(ii) if ¬P then S has *no truth-value* (S is neither true nor false).

The claim was that in order to assign a truth-value (true/false) to S, P *must* be true. Therefore, the conclusion that both the sentence "The King of France is bald" and its negative counterpart "The King of France is not bald" *must* imply that "There exists a King of France".

If both S and the negation of S imply P, then *modus tollens* is assumed not to hold for this kind of implication, and presuppositions are clearly distinguished from *classical* entailments.

This presentation aims at proposing a revision of the notion of semantic presupposition. Following Russell (1905; see also Sellars 1954), I argue that presupposition failure does not result in a lack of truth value, but rather in falsity, and that semantic presuppositions are classical entailments.

Moreover, I claim that most standard cases of presupposition as discussed in the literature, but crucially not all cases, are classical entailments. I contend that factive verbs offer a paradigmatic example of this distinction, as the factivity related to *know* is semantic, hence a classical entailment, whereas the factivity related to *regret* is a merely pragmatic phenomenon.

Keywords – Presupposition, Classical entailment, Constancy under negation, Factivity

References

Keenan, E. L. (1971). Two kinds of presupposition in natural language. In: C.J. Fillmore & D.T. Langendoen (eds.), *Studies in Linguistic Semantics*, 45-52. New York: Holt, Rinehart and Winston, Inc.

Russell, B. (1905). On denoting. *Mind*, New Series 14: 56, 479-493.

Sellars, W. (1954). Presupposing. *The Philosophical Review* 63: 2, 197-215.

Stalnaker, R. (1972). Pragmatics. In: D. Davidson & G. Harman (eds.), *Semantics of Natural Language*, 380-397. Dordrecht-Boston: D. Reidel Publishing Company.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic* 2, 447-457.

Stalnaker, R. (1974). Pragmatic presuppositions. In: M. K. Munitz & P. K. Unger (eds.), *Semantics and Philosophy*, 197-213. New York: New York University Press.

Stalnaker, R. (1999). *Context and Content. Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy* 25, 701-721.

Strawson, P. F. (1950). On Referring. *Mind*, New Series 59: 235, 320-344.

Strawson, P. F. (1952). *Introduction to Logical Theory*. London: Methuen & Co.

van Fraassen, B. C. (1968). Presupposition, implication, and self-reference. *The Journal of Philosophy*, 65: 5, 136-152.

Semantic externalism(s) and content transparency

We examine the impact of various forms of content externalism on a transparency requirement borrowed from Boghossian (1994) and adapted to utterance contents. We conclude that a model that distinguishes at least two levels of meaning is required to account for what we take to be speakers' intuitions about contents.

This paper is part of a wider project exploring a centred-worlds-based model of relativist semantics, on which contents are purely descriptive, with objects and individuals, not just worlds and times, occurring as elements of the circumstance of evaluation. One purported advantages of such a model is that it identifies a level of content which stands a much better chance of achieving transparency (Boghossian 1994, 2011) than models that assume Russellian propositions, which contain individuals as semantic values of referential terms. In this paper, however, our focus is on assessing the impact of various versions of content externalism on the transparency requirement, but as applied to linguistic or 'utterance' content, not mental content. This way, we are diverting Boghossian's requirement from its intended initial object. We will distinguish four variants of externalism, Wittgensteinian, Burgean, Putnamian, and extended-indexicalist. These labels are indicative, and reflect only our preferred interpretation of the views of the above authors.

We take it that Wittgenstein (1953) defended an early variant of 'social externalism', as captured by his thesis that "meaning is use". Consequently, meaning is distributed across the whole set of linguistic practices of a community of speakers and is never encapsulated in a precise definition in terms of necessary and sufficient conditions. Tyler Burge's (1979) anti-individualism embodies a more recent variety of social externalism. Burge notoriously maintained that the meaning of the word *arthritis* is captured by the best description offered by contemporary scientific experts, a description that is not available to non-expert speakers, who defer to the experts for the meanings of the words they do not master completely. A third variety, 'physical externalism', which had been advocated a few years before by Saul Kripke (1972) and Hilary Putnam (1975), held that the meanings of words like *gold* or *water* depend on the actual nature of the very substances the words are intended to denote, whatever the descriptions provided by the current scientific experts. The implication is that experts too fall short of perfect mastery of certain categories of terms used by the community.

We propose a fourth variety of externalism, dubbed 'extended indexicalism'. Even though he argued for a "(social) division of linguistic labor", thus apparently endorsing deference to current scientific experts, Putnam advocated physical externalism, i.e. an indexical rather than descriptive theory of reference for natural-kind terms (Liu 2002). Drawing upon him, we consider the possibility of doing justice to certain 'realist intuitions' of speakers by arguing that social externalism, in its two variants, might always be made subordinate to some indexical externalism. On this hypothesis, all descriptive terms (even sortal or abstract ones, say, *freedom*) are seen as intended to capture phenomena that have a certain nature, which might lend itself to rigorous scientific investigation. It is in this 'nature' of the phenomenon, of which putative future experts may improve our knowledge, that the meaning of these descriptive terms resides. Current usage, or, when available, descriptions provided by current experts only approximate this meaning, in the same way that the descriptions that ordinary speakers have 'in the head' do (though less satisfactorily).

Now, Boghossian (1994) already pointed out that externalist accounts clashed with his transparency requirement on mental contents, and saw in this a reason to question the validity of the externalist theses. Boghossian (2011: 457) formulated a twofold transparency requirement: "(a) If two of a thinker's token thoughts possess the same content, then the thinker must be able to know a priori that they do; and (b) If two of a thinker's token thoughts possess distinct contents, then the thinker must be able to know a priori that they do". To assess transparency of utterance contents, we replace *two of a thinker's token thoughts* by *two contents as understood by a speaker/addressee*. In the following, we investigate whether such a clash also arises in connection with linguistic contents. Can it be maintained that speakers 'know what they mean', because they understand the meanings of the words they use – because these meanings are 'in their heads'? We sketch answers to this question for each form of externalism identified above. We then discuss a fourth variety of externalism and assess it likewise.

Consider a speaker, Anna, who does not herself master the concept ARTHRITIS, as she assumes that arthritis can affect soft tissue as well as joints. Anna says (1):

(1) Emile Ajar has arthritis in his thigh.

On an internalist view of content, the truth-conditions of (1) represented as Russellian propositions – which include the referents of referential expressions – will be something like (2).

(2) [[1]] =  has a painful ailment affecting soft tissue or joints in 

On such a basic internalist account, on which content is determined by the meanings stored in the speaker's mental lexicon, (1) may come out true, thus clashing with strong intuitions that (1) is false: arthritis, it feels, *cannot* affect soft tissue. This provides an argument in favour of some version of externalism. To capture the intuitions behind externalism à la Wittgenstein or Burge, the content of an utterance of (1) can be represented as in (3), which gets the truth-conditions right. $D_x(\sigma)$ is a notational device borrowed from Recanati (2000), which endows an expression σ with a different ('deferential') character, the semantic value that σ has for a community of users of some language L or for some relevant group of experts.

(3) [[1]] =  has $D_{ENGLISH/EXPERTS}$(arthritis) in 

This means that the concept expressed by the token of *arthritis* in (1) is the concept that usage has fixed or, alternatively, that the best experts have fixed. On Burge's picture some speakers, the experts, may entertain the very concept corresponding to arthritis; on Wittgenstein's, the concept is not even possessed by anyone. In both cases our adapted transparency requirement is violated. If Anna said (or heard) (4) instead of (1),

(4) Emile Ajar has a painful ailment affecting soft tissue or joints in his thigh,

she would not be able to know a priori that (4) has a different content from (1).

The kind of physical externalism advocated by Kripke or Putnam affects natural-kind terms. As soon as an utterance contains one, the semantic value of such a term is fixed by 'the world'. Thus, if Anna says (5)

(5) Gold is less dense than lead,

her understanding of the words *gold* and *lead* has no impact on the content of (5). Here again, she fails to have transparent access to the content of the utterance.

If one understands contents descriptively, as on the extended relativist model we are exploring, transparency is violated as well, on all three forms of externalism. On this model, the content of (1) will, very roughly, be represented as (6):

(6) [[1]] = The F has ailment G in the F's thigh

However, to respect (social) externalist intuitions, there must be something in (6) that indicates that the ailment G has to be understood the way usage or the experts fix it. This necessarily introduces an indexical, referential, element that prevents the content from being entirely descriptive, hence from being transparent. Extended indexicalism obviously falls prey to the same problem.

The conclusion is clear: on all forms of externalism we have examined, content fails to be transparent. Yet, there must be a level of content that achieves *some* degree of transparency; otherwise it becomes impossible to account for the role of utterance contents in the behaviour of cognitive agents. The lessons of externalism must be reconciled with a moderate variant of the transparency requirement. We think we can achieve this extended indexicalism and its corollary theory of semantic deference (Author et al. 2007) are incorporated within the relativist semantic model we are currently exploring. This model allows us to distinguish between an 'ideal semantic evaluation' – based on public meanings to which, consciously or not, speakers defer – and a 'subjective evaluation' made by speakers themselves, based on their own conceptions of the meanings and referents of the expressions they are using. The former evaluation requires the identification of a 'language parameter' (corresponding to the *x* in Recanati's deferential operator), i.e. the community of users to whom a speaker defers when performing a given language act. The latter evaluation is what supports a speaker's own theoretical and practical reasonings.

References

Boghossian, P. (1994). The transparency of mental content. *Philosophical Perspectives* 8, 33-50 ▏**Boghossian**, P. (2011) The Transparency of Mental Content Revisited, *Philosophical Studies* 155, 457-465 ▏**Burge**, T. (1979) Individualism and the Mental, *Midwest Studies in Philosophy* 4, 73-121 ▏**Kripke**, S. (1972/1980) *Naming and Necessity*, Cambridge, MA.: Harvard University Press ▏**Liu**, JL (2002), Physical externalism and social externalism: are they really compatible?, *Journal of Philosophical Research* 27, 381-404 ▏**Putnam**, H. (1975) The meaning of 'meaning', in Mind, Language and Reality. Philosophical Papers, volume 2. CUP, 215-271 ▏**Recanati**, F. (2000) *Oratio Obliqua, Oratio Recta: An Essay on Metarepresentation*. Cambridge, MA.: MIT Press ▏**Wittgenstein**, L. (1953) *Philosophical Investigations*. Blackwell Publishing, 2001.

Gender presuppositions in conditionals

Preliminary generalizations Cooper (1983) observes that free 3sg pronouns have *indexical* gender presuppositions: they presuppose that their descriptive content (human male/female, nonhuman) is satisfied by their referents in the actual world. Thus, one cannot utter (1-a) felicitously to say of a female individual x that there is a possible circumstance in which x is male and American. Moreover, one cannot utter (1-b) felicitously in a context in which one knows that Bill is male, but Jones believes that he is female:

(1) a. ??He could have been an American (pointing at Scarlett).
b. ??Jones believes that she (pointing at Bill) is a university professor.

Pronouns anaphoric to proper names also have indexical presuppositions. Thus, (2-b) is infelicitous and so is (2-a) in a context where one knows Alex is male:

(2) a. ??Alex$_i$ could have recognized a conspirator near her$_i$.
b. ??John$_i$ didn't realize that Alex$_j$ was male. He$_i$ thought that she$_j$ liked him$_i$. (Sharvit 2008)

(The indexical presuppositions disappear when the pronoun is bound under a modal, as argued by Del Prete and Zucchi 2016, but we won't address this issue here).

The problem with conditionals Conditional (3), uttered in a context in which the conversational agents know that Sasha is female, supports the above generalization concerning the indexical presuppositions of pronouns anaphoric to proper names:

(3) If Sasha$_1$ were male, we would buy her$_1$/??him$_1$ a toy car.

However, this is not the case for (4) from Yanovich (2010), uttered in a context in which we don't know Sasha's gender:

(4) a. If Sasha$_i$ is male, I'll buy him$_i$ a doll.
b. If Sasha$_i$ is female, I'll buy her$_i$ a toy car.

In (4), the presuppositions of the pronouns must be met in the closest world in which Sasha is male/female. This contrast doesn't just depend on (4-a)-(4-b) being indicative conditionals. Indeed, indicative predictive conditional (5), like counterfactual conditional (3), requires the presupposition to be met in the actual world (M. Kaufmann, p.c.):

(5) If John$_i$ undergoes an operation to become female, we'll buy him$_i$/??her$_i$ a toy car.

***Actually* in conditionals** Jackson (1987) points out that, while counterfactual (6-a) makes perfect sense, the corresponding indicative in (6-b) is incoherent:

(6) a. If Oswald had not shot Kennedy, things would be different today from the way they actually are.
b. ??If Oswald did not shoot Kennedy, things are different today from the way they actually are.

The same contrast also holds between counterfactuals and predictive indicatives, as the following examples from Weatherson (2001) show:

(7) a. If Warren Beatty were to become the next president, things would be different from the way they actually will be.
b. ??If Warren Beatty becomes the next president, things will be different from the way they actually will be.

The moral Jackson draws from these examples is that indicatives, unlike counterfactuals, are not intensional: they do not introduce a distinction between the closest world in which the antecedent is true and the actual world. This is why the indicatives in (6-b) and (7-b) are incoherent: it cannot be that in the actual world things are different from the way they are in the actual world. Jackson's conclusion is that (6)-(7) are

evidence for the material conditional analysis of indicatives. This analysis, however, runs into well-known problems (see Sainsbury 2001, pp. 87-94 for a review). An alternative moral suggested by Weatherson (2001) and Nolan (2003) is this: when using a counterfactual, the speaker evaluates the consequent in the closest world in which the antecedent is true by taking for granted that this world is distinct from the actual world; when using an indicative, on the other hand, the speaker evaluates the consequent in the closest world in which the antecedent is true, *by considering this world as actual*. According to this alternative view, the reason why (6-b) and (7-b) are incoherent is that the world referred to by "actually" is the closest world w in which the antecedent is true and it can't be that in w things are different from the way they are in w.

The proposal In a Kaplanian framework, the suggestion by Weatherson and Nolan amounts to regarding indicative conditionals as *monsters*–they shift the context of utterance c to a context c' relative to whose world and time the antecedent is true (see Santorio 2012 for further evidence from the behavior of first person pronouns that indicative conditionals are monsters). We propose the following semantics (given a context c, "c_w" and "c_t" are short for the world and time of c):
indicative conditionals
$[\![if\ p, q]\!]_{c,g,w,t} = 1$ iff $[\![q]\!]_{c',g,c'_w,c'_t} = 1$, where c'_w is the world closest to w such that $[\![p]\!]_{c,g,c'_w,c'_t} = 1$, for some c' which, for all the speaker knows at $< c_w, c_t >$, might be the context she is in.
counterfactual conditionals
$[\![if\ p, q]\!]_{c,g,w,t} = 1$ iff $[\![q]\!]_{c,g,w',t} = 1$, where w' is the world closest to w such that $[\![p]\!]_{c,g,w',t} = 1$.
Now contrasts (6) and (7) follow: "actually" refers to the world of the context, which for indicatives, but not for counterfactuals, is the closest world at which the antecedent is true. Facts (4)-(5) are also expected. For indicatives, the indexical presuppositions of the pronouns in the consequent must be met at a pair $<c'_w, c'_t>$, which, for all the speaker knows in c, might be the world and time of utterance, where c'_w is the world closest to the actual world in which the antecedent is true at c'_t. In the context of utterance of (5), John's gender is known to the speaker to be male at the utterance time, thus John is male at $<c'_w, c'_t>$ and the presupposition of "her" isn't met. In the context of utterance of (4-a), Sasha's gender is not known, and Sasha is required to be male at $<c'_w, c'_t>$, thus the presupposition of "him" is met. By a parallel reasoning, the presupposition of "her" is met for (4-b). We'll discuss this proposal *vis a vis* recent proposals by Yalcin (2015), Mackay (2017) that argue against "actually" being an actuality operator.

References

Cooper, R. (1983). *Quantification and syntactic theory*. Reidel, Dordrecht.

Del Prete, F. and Zucchi, S. (2016). A unified non monstrous semantics for third person pronouns. Forthcoming in *Semantics and Prgmatics*.

Jackson, F. (1987). *Conditionals*. Blackwell, Oxford.

Mackay, J. (2017). Explaining the actuality operator away. *The Philosophical Quarterly*.

Nolan, D. (2003). Defending a possible-worlds account of indicative conditionals. *Philosophical Studies*, 116:215–269.

Sainsbury, M. (2001). *Logical forms*. Blackwell, Oxford, 2nd edition.

Santorio, P. (2012). Reference and monstrosity. *The Philosophical Review*, 121(3):359–406.

Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy*, 31:353–395.

Weatherson, B. (2001). Indicative and subjunctive conditionals. *The philosophical Quarterly*, 51(203):200–216.

Yalcin, S. (2015). Actually, *Actually*. *Analysis*, 75(2):185–191.

Yanovich, I. (2010). On the nature and formal analysis of indexical presuppositions. In *New frontiers in artificial intelligence: JSAI-isAI Workshops, LENLS, JURISIN, KCSD, LILI, Tokyo, Japan, November 19-20, 2009, Revised Selected Papers*, pages 272–291.

Names are Polyreferential

Abstract

Although the view that sees proper names as referential singular terms has widely been considered orthodoxy, there is a growing popularity to the view that proper names are predicates. This is partly because the orthodoxy faces two anomalies that Predicativism can solve: on the one hand, proper names can have multiple bearers, i.e., for most names there are many individuals called that name. But multiple bearerhood is thought to be problematic for the view that proper names are singular terms because a standard assumption about singular terms is that they are terms that have just one semantic referent. But if names have multiple bearers the issue arises of deciding which one of the bearers of a given name is its semantic reference. On the other hand, as Burge (1973) noted, proper names can have predicative uses. But the view that proper names are singular terms arguably does not have the resources to deal with Burge's cases.

Predicativism can straightforwardly deal with the anomalies mentioned before. If proper names are predicates then they are the kind of term that can be true of many individuals. Thus, there is nothing surprising about the fact that proper names can have multiple bearers. Furthermore, if proper names are predicates, then predicative uses of proper names are not problematic, there are just the standard case. As well as dealing with these anomalies, it is claimed that the predicate view can successfully account for *referential uses* of proper names while still treating them as predicates, thus providing a uniform semantics of proper names.

However, referentialists have counter-challenged Predicativism by calling attention to additional data concerning the behaviour of proper names that cannot be accounted for within the predicativist's unified semantics. These are Jeshion's examples — i.e., common noun uses of proper names that that are semantically different from the predicative uses introduced by Burge (Jeshion, 2015) — and adjectival and verbal uses (Delgado, 2011, 2016).

Thus, an adequate theory of proper names should account for multiple bearerhood and Burge's predicative uses but also should account for the additional data introduced by referentialists.

The view I defend accounts for multiple bearerhood in a way that preserves the referentialist's contention that proper names are fundamentally

referential singular terms. In my view, proper names are *polyreferential*, i.e., they are single words that have more than one referent. The view rejects the assumption that reference for proper names is a one-to-one relation between the name and its referent; instead I claim that reference for proper names is one-to-many: a name is referentially related to each one of its bearers. Thus, proper names are still singular terms even thought they are polyreferential: because their fundamental semantic property is still that of *referring*; and they contribute individuals to propositional content directly.

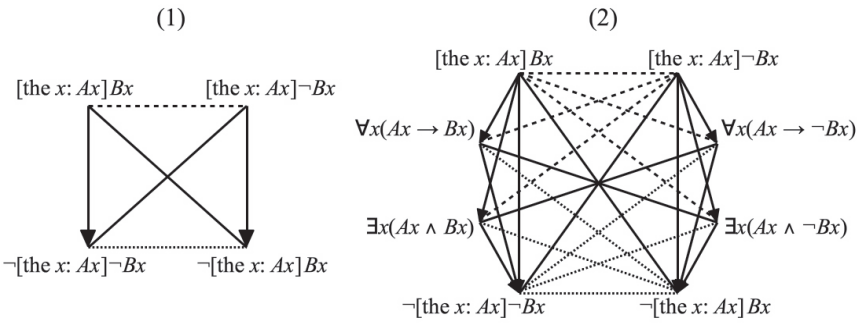
In addition, I argue that my account is superior to Predicativism as well as other referentialist's views, in adequately accounting for all the uses of proper names mentioned so far. I claim that proper names are referential singular terms fundamentally, and show that their referential meaning is the base meaning from which every other use is semantically derived.

The Logical Geometry of Russell's Theory of Definite Descriptions

In this paper I will study Russell's famous theory of definite descriptions (henceforth: 'TDD') in terms of the Aristotelian diagrams it gives rise to. Russell analyzed sentences of the form 'the A is B' in terms of the existence, uniqueness and universality conditions. I will first argue that each definite description gives rise to four logically distinct formulas, depending on the scope of the negation(s), and show that these four formulas jointly constitute a classical *square of opposition* (cf. Figure 1 below). The Boolean closure of this square is a so-called *Jacoby-Sesmat-Blanché hexagon*, which can be seen as highlighting the role of the existence and uniqueness conditions in Russell's TDD. Next, I will discuss the connection between the Aristotelian square for TDD and the traditional Aristotelian square for the categorical statements. After arguing that the latter is already implicitly present in the former, I integrate both into a single Aristotelian diagram, viz. a so-called *Buridan octagon* (cf. Figure 2 below), which can be seen as highlighting the role of the universality condition in Russell's TDD. Finally, I will study the exact influence of the existence and uniqueness conditions on Russell's TDD by moving from ordinary first-order logic (FOL) to two stronger logical systems, SYL and SYL*, and showing that this causes the Buridan octagon to turn into a so-called *Lenzen octagon* or to collapse into a classical *square* again.

There thus exists an extensive and multi-faceted interaction between Russell's theory of definite descriptions on the one hand, and logical geometry (i.e. the contemporary study of Aristotelian diagrams as objects of independent interest) on the other. Throughout the paper I will emphasize the fruitfulness of this interaction, both from the perspective of Russell's TDD and from that of logical geometry.

From the perspective of philosophy of language, I will illustrate the great heuristic value of this diagrammatic analysis by pointing out several new insights that are relevant for TDD. For example, constructing a complete square of oppositions leads us to consider the doubly-negated formula $\neg[\text{the } x\colon Ax]\neg Bx$ —in the more usual notation of first-order logic: $\neg\exists x(Ax \wedge \forall y(Ay \rightarrow y = x) \wedge \neg Bx)$. This formula does not seem to have been studied before in the literature on TDD, but I will show that it can be seen as a 'weak' version of the ordinary definite description statement $[\text{the } x\colon Ax]Bx$, and discuss some connections with the recent literature on self-predication principles (i.e. principles of the form 'the A is A'). Furthermore, while the difference between the formulas $[\text{the } x\colon Ax]\neg Bx$ and $\neg[\text{the } x\colon Ax]Bx$ is usually drawn in a strictly *syntactic* fashion (relying on the



notion of scope), I will make use of (i) bitstring semantics and (ii) Seuren and Jaspers's Principle of Complement Selection from natural language semantics to provide a more *semantic* characterization (as negations of $[\text{the } x\colon Ax]Bx$ relative to two different logical spaces). Finally, by studying logical systems in terms of the Aristotelian diagrams that they give rise to, we introduce a new *layer of abstraction*, which might be helpful for drawing connections between logics that *prima facie* have nothing to do with each other; for example, I will highlight an unexpected connection between SYL^* (cf. supra) and public announcement logic, by showing that both logical systems give rise to 'vertically flipped' Aristotelian squares.

On the other hand, from the perspective of logical geometry, I will emphasize that TDD is a rich source of philosophically interesting decorations for a wide variety of Aristotelian diagrams (e.g. the classical square of opposition, but also various types of hexagons, octagons, etc.). These Aristotelian diagrams serve as perfect illustrations of various logical phenomena that are studied more systematically in logical geometry. For example, we will see that a single family of Aristotelian diagrams can have several *Boolean subtypes*: some of the Aristotelian diagrams we construct turn out to have the same Aristotelian properties, but vastly different Boolean properties (e.g. Boolean closure size). Furthermore, we will encounter several manifestations of the *logic-sensitivity* of Aristotelian diagrams: when a single set of formulas (viz. the 4 definite description formulas and the 4 categorical statements) is analyzed in different logical systems (viz. FOL, SYL and SYL^*), it can give rise to Aristotelian diagrams that (i) belong to different Aristotelian families (Buridan octagon vs. Lenzen octagon vs. classical square), (ii) have different Boolean properties (having Boolean closures of size 2^6 vs. 2^5 vs. 2^3), and (iii) are even of different sizes (octagon vs. square).

The final conclusion, therefore, is that despite Russell's severe criticisms of Aristotelian logic, there exists a highly fruitful interaction between one of the cornerstones of precisely this logic—viz. the square of oppositions and its extensions (as studied today in logical geometry)—and Russell's own quintessential contribution to logical philosophy—viz. his theory of definite descriptions.

A direction effect on taste predicates

This paper reports some puzzling experimental results in the domain of predicates of personal taste and offers a tentative explanation of what might be going on. The results concern cases where people change their taste before they assess previous claims about what is tasty and what isn't. It turns out that people's responses significantly depend on whether they go from liking the thing in question to not liking it or the other way around. We suggest that these data can be explained in terms of a flexible form of relativism, according to which sentences like "This is tasty" are polysemous in that they sometimes express what standard forms of relativism predict and sometimes express what standard forms of contextualism predict.

We presented one group of participants the following instructions ("not-like-to-like"):

> **Picture yourself in the following scenario.**
>
> Yumble is a new brand of bubblegum. You have never had a Yumble. One day you decide to try one. You don't like the taste. You tell your friend Paul:
>
> "Yumble isn't tasty."
>
> A few weeks later, you and Paul meet at the check-out in the supermarket. Yumble hasn't changed its taste, but you have now come to like it. You take a pack from the shelf. Paul says:
>
> "That's funny, I have a clear recollection of you saying 'Yumble isn't tasty' last time we met!"

Subjects were then asked to rate how likely they would be to respond either by saying "What I said was true. Still, Yumble is tasty" or "What I said was false. Yumble is tasty." In the scenarios described, subjects significantly preferred the "false" response.

We presented another group of participants the following instructions ("like-to-not-like"):

> **Picture yourself in the following scenario.**
>
> Yumble is a new brand of bubblegum. You have never had a Yumble. One day you decide to try one. You like the taste. You tell your friend Paul:
>
> "Yumble is tasty."
>
> A few weeks later, you and Paul meet at the check-out in the supermarket. Yumble hasn't changed its taste, but you don't like it anymore. When you refuse to buy Yumble, Paul says:
>
> "That's funny, I have a clear recollection of you saying 'Yumble is tasty' last time we met!"

Subjects were then asked to rate how likely they would be to respond by saying either "What I said was true. Still, Yumble isn't tasty" or "What I said was false. Yumble isn't tasty." This time, subjects significantly preferred the "true" response. (We obtained similar results for cases where we replaced "Yumble is tasty" by "Yumble tastes good" and "Yumble isn't tasty" by "Yumble tastes bad.")

These results are puzzling and no extant theory immediately predicts them. Take, for instance, a simple form of contextualism, according to which "X is/isn't tasty" expresses the proposition that

X is/isn't tasty according to the speaker's current taste standard. On such a view, the "true" response should be preferred in both cases. Take, on the other hand, a simple form of relativism, according to which "X is/isn't tasty" expresses a relativistic proposition to the effect that X is/isn't tasty. Whether one judges such a proposition as true or false depends on one's present taste standard. Hence, the "false" response should be preferred in both of the above cases.

We suggest the following view to explain the data: "X is/isn't tasty" sometimes express the suggested contextualist proposition and sometimes expresses the suggested relativistic proposition. A formal implementation of such a position can be found e.g. in (Stephenson, 2007). In the like-to-not-like condition, the contextualist proposition is expressed so that the "true" response is preferred. In the not-like-to-like condition, the relativistic proposition is expressed so that the "false" response is preferred.

Up to this point, the suggested account may seem ad hoc. Even if "X is/isn't tasty" has a relativist and a contextualist interpretation, why do we choose different interpretations in the two conditions described? To explain this, we appeal to familiar accounts of the phenomenon of "negative strengthening" (e.g. (Horn, 1989: sec. 5.3)). Here is a well-attested observation: Suppose I say "It's not clean." This is naturally interpreted as conveying that it's actually dirty. Suppose I say, "It's not dirty." This is naturally interpreted as conveying just that it's not dirty, leaving it open whether it might not be clean either. Why is that so? The basic explanation Horn (1989) and others offer is that, when we want to convey a negative judgement, we often package it in an indirect way so as to appear politer. In the case at hand, we say "It's not clean" instead of "It's dirty" because the former has at least one interpretation were it is less negative than it is meant to be. This suggests the following principle:

> NEG When faced with an evaluative judgment, people tend to choose the least positive of all available interpretations.

If people abide by NEG as a (defeasible) default, we can explain the suggested choice of interpretations in the above "tasty" scenarios.

Consider the case where our subjects start out by saying "Yumble is tasty." On the view proposed, this claim has a contextualist and a relativist reading. The contextualist reading is less positive than the relativist one. For on the contextualist reading, the speaker just speaks about herself. On the relativist reading, she makes a proposal to update the common ground that can be accepted only if everybody likes Yumble (see e.g. (Stephenson, 2007: sec. 5)). By NEG, we choose the contextualist reading. Correspondingly, we end up judging what is said as true later when we've changed our taste standard.

Consider now the case where our subjects start out by saying "Yumble isn't tasty." Once more, this has a contextualist and a relativist reading. This time, though, the relativist reading is less positive. On the relativist reading, the proposition expressed can be added to the common ground only if no one likes Yumble. On the contextualist reading, the speaker is merely giving to understand that she herself doesn't like Yumble. By NEG, we choose the relativist interpretation. Correspondingly, we end up judging our claim to be false later when we have changed our taste standard.

Bibliography

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

Stephenson, T. (2007), 'Judge Dependence, Epistemic Modals, and Predicates of Personal Taste', *Linguistics and Philosophy*, 30/4: 487–525.

Open-textured and Dynamic Meaning as a Ground for a New Metasemantic Framework for Natural Languages.

The idea that languages are shared by speakers is intuitively acceptable for lay persons as well as for many philosophers of language. In fact, we often act in such a way that we (at least seemingly) assume that natural languages are shared – we take Spanish language lessons, we talk about German speaking countries and we consider ourselves to be English speakers. Such a way of talking about natural languages seem to suggest that a language is a set of grammar rules and expressions with stable and determinate meanings which can be learned and shared by members of a community.

In this talk, I will argue that such a view cannot be used as an accurate representation of natural languages because of two reasons: a) meanings of expressions in natural languages are open-textured and b) meanings of expressions in natural languages are dynamic. As I will try to show, if we take a) and b) seriously, then we need to find a new metasemantic framework for natural languages which does not rely on the notion of shared language. A new metasemantic account which can give us a clear idea of what natural languages are, how speakers use them in communication, what do we do when we learn a language and what it means to belong to a linguistic community.

Several authors argued that meanings of expressions in natural languages are open-textured. While there are some differences, their argumentation follows a similar pattern: If you want to determine the meaning of an expression in a natural language, you have to focus on how speakers use the expression in various circumstances. There are always circumstances for which the use of an expression has not been considered and decided before. Therefore, meanings of expressions in natural languages are always underdetermined = meanings of expressions in natural languages are *open-textured*.[1]

Moreover, when the use in the undetermined circumstances has to be decided, the decision is always made by particular speakers in particular situations and the decision depends on the *accidental features* of these situations.[2] The accidental character of these decisions has two important consequences. Firstly, different speakers may make different (even contradictory) decisions as they may not know about previous decisions of other speakers. Secondly, as these decisions depend on accidental features, most of them can hardly be taken as constitutional for meaning and binding for other speakers. As a result, a meaning of an expression in a natural language can vary from person to

[1] The term "open texture" was coined by Waismann (1945) and it has been discussed recently in Gauker (forthcoming) and Shapiro (2006). A similar idea can be found in Wilson (1982). The argument as presented here probably does not fit Shapiro's view as he understands open texture as a kind of vagueness.
[2] Similar arguments can be found in Gauker (forthcoming) and Wilson (1982, 2006).

person, from conversation to conversation. In other words, meanings of expressions in natural languages are *dynamic*.[3]

The idea of a shared and stable language is quite common within the philosophy of language. David Lewis (1969, 1973) depicts language as an outcome produced by speakers conforming to conventions. For Lewis, other speakers' conforming to a convention is a good reason to conform for any individual and this leads to the (almost) overall acceptance of conventions within community. For normative inferentialists,[4] a language is an outcome of interlocking normative attitudes of speakers and normative attitudes harmonize within communities under the pressure of corrective behaviour.

I agree that such a view is common and popular for good reasons. If we have a notion of shared language we can easily explain what makes someone a member of a linguistic community – her sharing of the language; we can easily explain the aim of language learning – to learn the shared language; and we can easily explain why communication can be successful – because we use the pre-established and pre-learned lexicon. However, if meaning in natural languages is open-textured and dynamic, the idea of shared language does not make any sense. There is no stable and determinate lexicon to be shared.

The question is if it is possible to give a coherent view on natural languages which is compatible with the idea of open-textured and dynamic meaning. And I believe there is one. We have to accept the successfulness of communication as a criterion for including someone within a linguistic community. This is a deliberately vague criterion: How successful communication must be to say that someone is a member of a community is a matter of practical decision. In other words, linguistic communities are flexible with very vague boundaries. The successfulness of communication can play also the role of the aim of language learning: There is no final point to be reached, no shared language to master. Language learning is a never-ending process of creating and improving our strategies for (successful) communication in diversified environment. What makes our communication successful is our ability to discuss, modify and agree upon our communicational strategies within conversations. And so what is a natural language after all? Well, "there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed…"[5]

[3] For a similar view supported by many examples of meanings changing within conversations see Ludlow (2014).
[4] Brandom (1994, 2000); Peregrin (2014).
[5] Davidson (1986).

References

Brandom Robert (2000) *Articulating Reasons: An Introduction to Inferentialism*.

Brandom Robert (1994) *Making It Explicit: Reasoning, Representing, and Discursive Commitment*.

Davidson Donald (1986) A Nice Derangement of Epitaphs.

Gauker Christopher (forthcoming) Open Texture and Schematicity as Arguments for Non-referential Semantics. In: Sarah-Jane Conrad and Klaus Petrus, eds., *Meaning, Context, and Methodology*.

Lewis David (1969) *Convention: A Philosophical Study*.

Lewis David (1973) Languages and Language.

Ludlow Peter (2014) *Living Words: Meaning Underdetermination and the Dynamic Lexicon*.

Peregrin Jaroslav (2014) *Inferentialism: Why Rules Matter*.

Shapiro Stewart (2006) *Vagueness in Context*.

Waismann Friedrich (1945) Verifiability.

Wilson Mark (1982) Predicate Meets Property.

Wilson Mark (2006) *Wandering Significance: An Essay on Conceptual Behavior*.

Do we hear meanings (and should we care whether we do)?

Upon listening to an utterance in a language she understands, a hearer typically becomes consciously aware of that utterance's meaning (Fodor, 1983; Recanati, 2004; Smith, 2010). Experiences of utterance understanding are what surface to hearer's consciousness in the course of language comprehension (Fodor, 1983; Hunter, 1998; Smith, 2010). They are often illustrated with the following contrast case. Suppose you hear someone uttering a sentence in your native language, for instance "Penguins spend about half of their lives in the oceans". Now suppose everything is the same except that this person is speaking a language that is totally unfamiliar to you. This time, although the utterance expresses the same thing you have no idea what you have been told. The experience of listening to speech in a language you do not know differs from the experience of listening to speech in a language you understand and speak fluently (e.g. Siegel, 2006; Bayne, 2009; O'Callaghan, 2011; Dodd, 2014; Brogaard, 2016). Such *experiences* play a vital role in linguistic communication and they are a crucial basis for the beliefs we acquire as a consequence (Lackey, 2008; Audi, 2013; Peet, 2016). It is, therefore, of great importance to understand the *nature* and the *epistemic role*(s) of such conscious states of language understanding.

Recently the nature of experiences of utterance understanding has been a hotly debated topic. One influential position in this debate is the *semantic perceptual view*, which claims that properties like *having meaning x* can be perceived (Siegel, 2010; Bayne, 2009; Brogaard, 2016). As a result, recent debates in philosophy of language understanding have focused on the question of whether we can *hear meanings*, i.e. whether we can conceive of experiences of understanding as *perceptual*. Visual experiences of seeing shapes and colours are paradigm examples of perceptual experiences. Several features of experiences of language understanding suggest that they are interestingly similar to such paradigm perceptual experiences (Hunter, 1998; Fricker, 2003; Dodd, 2014). First, experiences of language understanding need not be veridical: a subject can have illusory meaning experiences. A classic example of meaning-illusion is found in Putnam's scenario involving an ant that (accidentally) traces a line in the sand which has the shape WINSTON CHURCHILL (Putnam, 1981). We can't help but experience it as a written proper name. Analogous examples involving full utterances are easy to imagine (e.g. Azzouni, 2013; p. 92). Moreover, hearers typically have no control over whether in listening to (or reading) an utterance they have the experience of understanding it or not. The involuntary, automatic nature of such experiences suggests that they are probably produced by fast, specialized mechanisms, similarly to the way perceptual experiences are produced (Fodor, 1983). Despite these apparent similarities, some philosophers have criticized the view that experiences of understanding are perceptual by emphasizing dissimilarities between such experiences and perceptual experiences (O'Callaghan, 2015; Prinz, 2011).

Can we *hear meanings,* just as we can see objects, and are experiences of language understanding *perceptual*? We still seem to lack a convincing answer. Importantly, it is unclear how we should answer this question in the light of facts about the psychological mechanisms underlying such states. Most philosophical debates have focused on their phenomenology (e.g. Siegel, 2010; Reiland, 2015) and have proceeded largely in isolation from relevant empirical results. In this paper I will argue that *once we acknowledge several facts about the contents and psychological basis of such experiences,* the likely answer to the question of whether we hear meanings is *negative* one. To support this claim I will develop and discuss two sets of considerations that speak against the idea that we can hear meanings. I will also suggest that the question may be of little importance and that the debate should focus instead on the important issue of their epistemic role(s).

First, I will investigate whether meaning is the kind of thing that could be perceived. I will propose that on several influential accounts of meanings (e.g. as abstract structured objects, Heim & Kratzer, 1998; as speakers' complex intentional states, Grice, 1989) the answer to this question is likely to be 'no'. The views that we perceive speakers' mental states or abstract entities are equally controversial. Arguably, if the claim that experiences of understanding are perceptual is to have any substance, its proponents need to provide an account of the objects that would be perceived in language understanding. Without it we have little reasons to take seriously the claim that we hear meanings.

Second, I will investigate whether the kind of process that delivers experiences of language understanding is likely to be described as *perceptual*. I will provide empirical evidence suggesting that the psychological basis of experiences of utterance understanding goes well *beyond* perceptual mechanisms such as those involved in speech perception (Carbonell & Lotto, 2015; Cutler, 2012; Schneider et al., 2014) and speech sound learning (Brogaard, 2016). Drawing on facts about their underlying psychological mechanisms I will propose that experiences of utterance understanding are complex entities involving: non-semantic (phonological, morphological etc.), semantic and pragmatic properties. To this effect, their production involves not only perceptual, but also cognitive processes (Smith, 2010; Lebois et al., 2015). The above evidence concerning psychological mechanisms behind such experiences seems to undermine the idea that they are best conceived as perceptual.

Towards the end of my paper I will briefly discuss one recent proposal for motivating the claim that we hear meanings. Berit Brogaard (2016) has argued that we can become directly aware of meanings through auditory perception because experiences involved serve as *immediate justifiers* for our beliefs about what was said. The suggestion being, roughly, that we can derive the answer to the question about whether we hear meanings from the epistemic role they tend to play. I will argue that epistemological considerations cannot on its own support the claim that we hear meanings. But perhaps we need not settle on whether we actually do? It seems that nothing much depends on whether we hear meanings, so long as we have relevant experiences of language understanding. I will end by suggesting that given the above presented arguments we can leave aside the question of whether we can hear meanings and focus, indeed, on investigating the epistemic role(s) of such experiences, as they figure in linguistic communication.

References:

Audi, R. (2013). Testimony as a social foundation of knowledge. *PPR*, *87*(3), 507-531.
Azzouni, J. (2013). *Semantic perception: how the illusion of a common language arises*. OUP.
Bayne, T. (2009). Perception and the reach of phenomenal content. *Philosophical Quarterly*, *59*(236).
Brogaard, B. (2016). In Defense of Hearing Meanings. *Synthese* ed. L. Moretti & T. Piazza.
Carbonell, K.& Lotto, A. (2015). Speech is not special, again. *Multisensory & sensorimotor interaction in speech perception*, 105.
Cutler, A. (2012). *Native Listening. Language Experience and the Recognition of Spoken Words*. MIT
Dodd, J. (2014). Realism and anti-realism about experiences of understanding. *Phil Studies*, *168*(3).
Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.
Fricker, E. (2003). Understanding and knowledge of what is said. In:*Epistemology of Language*. OUP.
Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar* (Vol. 13). Oxford: Blackwell.
Hunter, D. (1998). Understanding and belief. *Philosophy and Phenomenological Research 58*(3).
Lackey, J. (2008). *Learning from words: Testimony as a source of knowledge*. OUP.
Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*(8), 1764-1801.O'Callaghan, C. (2011). Against hearing meanings. *The Philosophical Quarterly*, *61*(245), 783-807.
O'Callaghan, C. (2015). Speech Perception. In: *Handbook of the Philosophy of Perception*. Oxford.
Peet, A. (2016). Testimony and the epistemic uncertainty of interpretation. *Phil Studies*, *173*(2).
Prinz, J. (2011). The sensory basis of cognitive phenomenology. In:*Cognitive Phenomenology* (174-196). OUP
Putnam, H. (1981). *Reason, truth and history* (Vol. 3). Cambridge UP.
Recanati, F. (2004). *Literal meaning*. OUP.
Reiland, I. (2015). On experiencing meanings. *The Southern Journal of Philosophy*, *53*(4), 481-492.
Schneider, S. et al. (2014). Beyond the N400. *Cortex*, *53*, 45-59.
Siegel, S. (2010). *The contents of perception*. OUP
Smith, B. C. (2010). Speech sounds and the direct meeting of minds. In Nudds & O'Callaghan: *Sounds and perception* (183-210) OUP.

The Nature of the Fundamental Doxastic State

Suppose that Sally is a typical (human) doxastic agent. Like most of us, Sally believes that there is an external world filled with concrete objects. She also believes that the force needed to accelerate such an object is a function of its mass and acceleration, though she is somewhat more confident that there *are* external objects to accelerate than she is in any physicist's equations. She reserves some doubt that physicists have gotten something wrong, but she's highly confident that physicists exist.

I take it that my very partial description of Sally's overall doxastic state should be intelligible to the ordinary English speaker. I haven't used any esoteric jargon, and one doesn't have to be a philosopher to grasp what was said. But note that the description involved attributing to Sally three *prima facie* distinct kinds of doxastic state, distinguishable by the following characteristic forms:

(i) α believes that *P*,
(ii) α is highly [alt: fully / somewhat / a little / etc.] confident that *P*
(iii) α is more [alt: less / just as / etc.] confident that *P* than that *Q*

We can call the kind of state attributed by (i) an *outright belief*. It is perfectly natural, and very standard, to think of outright belief as involving a relation between a subject α and an individual proposition *P*. The kind of state attributed by (ii) we can call *partial belief*. On the face of it, partial beliefs seem to have more structure than outright beliefs: they look like they involve a relation between a subject α, a proposition *P*, and a degree *x*.[1] Finally, the kind of doxastic state picked out by an attribution of the form (iii) belongs to the class of *comparative beliefs*. This class includes states like α's taking *P* to be more likely than *Q*, just as likely as *Q*, at least as likely as *Q*, and so on. And again, comparative belief states seem to have a different structure than outright beliefs, involving this time a relation between a subject α and an ordered pair of propositions (*P*, *Q*).

I think it's fair to assume that Sally has all three kinds of doxastic state. (She may indeed have other kinds of doxastic state which aren't adequately captured by any attribution of the forms (i) through (iii).) We don't gain anything by insisting that one or more of these kinds of state don't *exist*; each seems like it could be used to provide an interesting description of how Sally takes the world to be. However, it's also entirely fair to think that Sally may have some of these states *purely in virtue of* having some of the others. There is no need to posit multiple distinct systems of doxastic representation all situated in Sally's head where we can get by with one, and a parsimonious account of the doxastic attitudes as a whole would plausibly posit a single *basic* or *fundamental* kind of state from which the others could be systematically derived. The question to ask, then, is which—if any—of the three kinds of doxastic state that we've distinguished is fundamental?

A lot has been written about the relationship between partial beliefs and outright beliefs. On the one hand, there are compelling reasons to treat partial beliefs as more fundamental than outright beliefs. Partial beliefs are better placed than outright beliefs to account for the normative role of logic in reasoning (Christensen 2004); they help us to characterise the role of graded epistemic modals in communication in a way that outright beliefs seemingly cannot (Yalcin 2012); they are central to contemporary work on rational decision-making, and equally central to empirically successful models of (amongst other things) vision, language processing, categorisation behaviour, causal reasoning and other forms of common-sense reasoning (e.g., Anderson 1991; Chater and Manning 2006; Griffiths 2007; Kemp and Tenenbaum 2009; Griffiths and Tenenbaum 2011); and the additional structure that partial beliefs appear to have generate well-known problems with proposals for reducing partial beliefs to outright beliefs (e.g., beliefs about chances). On the other hand, there are deep and still unresolved problems with 'threshold' theories for reducing outright beliefs to partial beliefs.

This paper does not address the relationship between partial belief and outright belief. Instead, it looks at the relationship between comparative belief and partial belief. Amongst Bayesians, the (largely unspoken) orthodoxy has for decades been that comparative beliefs are more fundamental than partial beliefs. Call this position *ordinalism*; its primary contrast is with *cardinalism*, the position that

[1] We can remain somewhat neutral about the character of the degrees *x*. In orthodox Bayesian theories, it's assumed that *x* can take any value in the [0,1] interval of the reals (and nothing else). But this assumption doesn't seem to be baked into folk theorising about the mind, and it's certainly not universally adopted in empirical psychology. What matters for present purposes is just that the degrees—whatever they may be and however they may or may not be numerically represented—can be placed into some non-trivial ordering.

partial beliefs are more fundamental than comparative beliefs. According to ordinalism, it is possible to give a full and complete specification of a subject's doxastic state entirely in terms of comparative beliefs. When we attribute partial beliefs to a subject, we are ultimately describing her as having a (total) comparative belief structure which satisfies a certain property—e.g., for Sally to be *fully confident* that *P*, for instance, is for Sally to have comparative beliefs such that *P* is ranked at least as high as everything else, and just as high as an obvious tautology like *if P then P*.

Despite its venerable status, it is hard to find thorough discussion regarding the relative merits of ordinalism and cardinalism. Historically, ordinalism was motivated by a thorough-going behaviourism, of the form that no one still holds today. In the contemporary literature, proponents of ordinalism are generally much more concerned with showing that it's *possible* to reduce partial beliefs to comparative beliefs (under certain precisely specified conditions), but very little has been said about *why* we should want to. Moreover, cardinalism has strong *prima facie* appeal. If partial beliefs can be represented using real numbers—as they are on the vast majority of models—then once you fix all of the facts about a subject's partial beliefs, you *ipso facto* fix all of the facts about her comparative beliefs. From partial beliefs, you get comparative beliefs for free. So why would anyone be an ordinalist?

After setting up the debate, I will develop a novel line of argument in favour of ordinalism, which undercuts the simple case for cardinalism just outlined. The argument begins with the observation that any *minimal* specification of a subject's partial beliefs will force her comparative beliefs to satisfy certain basic structural conditions—conditions which, *arguably*, they need not satisfy. If so, then it is possible for a subject to have comparative beliefs in the absence of partial beliefs, implying that the latter cannot be more fundamental than the former. Given the additional observations that one cannot have partial beliefs in the absence of comparative beliefs, we have enough to rule out cardinalism as a viable answer to the basic question of fundamentality.

As we will see, the strength of the argument rests on just how *minimal* a specification of a subject's partial beliefs can be: the fewer the constraints we place on such a specification, the harder it will be to argue that α's comparative beliefs can conceivably fail to satisfy the requisite structural conditions. Drawing on the work of Lewis and Davidson, I conclude with some reasons to be optimistic about ordinalism.

Anderson, J.R. 1991. The Adaptive Nature of Human Categorization. *Psychological Review*. **98**, p409.

Chater, N. and Manning, C.D. 2006. Probabilistic Models of Language Procesing and Acquisition. *Trends in Cognitive Sciences*. **10**, pp.335-344.

Christensen, D. 2004. *Putting Logic in its Place: Formal Constraints on Rational Belief.* Oxford University Press.

Griffiths, T.L. 2007. Topics in Semantic Representation. *Psychological Review*. **114**, p211.

Griffiths, T.L. and Tenenbaum, J.B. 2011. Predicting the Future as Bayesian Inference: People combine prior knowledge with observations when estimating duration and extent. *Journal of Experimental Psychology: General*. **140**, p725.

Kemp, C. and Tenenbaum, J.B. 2009. Structured Statistical Models of Inductive Reasoning. *Psychological Review*. **116**, pp.20-58.

Yalcin, S. 2012. Bayesian Expressivism. *Proceedings of the Aristotelian Society*. **112**, pp.123-160.

It's a *wonder*-ful world. On our language about the attitudes

1. Propositionalism and sententialism

It is widely agreed that 'that'-clause attributions such as

(1) Robinson Crusoe believes that Poll is a parrot

(2) Robinson Crusoe fears that the cat is on the mat

are of the form *R(a,b)* and express the holding of a relation: a relation of belief between Robinson Crusoe and the denotation of the 'that'-clause 'that Poll is a parrot' in (1), a relation of fear between Robinson Crusoe and the denotation of 'that the cat is on the mat' in (2), and so on. Since the same 'that'-clause can be the object of the various different attitude predicates, as shown by the fact that we say that Robinson Crusoe believes what Friday knows, or that Robinson Crusoe will confess something Friday will find hard to believe, it is also widely agreed that 'that'-clauses always denote entities of the same kind, whatever they are, no matter in the context of what predicate they occur. In this talk, we will assume, in good company, that this is all correct.

The question we will focus on is the one that immediately arises from these assumptions, i.e. what these entities denoted by 'that'-clauses are. Traditionally, two main options have been put forward: propositions and sentences. Because propositions are the meanings of sentences, the two options are then to take the relations expressed as holding by sentences such as (1) and (2) to be either toward something that has meaning (sentences), or toward something that is meaning (propositions).

2. How to choose?

Each of sententialism and propositionalism has its problems but, as I will show, none of the usual criticism is conclusive so that, with Thomason, I will urge that "philosophical debate on the semantic status of complements of verbs such as *believe* has produced something like a stalemate" (1975: 235). How to adjudicate between the two accounts, then? Thomason suggested that since we are in a stalemate, "it may be well to consider evidence from fresh quarters that bears on this issue" (ibid.). The purpose of this paper is to follow Thomason's wise suggestion and in order to make progress in the old debate as to what 'that'-clauses denote, we will consider the evidence that comes from 'wh'-clauses and the attributions they occur in, such as

(3) Robinson Crusoe knows whether Poll will be back

(4) Robinson Crusoe wonders what Poll is thinking.

3. The evidence from the fresh quarters

We will first see why 'wh'-clauses and 'wh'-attributions can indeed count as fresh quarters that bear on the issue of what 'that'-clauses denote. Put differently, we will see that a homogeneous treatment of 'that'- and 'wh'-clauses is natural, very plausible and methodologically welcome. According to this homogeneous treatment, according to sententialism 'wh'-clauses denote questions intended as linguistic items, while propositionalism takes 'wh'-clauses to denote answers, which are propositions.

We will then see that there are good reasons to think that 'wh'-clauses denote questions, as in accordance with sententialism:

(i) the ascriber might not know the answer;

(ii) there might be no correct answer;

(iii) in some cases there might not be enough correct answers;
(iv) in other cases there might be too many answers.

4. Conclusion

Many discussed 'that'-attributions, such as

(1) Robinson Crusoe believes that Poll is a parrot.

Many discussed propositionalism for 'that'-clauses. A few discussed 'wh'-attributions, such as

(3) Robinson Crusoe knows whether Poll will be back.

This is surprising. In the end attributions like (3) are as common as attributions such as (1). A few discussed sententialism for 'that'-clauses. But nobody discussed sententialism for 'wh'-clauses. This is a real shame. First of all, because sententialism seems to be the account on the right track for 'wh'-attributions. Secondly, we cannot prove once and for all that 'that'- and 'wh'-clauses should be treated homogeneously but, as we will have seen, the extensions of sententialism and propositionalism to 'wh'-clause attributions we will have sketched are the natural results of how the two accounts treat 'that'-clauses. Therefore, even if not necessarily or automatically, the fresh quarters of 'wh'-clause attributions and of the attitudes those attributions attribute seem to show that sententialism might be the account on the right track also for 'that'-clause attributions. The old, famous debate about what 'that'-clauses denote then seem to have a winner, and the winner is, quite unexpectedly, sententialism.

References

Bach, K. Questions & Answers (ms.).
Boër, S.E. 1978. 'Who' and 'Whether': Towards a Theory of Indirect Question Clauses. *Linguistics and Philosophy*
Braun, D. 2006. Now You Know Who Hong Oak Yun is. *Philosophical Issues*
Farkas, K. 2016. Practical Know-Wh. *Noûs*
Forbes, G. forthcoming. Content and Theme in Attitude Ascriptions. In A. Grzankowski, Montague, M. (Eds). *Non-Propositional Intentionality*
Groenendijk, J. & Stokhof, M. 1982. Semantic Analysis of *Wh*-complements. *Linguistics and Philosophy*
Hamblin, C.L. 1958. Questions. *Australasian Journal of Philosophy*
Karttunen, L. 1977. Syntax and Semantics of Questions. *Linguistics and Philosophy*
Lewis, D.K. 1982. 'Whether' Report. In Pauli, T. (Ed.). *Philosophical Essays Dedicated to Lennart Åqvist on his Fiftieth Birthday*
Masto, M. 2016. Knowledge -*the* and Knowledge –*wh*. *American Philosophical Quarterly*
Prior, A.N. 1971. *Objects of Thought*
Stanley, J. & Williamson, T. 2016. Skill. *Noûs*
Stout, R. 2010. What You Know when You Know an Answer to a Question. *Noûs*
Thomason, R.H. 1975. Necessity, Quotation, and Truth: An Indexical Theory. *Philosophia*

An Action-Based Theory of Picture Perception

On the one hand, a crucial result of philosophy of perception has been the avowal of the presence of crucial basic relations between vision and action (henceforth: BVS): (1) Vision is crucial for action guidance (Gibson 1979/1986; Milner and Goodale 1995/2006); (2) Vision is a form of action: movement leads the observer to visually appreciate structured and familiar patterns of change in the sensory stimulation with respect to the way this movement is performed (O'Regan and Noë 2001; Noë 2004). (1) recalls the notion of 'vision-for-action', which subserves the perception, conscious (Nanay 2013; Ferretti 2016b; Briscoe 2009) or unconscious (Brogaard 2011), of action possibilities offered by objects (Ferretti 2016a, 2016b, 2016c); (2) is related to the notion of 'sensorimotor understanding', which can also be conscious or unconscious (Noë 2004). From the BVS we can build several 'action-based theories of perception' (Briscoe and Grush 2015). On the other hand, one of the most important enterprises of philosophy of perception is to define what perceptual state we are in during picture perception (Nanay 2011, 2017; Hopkins 2012; Ferretti 2016a, 2016c) and what makes it different from face-to-face perception. Unsurprisingly, the literature on picture perception and the one about the BVS remain completely separated – with some brief exception (Ferretti 2016b, 2016c; Nanay 2010a, 2011a): there is a clear sense in which object in pictures are not salient for action. Here I argue that pictures and action are strictly related. Here is the rationale. First of all, we know that: (a) Perception is linked to action – from (1) and (2). And that: (b) Pictorial representations are perceptual (Nanay 2011, 2015, 2017; Ferretti 2016a, 2016c, 2017). (a) and (b) lead to the following question: if perception – not only its development, but also its exercise (Noë 2004) - grounds on action, how can we account for picture perception without mentioning the BVS? Here I develop the first '*action-based theory of picture perception*' (henceforth: ABP), which is able to recollect these two crucial research fields in the current philosophy of perception and to explain why '*the pictorial needs the motoric*'. Note that (1) and (2) are related to the following two ideas: (3) Vision is always *egocentric* - it is related to the observer's point of view (see Briscoe 2009). And, (4) Action guidance (cfr. 1) is performed toward objects presented within the observer's *peripersonal* action space (see Ferretti 2016a, 2016b, 2016c). (3) and (4) are crucially related to (1) and (2). I first address this relation (§1). Then, I suggest (§2) that the BVS are crucial to explain how picture perception is possible and why it is different from face-to-face perception. Thus, an account of picture perception cannot be conceived apart from the BVS. And this explains why we should go for an ABP. Before cashing out this account, I specify some background assumptions. Picture perception is given by the *simultaneous* occurrence of (i) perception of a surface (S) and (ii) perception of the depicted object (D). Since evidence suggests that both (i) and (ii) can be unconscious (Nanay 2011, 2017), and since, in everyday life, most of the time, (i) is unconscious, while we are visually consciously focused only on (D) (Nanay 2011, 2017), we can endorse that, in usual picture perception (i) is usually unconscious while (ii) is usually conscious (Ibid.). But an unconscious (i) can influence the way in which we obtain a conscious (ii) (Nanay 2017: 5.1). In (§3) I embed this idea of simultaneity within an ABP, by linking simultaneity to the BVS. My claim is that picture perception is possible because we can perform the representations described by the BVS on (S), but we can't equally perform them on (D): (α) we visually represent that there is a (S) we can act on (1), which is represented in egocentric terms (3) and localized as falling within our peripersonal action space (4), and with respect to which we can have a sensorimotor understanding (2). In (§4), I explain the implications of (α). Of course, in usual picture perception: (β) D can be quasi-egocentrically consciously represented (Briscoe 2009: 447) (3), and unconsciously represented as offering action properties (Ferretti 2016c) (1) because they can be unconsciously localized as apparently falling in the peripersonal action space (Id.) (4). But when (α) is the case, even if (β), there is no risk to be deceived that D is a real, present one, like in trompe l'oeil-like situations. When (α) is the case: (α1) there is no sensorimotor understanding of (D), but only of (S) (Nanay 2010): we can change spatial position in relation to (S), but this spatial shift does not reflect any change in (ii) (Nanay 2010, 2011: 471, 2017: 5.1; Hopkins 2012: 657); (α2) action possibilities, as well as peripersonal and *absolute* egocentric localization, can be attributed, at least unconsciously (but also consciously) only to S (Ferretti 2016c). Conversely, when (α) is not the case, because (S) is not visible, (D) looks, to our visual consciousness, like a real, present object we can interact with (cfr. 1) and which falls in our peripersonal action space (cfr. 4) and concerning which we have the impression of sensorimotor understanding (cfr. 2) and *absolute* egocentric localization (cfr. 3) (Ferretti 2016c; Vishwanath 2014). This is what actually happens in the case of trompe l'oeil perception: the representations described by the BVS are related to (D), not (S), because (S) cannot be tracked by our visual system (Id.). In (§5) I point out that, though it is not necessary to explain whether the representations described by the BVS concerning (i) - (cfr. α) - have to be conscious or unconscious, it seems reasonable that they are at least *unconscious*. Indeed, following the idea of *simultaneity* exposed above, during trompe l'oeil perception, not only there is no conscious (i), but we do not have also an unconscious (i): i.e., we cannot rely on any (i). If we had an *at least* unconscious (i), this would equate - following the notion of simultaneity reported above - normal and trompe l'oeil picture perception concerning the behavior of (i) (which is, for both of them, unconscious) and (ii) (which for both of them is conscious). But this would leave unexplained why, nonetheless, only the trompe l'oeil fosters the feeling of presence. So, trompe l'oeil is obtained due to the lack of any (i). Therefore, the representations described by the BVS we need to perform on (i) in order not to fall in a trompe l'oeil situation must be, *at least*, unconscious. And this allows, in usual picture perception, to remain consciously focused on (D), with the possibility to shift consciousness on (S); something not possible with trompe l'oeil (Ferretti 2016c; Nanay 2015). Thus, that an unconscious (i) can influence our conscious (ii) also holds for the representations described by the BVS: the presence of *at least* unconscious representations described by the BVS related to S (α) avoids falling in illusory pictorial situations *à la* trompe l'oeil, in

which our conscious vision is fooled that, in (ii), (D) is real and present. As a result, in usual picture perception, we consciously perceive the possibility of action performance (1), peripersonal (4) and *absolute* egocentric localization (3) and sensorimotor understanding (2) only concerning (S) – which is the unique component perceived as real and present - but not with (D) (Ferretti 2016c); though, again, (β) can coexist with (α). Accordingly, the perceptual representations described by the BVS occurring in (α) can sometimes be conscious (Nanay 2011, 2015, 2017). This is compatible with our empirical informed accounts of picture perception (Nanay 2011, 2015; Ferretti 2016a, 2016b, 2017). Summing up, the at least unconscious representations described by the BVS register the presence of (S) we can perform them on. This is sufficient to avoid the impression of the possibility to consciously perform them on (D). Therefore, the BVS are not only crucial in face-to-face perception, but they are also crucial for us to reach picture perception – and when they are fooled in picture perception, we fall victim of trompe l'oeil illusions. Thus, we need an ABP to explain the peculiar nature of pictorial experience. And while in face-to-face perception they are (only) related to a real object, in picture perception they are crucially related to (S), which visually encodes D (cfr. α) – though they can be, to some minimal extent, also related to D (β)[1].

References

Briscoe, R. 2009. Egocentric spatial representation in action and perception. *Philosophy and Phenomenological Research*, *79*, 423–460.

Briscoe R. Grush R. 2015. Action-based Theories of Perception. In Zalta E.N. (ed.), The Stanford Encyclopedia of Philosophy plato.stanford.edu/archives/fall2015/entries/action-perception/.

Brogaard, B. 2011. Conscious vision for action versus unconscious vision for action? Cognitive Science, 35, 1076–1104.

Ferretti G. 2017. Pictures, Emotions, and the Dorsal/Ventral account of Picture Perception. *Review of Philosophy and Psychology*, DOI 10.1007/s13164-017-0330-y.

Ferretti, G. 2016a. Pictures, action properties and motor related effects. *Synthese Special Issue Neuroscience and Its Philosophy*, 193(12): 3787–3817. doi:10.1007/s11229-016-1097-x.

Ferretti, G. 2016b. Through the forest of motor representations. *Consciousness and Cognition*, 43: 177–196. doi:10.1016/j.concog.2016.05.013.

Ferretti, G. 2016c. Visual feeling of presence. *Pacific Philosophical Quarterly*, doi:10.1111/papq.12170.

Hopkins, R. 2012. Seeing-in and seeming to see. *Analysis*, 72, 650–659.

Milner, A. D. and Goodale M. A. (1995/2006). *The visual brain in action*. 2nd edition, Oxford: Oxford University Press.

Nanay, B. 2010. Transparency and sensorimotor contingencies: Do we see through photographs? *Pacific Philosophical Quarterly*, 91: 463-480.

Nanay, B. 2017. Threefoldness. *Philosophical Studies*, 1–20. doi:10.1007/s11098-017-0860-2.

Nanay, B. 2011. Perceiving pictures. *Phenomenology and the Cognitive Sciences*, 10: 461–480.

Nanay, B. 2015. Trompe l'oeil and the dorsal/ventral account of picture perception. *Review of Philosophy and Psychology,* 6: 181–197. doi:10.1007/s13164-014-0219-y.

Noë, A. 2004. *Action in Perception.* Cambridge, MA: MIT Press.

O'Regan J.K., Noë A., 2001, A Sensorimotor Account of Vision and Visual Consciousness. *Behavioral and the Brain Science*, 24, pp. 939- 1031.

Vishwanath, D. 2014. Toward a new theory of stereopsis. *Psychological Review,* 121(2): 151–178. doi:10.1037 /a0035233.

[1] This work was supported by the 'Fondazione Franco e Marilisa Caligara per l'Alta Formazione Interdisciplinare'.

"Language and Conceptual Development: a Philosophical Review of the Empirical Research"

In the past thirty years a lot of empirical work has been done in order to highlight the relation between language and conceptual organization. There is a common sense assumption according to which language shapes conceptual knowledge, a claim that is weaker than the idea that language shapes thought, but that belongs to the same family. Efforts to demonstrate this hypothesis led the first works on this topic. In particular, a field that seems promising is the one of concepts acquisition and concepts development.

It has been proved that very young infant (prelinguistic) are able to form categories on the basis of the perceptual features of objects. It is possible that language shapes those categories by enabling the grouping together of perceptually distant objects and the separation of similar objects into different categories as well as forming the basis for inferences about hidden properties. At the current state of research there is no consent about the exact role of language in the process of categorization in young children. This paper aims to offer a critique review of the empirical work done so far.

It is widely accepted that there is a facilitative role of naming on categorization and induction in infants within their first year of life (e.g., Dewar & Xu, 2007; Waxman & Booth, 2001); labels do have an impact on category formation in young infants (Waxman, 1999, 2003; Waxman & Booth, 2003; Waxman & Markow, 1995). It is demonstrated that providing 13-month-old infants with the same name for a set of distinct objects (e.g., four different animals) supported the formation of a unified category (e.g., animal). The same effect cannot be achieved if, instead of nouns, the visual stimuli are accompanied by non-linguistic sounds, like tones and buzzers (Fulkerson & Waxman, 2007; Ferry, Hespos, & Waxman 2010). Infants are sensitive to statistical correlations between labels and category instances, but this does not explain the nature of the mechanisms underlying the process.

Two alternative accounts have been advanced to account for the effects of nouns on categorization and induction: supervised name-based categorization and unsupervised feature-based categorization. One possibility, the one suggested by Waxman and colleagues, is that labels facilitate categorization because they act as invitations to form categories by highlighting the commonalities between objects. Labels do impact the process of categorization because different objects are given

the same name and objects named in the same way name belong to the same category. According to this hypothesis, labels are nouns and they have a referential use.

An alternative approach assumes that labels are additional feature-values taken into account by infants during the process of category formation (Sloutsky, Kloos, & Fisher, 2007). On this view labels have the same status as other features, but not names, and are handled in the same manner and as part of the same statistical computation as other features, they are nonsupervisory. Like other features, they may vary in their salience and thereby have a greater or lesser impact on the outcome of computations (Gliozzi, Mayor, Hu, & Plunkett, 2008). Similar distinctions have been found in studies about concept formation in adults (Yamauchi & Markman, 2000).

It seems that there is a double possible role of labelling: one is the proper referential use, the other one is being merely a perceptual feature.

Making sense of self talk

People talk not only to others but also to themselves. The self talk we engage in may be overt or covert, and supports a variety of higher cognitive functions, including reasoning, problem solving, planning and plan execution, attention, and motivation (Winsler 2009). When talking to herself, a speaker borrows linguistic devices from her mother tongue, originally designed for interpersonal communication, and employs them to communicate with herself. This transition from social talk to self talk is reflected in our folk psychology: wondering if it will rain tomorrow, I "ask myself" whether it will rain; I "tell myself" to do the dishes and "promise myself" an espresso when I'm done; I "remind myself" to lock the door when leaving my office; and so on. Asking, telling, promising, and reminding are originally social acts, but when addressed not to others but oneself, they come to function as modes of thinking: wondering, making up one's mind, motivating oneself, and so on. How is this possible?

To answer that question, we need a theory of language and communication that allows us to explain how the same linguistic devices can be used to communicate with others and oneself. According to the received view, language primarily serves to transfer information between people: communication is information exchange. There can hardly be any doubt that this is what language does, and that it is important. However, the information-based view fails to explain the apparent continuity between social talk and self talk, and the developmental transition from the former to the latter; for what could be the point of exchanging information with oneself? Therefore, I propose that instead of focusing on information exchange, communication is better viewed as a form of social interaction in which speakers negotiate commitments between each other. On this account, the primary purpose of social talk is to make commitments to others, while self talk serves to make commitments to oneself.

Viewing commitment as a three-place relation C between two individuals, *a* and *b*, and a propositional content φ, I read $C_{a,b}(\varphi)$ as "*a* is committed to *b* to act in accordance with φ", and hypothesise that every speech act creates a commitment of this form, where *a* is the speaker and *b* the addressee. Promising is the paradigm case. If Don promises Mel to do the dishes, for instance, his speech act brings it about that $C_{\text{Don,Mel}}$(Don will do the dishes). Due to Don's committing himself in this way, Mel becomes entitled to act on the premiss that Don will do the dishes, and thus Don's commitment helps Mel to coordinate her actions with Don's. That's what commitments are for: they are coordination devices. Other speech act types are analysed along the same lines.

Commitments divide into two main categories: telic and factual (cf. Walton and Krabbe 1995). Don's promise creates a telic commitment: it is his goal to do the dishes. Statements create factual commitments. If Don tells Mel, "Vlad sent me such a sweet Christmas card", the content of his commitment is a proposition about the past, but it constrains his *future* actions, for Don is now committed to act in the future on the premiss that Vladimir sent him a Christmas card in the past.

Whereas in social talk, speakers make commitments of the form $C_{a,b}(\varphi)$, where $a \neq b$, in self talk, speaker and addressee coincide, and therefore $a = b$. Thus, self talk creates *private* commitments, whose purpose is to enable the speaker to coordinate his own actions. For instance, by promising himself, "I'll do the dishes", Don makes a commitment to himself to do the dishes, which enables him to coordinate his own actions, for instance, by avoiding to act in a way that might prevent him from doing the dishes.

Like all commitments, private commitments are either telic or factual. Private telic commitments are intentions: if Don is committed to himself to do the dishes, then he intends to do the dishes (cf. Bratman 1987). Private factual commitments are beliefs: if $C_{a,a}(\varphi)$, and a's commitment is factual, then a accepts that φ is the case, and acts accordingly. To illustrate how self talk might give rise to such commitments, consider the following case. Mel is trying to remember where her phone is. She weighs and rejects various possibilities, until she finds herself left with one option only, and concludes: "It's in the kitchen." By telling herself that her phone is in the kitchen, Mel addresses a statement to herself, thus creating a private factual commitment to act in accordance with the premiss that her phone is in the kitchen. That is to say, Mel has formed the belief that her phone is in the kitchen.

To sum up, unlike the received view which defines communication as information exchange, the commitment-based view provides a unified treatment of other-directed and self-directed speech acts. It predicts that addressing speech acts to oneself is a way of creating or renewing private commitments, that is, commitments of the form $C_{a,a}(\varphi)$, which will be either intentions or beliefs. This explains how what is originally a social act can come to function as a mode of thinking.

Although I have contrasted my commitment-based approach to communication with the information-based approach, and argued that the latter is ill-equipped for dealing with self talk, this was not meant to imply that the two views are incompatible. Just as it can hardly be denied that speech acts cause commitments on the part of the speaker, there is little reason to doubt that they carry information between people. If this much is true, the two views are trained on different aspects of communication, and complement each other. Still, I argue that commitments come first, both in communication with others and oneself.

References: Bratman, M. E. 1987: *Intention, plans, and practical reason*. Cambridge, Massachusetts: Harvard University Press. • Walton, D. N. and E. C. Krabbe 1995: *Commitment in dialogue: basic concepts of interpersonal reasoning*. Albany: State University of New York Press. • Winsler, A. 2009: Still talking to ourselves after all these years: a review of current research on private speech. In A. Winsler, C. Fernyhough, and I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation*, pp. 3–41. Cambridge University Press.

Introspection without Judgment

This paper is about introspection of phenomenal states, i.e. the process through which one can form beliefs about the phenomenology of one's current conscious experience. In thinking about this kind of introspective process, we can distinguish the *act* of introspecting from the *product* of that act. Much of the literature on introspection seems to assume that the product of an introspective act is always an introspective judgment, or belief. Accordingly, the debate revolves around the nature and epistemic features of introspective judgments. I suggest that reflection on these questions would benefit from a more fine-grained characterization of the products of introspective acts. We should think of the immediate product of an introspective act (a product is *immediate* iff it does not depend on any intermediary products) as an introspective *state*, where 'introspective state' is a more general and more theoretically neutral notion than 'introspective judgment'. It may be, of course, that all introspective states are judgments. Here, however, I argue that there is a kind of introspective state that is not a judgment. I suggest that two different kinds of introspective state should be distinguished: one conceptual and possibly propositional (and thus potentially judgment-like in important respects), the other nonconceptual and non-propositional (hence not a judgment). I call them, respectively, fact-introspection and thing-introspection. The purpose of this paper is to argue that thing-introspection is a real psychological phenomenon.

The fundamental difference between thing-introspection and fact-introspection is that the former is a conceptual mental state, whereas the latter is a nonconceptual mental state. Fact-introspection requires the deployment of (some of) the phenomenal concept(s) that are associated with the introspected experience: it involves recognizing the introspected experience as a token of a particular experience type. Thing-introspection, instead, is nonconceptual. When you thing-introspect your current conscious experience, you need not possess any phenomenal concept and indeed no phenomenal concept is applied when you merely thing-introspect. Although they often co-occur, thing-introspection does not entail fact-introspection. It follows that thing-introspecting is sufficient for being in an introspective state.

My argument for the existence of thing-introspection is an argument from phenomenal-concept acquisition. Here is the main line of the argument:

(*P*1) (Almost) all phenomenal concepts are acquired.
(*P*2) For most phenomenal concepts, if they are acquired, they are acquired by introspection.
(*P*3) If all introspective states are conceptual, then it is not the case that most phenomenal concepts are acquired by introspection.
(*C*) Not all introspective states are conceptual.

Denying (*P*1) would entail an implausible version of nativism about phenomenal concepts. To be sure, plausible versions of nativism are defended in the philosophical and cognitive-science literature (e.g. Susan Carey's (2009) core cognition theory). However, none of them implies that *a great number* of phenomenal concepts are possessed from birth. Indeed, all they imply is that *a few* conceptual representations are innate.

As for (*P*3), it seems that its denial cannot be an option, on pain of circularity. For if having an introspective state depends on deploying a certain concept, then that concept must be possessed by the subject prior to having the relevant introspective state. Accordingly, it cannot be acquired by way of that very introspective state.

The premise which most is in need of defense is (*P*2). The most natural way to reject it would be to defend a sort of transparency theory of phenomenal-concept acquisition and maintain that phenomenal concepts are acquired not by *introspectively* attending to one's experience, but rather by *perceptually* attending to external objects and their properties. Now, although this transparency account may be viable for the acquisition of phenomenal concepts associated with perceptual experiences, it strikes me as deeply problematic when it comes to other kinds of experiences, such as algedonic experiences, bodily sensations such as tickles and itches, emotions, moods, and

imagination. The transparency theorist of phenomenal-concept acquisition is committed to four increasingly controversial claims:

1) All those experiences are intentional.
2) For any of those experiences, its phenomenal character reduces to its intentional content.
3) Introspection of any of those experiences involves attending not to the experience itself, but to its intentional object.
4) The phenomenal concepts associated with those experiences are acquired by attending not to the experience itself, but by attending to its intentional object.

Claims 1-3 have all received objections. Here are some of them. As for 1, consider algedonic experiences. If it is now a widespread idea that what pain represents is some kind of bodily damage, one may wonder what the representational content of *pleasure* might be, since surely it does not seem to be some kind of *bodily flourishing* (Massin 2013). As for 2, for one thing, there are some aspects of algedonic and emotional experience that do not seem to fit smoothly a representationalist account, namely their affective or evaluative component (Aydede and Fulkerson 2014). Moreover, some experiences, such as moods, simply seem to escape a representationalist reduction (Kind 2014). As for 3, a vast literature has been produced against the transparency account of introspection (Block 1996, Kind 2003, Smith 2008, Siewert 2004) and indeed various arguments in defense of the claim that introspection involves attending to one's experience have been provided (Goldman 2006, Petitmengin 2006, Hill 2009, Siewert 2012). To my knowledge, there are no direct arguments against 4 in the existing literature. However, if credence in either of 1-3 is weakened by the arguments just mentioned, so will be that in 4. Moreover, there seem to be independent reasons to reject 4. Except perhaps perceptual experiences, it seems much more plausible that the concepts we use to think about our experiences are formed by attending to properties of the experience itself, rather than to what the experience is about. Thinking introspectively about the anger I feel is different both from thinking about the person who causes my anger experience and from thinking about the bodily states that accompany my anger. Accordingly, it seems much more plausible that the concept ANGER EXPERIENCE is acquired by attending to the anger experience itself, rather than to the cause of the anger or some bodily states: it would be weird if we thought about the properties of our experiences through concepts that were originally formed by attending to something other than phenomenal properties. I do not take the foregoing to *establish* that a transparency account of phenomenal-concept acquisition cannot be provided. More modestly, I suggest that quite a lot of work should be done to make such an account viable. Accordingly, my overall thesis should be better characterized as the conditional: unless a viable transparency account of phenomenal-concept acquisition is provided, not all introspective states are conceptual.

I conclude that there is a conspicuous number of introspective concepts such that they must be acquired by introspection. On a conceptualist framework, there is no viable way in which such concepts may be acquired. If we are to deny that they are innate, we must accept that they must be acquired via a nonconceptual form of introspection. Therefore, thing-introspection exists.

References

Aydede, Murat, and Matthew Fulkerson. 2014. "Affect: Representationalists' Headache." *Philosophical Studies* 170 (2): 175–198.

Block, Ned. 1996. "Mental Paint and Mental Latex." *Philosophical Issues* 7: 19–49.

Carey, Susan. 2009. *The Origin of Concepts*. Oxford; New York: Oxford University Press.

Goldman, Alvin I. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Hill, Christopher S. 2009. *Consciousness*. Cambridge: Cambridge University Press.

Kind, Amy. 2003. "What's so Transparent about Transparency?" *Philosophical Studies* 115 (3): 225–44.

———. 2014. "The Case against Representationalism about Moods." In *Current Controversies in Philosophy of Mind*, edited by Uriah Kriegel, 113–34. London; New York: Routledge.

Massin, Olivier. 2013. "The Intentionality of Pleasures and Other Feelings, a Brentanian Approach." In *Themes from Brentano*, edited by Denis Fisette and Guillaume Fréchette, 307–337. Amsterdam: Rodopi.

Petitmengin, Claire. 2006. "Describing One's Subjective Experience in the Second Person: An Interview Method for the Science of Consciousness." *Phenomenology and the Cognitive Sciences* 5 (3–4): 229–69.

Siewert, Charles. 2004. "Is Experience Transparent?" *Philosophical Studies* 117 (1–2): 15–41.

———. 2012. "On the Phenomenology of Introspection." In *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar, 129. Oxford: Oxford University Press.

Smith, A. D. 2008. "Translucent Experiences." *Philosophical Studies* 140 (2): 197–212.

Bochum – Submission

On the location of pain

Two views on the nature and location of pain are usually contrasted. According to the first, pain is essentially a central state experience, and its bodily location is illusory, being the result of some central projection. According to the second, pain is a perceptual state, and its location is to be identified with the part of the body in which pain is felt. Against this second view, the cases of phantom, referred and chronic pain have been marshaled: all these cases apparently show that one can be in pain while not having anything wrong in her body. Pain bodily location, then, would be illusory.

I shall argue that the location of pain is to be identified with the perceptual *and* the representational content synchronized and coordinated with each other. Basically, it should be identified with the body and the body map (this encompassing both body image and body schema). A crucial role in such identification is played by demonstratives, indexicals and deictics. In reporting that we feel pain *here*, the deictic directly refers to the bodily part as coinciding with the part as represented. So, pain location is not illusory. An argument against a pure experientialist view is advanced, purporting to show some unviable consequences of that approach. Such argument hinges on having a perceptual and a phenomenal experience of one's body being pierced by a painful stimulus. The purely experientialist view, will have that the content of perception is logically distinguished from the content of experience, leaving unexplainable why we feel pain in the same location in which we see our body to be damaged. The view that I will advocate, on the contrary, insists that the deictics hold together the perceptual and the representational content.

The difficult cases for the perceptual view, phantom, chronic and the like, are explained as effects of the de-synchronization of the perceptual and the representational component, as already advocated by neuroscientist Vilayanur Ramachandran.

QUDs and Context-Sensitivity

The 'question under discussion' (or 'QUD') framework is a novel pragmatic framework that draws upon work in the semantics of questions in order to account for a range of pragmatic phenomena such as the use of prosodic focus in English (Roberts 1996), discourse moves (Ginzburg 1996), and non-sentential utterances (Ginzburg 2012). More recently, the QUD framework has been used to help resolve key issues in the semantics/pragmatics debates: notably Schaffer & Szabó (2013) use the framework to account for the context-sensitivity of knowledge attributions, while Schoubye & Stokke (2015) appeal to the framework in their account of how 'what is said' by a sentence is fixed within a context. In this brief talk, I want to suggest that this extension of the framework to explain context-sensitive content is in fact illegitimate. I will do so by first focusing on Schaffer & Szabó's account of knowledge attributions. One particular commitment of their view is that a knowledge attribution cannot be true if the question under discussion presupposes something false. I present a counter-example where this possibility is realised.

There are possible replies that Schaffer & Szabó could make. However, I will then argue that the limitations of the possible responses available to Schaffer & Szabó highlights a deeper worry for their account, and indeed, any account that attempts to capture context-sensitivity via appeal to the QUD framework. This is that in order for the QUD framework to provide a substantive explanation of our judgements about context-sensitive content, we need a principled account of how the QUD is determined for any given discourse. However, the only plausible account of how this happens appeals to the very same intuitions about meaning that the framework is beng deplyed to explain. So while the QUD framework may provide a useful way of thinking about certain pragmatic phenomena, it cannot be used to capture our intuitions about context-sensitive content in the manner that Schaffer & Szabó suggest.

The untenability of nativism in current research on mindreading

The nativist-constructivist debate constitutes a considerable part of current research on mindreading. A number of researchers (e.g. Peter Carruthers, Evan Westra, and Alan Leslie) are known for their nativist position in the debate and take issue with constructivist views proposed by other researchers, with Henry Wellman, Alison Gopnik, and Ian Apperly at the forefront (Carruthers, 2013, 2015, 2016; Carruthers, Laurence, & Stich, 2006; Leslie, 1994; Scholl & Leslie, 1999). More specifically, Westra and Carruthers (2017) propose a nativistic explanation of Theory of Mind Scale study results that Wellman et al. see as supporting constructivism (Shahaeian, Peterson, Slaughter, & Wellman, 2011; Wellman, 2014; Wellman, Fang, Liu, Zhu, & Liu, 2006; Wellman, Fang, & Peterson, 2011; Wellman & Liu, 2004). While allowing for development of the innate mindreading system, Westra and Carruthers (similarly to Chomsky's universal grammar) base their argumentation essentially on a competence-performance gap, claiming that cross-cultural differences in Theory of Mind Scale progression as well as discrepancies between infants' and toddlers' results on verbal and non-verbal false-belief tasks are fully explainable in terms of acquisition of other, pragmatic, cognitive developments, which are said to allow for an expression of the innately present theory of mind understanding.

The goal of the present paper is to bring together arguments against this nativistic view. The following issues (divided into two main parts) will be discussed and demonstrated to undermine the nativistic conceptions:

(1) The first and main focus of this paper is that the competence model offered by the nativism relies on a misconceived competence-performance gap. Westra and Carruthers (2017) argue that variance in ToM development across cultures stems from performance differences (caused, e.g., by pragmatic factors), with the underlying competence innately present. Such a view makes misguided theoretical assumptions as it wrongly conflates systematic description with explanation; it takes the description of the set of possible performances (what happens or can happen), which is still in need of an explanation (how it happens), to be an actual explanation of those performances (cf. Allen & Bickhard, 2013). More specifically, ToM nativists take the capacity description of belief attribution, which still needs an explanation, and turn it into the explanation itself. Such a move, as Allen and Bickhard (2013) demonstrate, leads to an artificial methodological problem where one seeks to eliminate "mere performance obstacles" so that the reified competence, in this case the desire-belief understanding, can be "expressed." In effect, instead of providing an actual *explanation*, this line of thought only demonstrates that certain *performances* (which they mistake for an explanation) are possible at younger ages. In reality, the actual explanation of these performances seems to be much more complex and culture-dependent than just a simple innate mechanism expressing itself in performance; this paper discusses research on cognitive development across cultures that supports this view (Gut & Cheung, 2017; Carpendale & Lewis, 2006; Gut & Wilczewski, 2015; Kallberg-Schroff M. & Miller J.M, 2014; Kobayashi, Glover, & Temple, 2006; Mayer & Trauble, 2012; Naito, 2014; Nawaz, Hanif, & Lewis, 2014; Pyers & Senghas, 2009).

(2) A set of other arguments against nativism will be presented in the second part of the paper. (a) The concept of innateness is argued to be inaccurate while trying to explain ontogeny; it has been dropped in many biological sciences altogether and many developmental psychologists advocate for doing the same in cognitive psychology. For one, adaptionist stories that cognitive psychologist often invoke are questionable – there are other evolutionary processes than adaptation (see Gould & Lewontin, 1979; Racine, 2013). Second, the neo-Darwinian notion of adaptation clearly adopted by 'nativists' in ToM research is used for population-scale phylogenetic analyses and is argued to be inadequate for accounts of ontogeny (see Carpendale & Wereha, 2013; Racine, 2013). And finally, and most importantly, reality of development is a complex interaction of changing elements that is belied by the simplistic notion of 'the innate' (Carpendale & Wereha, 2013; Lewis, Carpendale, & Stack, 2013; Mameli & Bateson, 2011; Oyama, Smith, Weintraub, & Lewontin, 2000). (b) The purported innate mindreading conceptual system posited by Carruthers ascribes adult-like understanding to infants, ignoring the difference between first- and second-order understanding, between what can be called 'presentation' and 'representation' (cf. Allen & Bickhard, 2013). (c) Advances in neurobiology speak strongly against any conceptual knowledge present early in infancy; the neocortex, where conceptual knowledge finds its correlates, is said to be largely equipotential at birth (Elman, 1996; Mareschal,

2007b, 2007a; Westermann et al., 2007). (d) The nativist interpretations of dishabituation studies are excessively charitable; they extend results of studies done with 15-month-olds to conclusions about innateness, whereas in reality at that age there has been plenty of time for construction of the skill (cf. Carpendale, Hammond, & Atwood, 2013). (e) Looking-time experiment paradigm used in non-verbal false belief tasks that provide the main support for the nativist argumentation has been criticized on methodological grounds (see Allen & Bickhard, 2013; Nelson, 2007; Ruffman & Mele Taumoepeau, 2014; Strijbos & Bruin, 2013). In the light of the presented arguments, nativism in theory of mind research is concluded to be an untenable position.

References

Allen, J. W., & Bickhard, M. H. (2013). Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate. *Cognitive Development*, *28*(2), 96–133. https://doi.org/10.1016/j.cogdev.2013.01.002

Carpendale, J. I. M., & Lewis, C. (2006). *How children develop social understanding*. *Understanding children's worlds*. Malden, Mass., Oxford: Blackwell.

Carpendale, J. I., Hammond, S. I., & Atwood, S. (2013). A Relational Developmental Systems Approach to Moral Development. In R. M. Lerner & J. B. Benson (Eds.), *Advances in Child Development and Behavior. Embodiment and Epigenesis: Theoretical and Methodological Issues in Understanding the Role of Biology within the Relational Developmental System - Part B: Ontogenetic Dimensions* (Vol. 45, pp. 125–153). Elsevier. https://doi.org/10.1016/B978-0-12-397946-9.00006-3

Carpendale, J. I., & Wereha, T. J. (2013). Understanding Common Developmental Timetables across Cultures from a Developmental Systems Perspective. *Human Development*, *56*(3), 207–212. https://doi.org/10.1159/000351276

Carruthers, P. (2013). Mindreading in Infancy. *Mind & Language*, *28*(2), 141–172. https://doi.org/10.1111/mila.12014

Carruthers, P. (2015). Mindreading in adults: Evaluating two-systems views. *Synthese.* Advance online publication. https://doi.org/10.1007/s11229-015-0792-3

Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology*, *7*(1), 141–162. https://doi.org/10.1007/s13164-015-0259-y

Carruthers, P., Laurence, S., & Stich, S. P. (Eds.). (2006). *The innate mind.*

Elman, J. L. (Ed.). (1996). *Neural network modeling and connectionism*. *Rethinking innateness: A connectionist perspective on development*. Cambridge Mass.: MIT Press.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society B: Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Gut, A., & Wilczewski, M. (2015). Is Mindreading a Universal or Culture-Specific Construct? In *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science Volume 1419 of CEUR Workshop Proceedings, CEUR-WS.org, 2015. [doi], s. 499-505* (pp. 499–505).

Gut, A., & Cheung, H. (2017). Theory of Mind in Hong Kong and Poland. (research in progress).

Kallberg-Schroff M., & Miller J.M. (2014). Culture and the development of Theory of Mind. In O. N. Saracho (Ed.), *Contemporary Perspectives in Early Childhood Education. Contemporary perspectives on research in theory of mind in early childhood education* (pp. 343–364). Charlotte, North Carolina: Information Age Publishing.

Kobayashi, C., Glover, G. H., & Temple, E. (2006). Cultural and linguistic influence on neural bases of 'Theory of Mind': An fMRI study with Japanese bilinguals. *Brain and Language*, *98*(2), 210–220. https://doi.org/10.1016/j.bandl.2006.04.013

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Domain Specificity in Cognition and Culture* (pp. 119–148). Cambridge University Press.

Lewis, C., Carpendale, J. I., & Stack, J. (2013). Anticipation and social interaction: Commentary on "Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate" by J. Allen and M. Bickhard. *Cognitive Development*, *28*(2), 159–163. https://doi.org/10.1016/j.cogdev.2013.01.008

Mameli, M., & Bateson, P. (2011). An evaluation of the concept of innateness. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *366*(1563), 436–443. https://doi.org/10.1098/rstb.2010.0174

Mareschal, D. (Ed.). (2007a). *Oxford series in developmental cognitive neuroscience*. *Neuroconstructivism: Perspectives and prospects*. Oxford: Oxford University Press.

Mareschal, D. (Ed.). (2007b). *Oxford series in developmental cognitive neuroscience: / Denis Mareschal ... ; Vol. 1. How the brain constructs cognition*. Oxford: Oxford Univ. Press.

Mayer, A., & Trauble, B. E. (2012). Synchrony in the onset of mental state understanding across cultures?: A study among children in Samoa. *International Journal of Behavioral Development*, *37*(1), 21–28. https://doi.org/10.1177/0165025412454030

Naito, M. (2014). From Theory of Mind to Theory fo Relation: Sociocultural Perspectives on Japanese Children's Social Understanding. In O. N. Saracho (Ed.), *Contemporary Perspectives in Early Childhood Education. Contemporary perspectives on research in theory of mind in early childhood education* (pp. 381–408). Charlotte, North Carolina: Information Age Publishing.

Nawaz, S., Hanif, R., & Lewis, C. (2014). 'Theory of mind' development of Pakistani children: Do preschoolers acquire an understanding of desire, pretence and belief in a universal sequence? *European Journal of Developmental Psychology*, *12*(2), 177–188. https://doi.org/10.1080/17405629.2014.973843

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory / Katherine Nelson*. Cambridge, Mass., London: Harvard University Press.

Oyama, S., Smith, B. H., Weintraub, E. R., & Lewontin, R. (Eds.). (2000). *The Ontogeny of Information*: Duke University Press.

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: evidence from learners of a new sign language. *Psychological Science*, *20*(7), 805–812. https://doi.org/10.1111/j.1467-9280.2009.02377.x

Racine, T. P. (2013). How Useful Are the Concepts "Innate" and "Adaptation" for Explaining Human Development. *Human Development*, *56*(3), 141–146. https://doi.org/10.1159/000350925

Ruffman, T., & Mele Taumoepeau. (2014). When and How Does a Theory of Mind Arise? In O. N. Saracho (Ed.), *Contemporary Perspectives in Early Childhood Education. Contemporary perspectives on research in theory of mind in early childhood education* (pp. 45–63). Charlotte, North Carolina: Information Age Publishing.

Scholl, B. J., & Leslie, A. M. (1999). Modularity, Development and 'Theory of Mind'. *Mind and Language*, *14*(1), 131–153. https://doi.org/10.1111/1468-0017.00106

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*(5), 1239–1247. https://doi.org/10.1037/a0023899

Strijbos, D. W., & Bruin, L. C. de. (2013). Universal belief-desire psychology?: A dilemma for theory theory and simulation theory. *Philosophical Psychology*, *26*(5), 744–764. https://doi.org/10.1080/09515089.2012.711034

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. *Oxford series in cognitive development*. Oxford, New York: Oxford University Press.

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of Theory-of-Mind Understandings in Chinese Children. *Psychological Science*, *17*(12), 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential Progressions in a Theory-of-Mind Scale: Longitudinal Perspectives. *Child Development*, *82*(3), 780–792. https://doi.org/10.1111/j.1467-8624.2011.01583.x

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Westermann, G., Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M. W., & Thomas, M. S. C. (2007). Neuroconstructivism. *Developmental science*, *10*(1), 75–83. https://doi.org/10.1111/j.1467-7687.2007.00567.x

Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, *158*, 165–176. https://doi.org/10.1016/j.cognition.2016.10.021

Abstract– Cause for Concern: Searle's Background Theory and the Context Sensitivity of Thought

Really Radical Contextualism (RRC) is the view that either the truth conditions (content) or extension of a sentence are not determined solely by its meaning. Some information from the context is also required. One proponent of this view is Searle, who seems to focus on the underdeterminacy of truth conditions. Searle claims that the context should be understood as a Background of assumptions that we bring to bear in understanding natural language sentences. What makes Searle "really radical" is that he thinks the Background is necessary for any intentional state to have content. This seems to include thoughts, as well as natural language sentences. It is the claim about thoughts that I will investigate here.

I'll raise two putative problems for this view. The first is that it is not clear what the Background does to help determine truth conditions. The second is that RRC seems to be in danger of a reductio ad absurdum. On the one hand it seems as though we do not have to undergo a process of understanding our own thoughts. There is a risk of a regress if our thoughts undergo a process in order to be understood. If a thought does not determine its own truth conditions, then it may need to be reinterpreted, yielding a new thought. But then this thought will also need to undergo a process of being understood, and so on for each new thought, as the RRC takes it that each thought's content will be underdetermined. So we should avoid postulating a process of understanding our own thoughts. However, Searle claims that thoughts do not determine their truth conditions without a context. This implies that there is a process of understanding thoughts, which can take the truth context into account, which conflicts with the claim that there is no such process.

In response to these problems I suggest that we should understand Searle's view in terms of Perry's concerning relation. An assertion is concerned with what its truth is relative to. The significance of this relation is that what an assertion is concerned with need not be articulated in the sentence that is uttered. In Perry's terminology, it need not be about what it is concerned with. So the Background is taken to cause a given sentence to concern a particular proposition, or possible state of affairs, rather than to be about a proposition, when uttered. This seems to make the Background comparable to a set of parameters for the evaluation of an utterance.

The advantage of such a view is that it avoids having to produce further sentences to take the Background into account. This avoids the threat of a regress, and thereby allows that a process of understanding one's thoughts can take place, without triggering a regress. The process does not result in a new sentence which tries to be about its truth conditions. It also gives us an account of what the Background is doing, namely, functioning as a set parameters with which to evaluate a given utterance. This seems to turn the Background view into a species of relativism.

I will argue, however, that despite avoiding these initial criticisms, the view is not compatible with Searle's RRC. This is because relativism seems to be incompatible with Searle's version of RRC. The relativist view seems to take the truth conditions to have already been determined, whereas on Searle's view it is the truth conditions that need to be determined. For relativists it is just the extension that needs to be determined by the context. So if we should understand RRC as using a Background to affect a concerning relation, then RRC should be understood in terms of the extension being underdetermined by the meaning alone, rather than a focus on truth conditions. In which case, it seems that we have a plausible account of RRC, with just a minor change in emphasis. This, I argue, suggests that the thoughts that we entertain could be context sensitive. Their extension is dependent on the meaning of the terms, the manner in which they are compiled and on the Background.

Deferring to Future Speakers

Temporal externalism is the view that the contents of thoughts and utterances can partly be determined by contingent linguistic and/or conceptual developments that take place after the time of utterance. It is, in other words, a version of semantic externalism, but one that has received considerably less attention in the literature than other externalist views of content. (For a statement of and argument for temporal externalism, see Jackman 1999; for criticism, see Brown 2000, for responses to the criticism see Stoneham 2003 and Jackman 2005.) This paper is in three parts. First, I want to clarify exactly what it would take for temporal externalism to be true. Second, I want to suggest that the example standardly discussed in debates about temporal externalism is perhaps not the most convincing one we can find, and suggest a new example which, I think, gives stronger support for temporal externalism. Finally, I will argue that the clarification given in the first part can be used to resolve a certain *prima facie* problem for temporal externalism.

First: what would it take for temporal externalism to be true? It will not suffice that speakers defer to future speakers in the way that ordinary users of 'water', in 1750, might be taken to have been deferring to future scientists. Here, the role of future speakers was merely instrumental: the content of 'water' was (assuming we go along with Putnam) partly determined by the physical environment, there and then; it is just that no suitable experts were available at the time. For an interesting version of temporal externalism to be true, speakers using some referring expression E should intend to talk about an entity, property, or a kind, and these intentions should be determinate enough for the term to refer. At the same time, however, the speakers' intentions should be less than fully determinate as to which entity, property, or kind, or which *specific type* of entity, property, or kind is in question; the resolution of this matter is left to future speakers. Drawing on a general dispositionalist view of reference recently suggested by various authors (Author's article 1, Author's article 2, Johnson & Nado 2014, Cohnitz 2015), I will argue that such deferential intentions need not be thought of as *conscious* intentions to let future speakers decide the meanings of one's terms. Rather, they can be understood (very roughly) as dispositions to accept and go along with a range of different interpretations for utterances containing the term in question, without contesting them.

If this (suitably precisified) is what would be required for temporal externalism to be true, *is it* true? Can we find plausible examples of cases where speakers in fact do defer to future speakers in the way sketched above? In the second part of my paper, I will argue that the standard example used in discussions of temporal externalism is not the most convincing one we can find, and suggest an improvement. The standard example was adopted from Wilson (1982) by Jackman (1999), and involves "Grant's Zebra", a term introduced around 1820 for a kind of zebra native to Kenya. As things actually turned out, "Grant's Zebra" denotes a *subspecies* of the plains zebra, and its application is based on morphological features. But, according to Jackman, had the historical details been suitable different shortly after the introduction of the name, it might have turned out to denote the *entire species*, the plains zebra (and to be applied based on interbreeding), rather than merely a subspecies. But these two alternative courses of events are exactly identical up to and including the introduction of the name (and for a short time afterwards): therefore we should say, according to Jackman, that the contents of early utterances of "Grant's Zebra" were determined by contingent events that had not yet taken place.

However, I find it psychologically implausible to suppose that the speakers introducing the name really had no clear intentions as to whether the name should apply on the basis of morphological

features or of interbreeding – and, if they did, it appears plausible to see the later developments as changes of meaning (as argued by Brown 2000). Moreover, biological taxa are notoriously problematic as examples in discussions of natural kinds, making the evaluation of the example difficult. In an attempt to get around these problems, I will present a new thought experiment, concerning proper names introduced to denote geographical areas, which in my view provides stronger support for temporal externalism than Jackman's example. In my example, explorers come ashore in new, unexplored territory, and introduce a new name for it. The intend the new name to denote *this land*, without a clear intention that it should denote a whole island rather than a region of an island, or vice versa. In such a situation, I will argue, it is quite plausible to let later developments concerning the exploration of the land (such as the presence or absence of other explorers, the order in which the area is explored, ans so on), including related linguistic developments not settled at the time the name was introduced, have a role in determining the referent of the name. By spelling out in detail two courses of events that could have taken place shortly after the name is introduced, I argue that this example provides stronger evidence for temporal externalism than the case usually discussed.

Finally, a *prima facie* problematic aspect of temporal externalism is that it appears to open the door for meaning skepticism. According to the view, contingent future events can partly determine the contents of my utterances, here and now. But what if events of the requisite kind never take place? For example, if all the explorers in Jackman's example of "Grant's zebra" were killed by a tropical disease shortly after 1820, it follows from temporal externalism that their utterances of the term did not have determinate meaning. Does it follow, then, that any use of a referring expression, here and now, might lack determinate meaning, due to unforeseen future events? In the last part of my paper I will argue that this worry is unfounded. The kinds of deferential intentions specified in the first part of my paper are, apparently, quite uncommon: in the overwhelming majority of cases, the meanings of our expressions are determined by facts and events that have already taken place. Moreover, I will argue that in the cases where we *do* have such deferential intentions toward future speakers, we should in fact *expect* that there are possible future courses of events that would render the meanings of the relevant expressions indeterminate.

References:

Brown, Jessica (2000): "Against Temporal Externalism". *Analysis* 60, 178-188.
Cohnitz, Daniel (2015): "The Metaphilosophy of Language". In Haukioja (ed.): *Advances in Experimental Philosophy of Language*, 85-108. Bloomsbury Academic.
Jackman, Henry (1999): "We Live Forwards but Understand Backwards: Linguistic Practices and Future Behavior". *Pacific Philosophical Quarterly* 80, 157-177.
Jackman, Henry (2005): "Temporal Externalism, Deference, and our Ordinary Linguistic Practice". *Pacific Philosophical Quarterly* 86, 365-380.
Johnson, Michael & Jennifer Nado (2014): "Moderate Intuitionism: A Metasemantic Account". In Booth & Rowbottom (eds.): *Intuitions*, 68-90. Oxford University Press.
Stoneham, Tom (2003): "Temporal Externalism". *Philosophical Papers* 32, 97-107.
Wilson, Mark (1982): "Pradicate Meets Property". *Philosophical Review* 91, 549-589.

Why computational explanations are not mechanistic: An argument from idealization

In recent years, the new mechanistic approach in the philosophy of science has become increasingly dominant, and attempts to expand its application by including other types of explanation in its scope have been made. One such attempt is the idea that computational explanation in cognitive science (and computation in general) is, in essence, mechanistic (Kaplan, 2011; 2017; Piccinini, 2015; Piccinini & Craver, 2011). A few counter arguments tried to rebut this claim (Chirimuuta, 2014; Rusanen & Lappi, 2016; Shagrir & Bechtel, 2017). For the most part, these arguments are based on the literature on explanation. However, a philosophical field that has been mostly overlooked in this discussion on the nature of computational explanation is that of models and modeling, which is still much more prevalent in the context of biology than in the context of cognitive science.

The goal of this talk is to introduce relevant insights from works on models and modeling into the debate on the nature of computational explanation. Specifically, I claim that computational explanation is distinct from mechanistic explanation in yet another aspect – in the idealization its models use. My claim is that computational explanations use a kind of idealization that uncovers formal, mathematical, or information-theoretic properties, while knowingly postulating false assumptions about the target domain in order to track these properties. This idealization, which I dub *formal* idealization, plays a part in delineating explanatory non-causal relations. It is dissimilar to *minimalist* idealization (Weisberg, 2007; 2013), which tracks the core causal factors of a system, and hence can be thought of as characteristic of mechanistic explanations that include idealization. I do not claim, however, that *formal* idealization is unique to computational explanation, but only that the fact that computational modelers use *formal* and not *minimalist* idealization provides a ground to distinguish computational explanation from the mechanistic framework.

Beyond paradoxes and bifurcations: Proposal for a unified theory of metaphor based on speaker's intention

There are two ways to understand the interpretation of metaphor within the contextualist strand of analytic philosophy of language. On the one hand, we as interpreters seem to understand many metaphors on the fly, such as 'on the fly', without any considerable cognitive effort. The theoretical description of this type of interpretation relies on technical terms like ad hoc concept construction (Carston, 2002) or modulation (Recanati, 2004). On the other hand, interpreting more poetic or philosophical metaphors often requires us to exercise our reflective abilities, resulting in interpretive activities that stretch over extended periods of time (Camp, 2006, Wearing, 2014). Take, for instance, the metaphor 'time is a pond in which the past bubbles to the surface' (Christoph Ransmayr). These two apparently different routes of interpretation lead some theorists to propose a bifurcated theory of metaphor (Bowdle & Gentner, 2005; Carston, 2010). Other theorists talk of a paradox of metaphor in view of the fact that most metaphors get interpreted according to the first route, while it is the second one that allegedly fits paradigm cases (Steen, 2008). Moreover, this first route does not seem to be distinctive of metaphor only but also covers cases of loose talk, thus threating the uniqueness of metaphor (Giora, 2008, Sperber & Wilson, 2008).

This paper aims to diffuse both the paradox of metaphor as well as attempts at constructing a bifurcated theory of metaphor. To this effect, it proposes an account according to which being a metaphor is a property of tokens of sentences in a given context. More precisely, what determines whether some sentence token is a metaphor depends on the structure of the intention of the speaker. One and the same sentence type may be uttered with or without such an intention. In the later case, we had better not speak of metaphor in order to avoid equivocation. Whether the relevant speaker uttered some sentence with the intention distinctive of metaphor is an epistemological question of interpretation. Interpreters may err both ways here, taking something to be metaphorical when it isn't, and vice versa. Yet epistemological questions of interpretation do not necessarily settle metaphysical questions of constitution (Neale 2016). Whether some utterance counts as metaphor depends on how it was intended, not on how it is interpreted. This is the conclusion I will argue for.

Bowdle, B. F., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review*, *112*(1), 193–216.
Camp, E. (2006). Metaphor and That Certain "Je Ne Sais Quoi." *Philosophical Studies*, *129*(1), 1–25.
Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
Carston, R. (2010). Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images. *Proceedings of the Aristotelian Society*, *cx*(3), 297–323.
Giora, R. (2008). Is Metaphor Unique? In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 143–160). Cambridge: Cambridge University Press.
Neale, S. (2016). Silent Reference. In G. Ostertag (Ed.), *Meanings and Other Things: Essays in Honor of Stephen Schiffer* (pp. 229–342). Oxford, New York: Oxford University Press.
Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.
Sperber, D., & Wilson, D. (2008). A Deflationary Account of Metaphors. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84–105). Cambridge: Cambridge University Press.
Steen, G. (2008). The Paradox of Metaphor: Why We Need a Three-Dimensional Model of Metaphor. *Metaphor and Symbol*, *23*(4), 213–241.
Wearing, C. (2014). Interpreting novel metaphors. *International Review of Pragmatics*, *6*(1), 78–102.

Generic Disjunctive Syllogism

Abstract

The semantics of bare plural generics is still a matter of much debate. While there are several worked-out proposals for giving the truth conditions of generics, somewhat less attention has so far been paid to the question what structural-logical principles, if any, govern generics. This paper takes a certain class of inferences involving generics, disjunction, and negation as a starting point to investigate some candidate principles. It argues that some of the most comprehensive truth-conditional accounts currently on the market are unable to account for the validity of what I call *generic disjunctive syllogism*.

This paper proceeds under the assumption that inferences such as (A) and (B) have valid readings and that, other things being equal, we want a semantic theory that predicts these validities (note that we are concerned with *validity*, not with *soundness*):

(A) (A1) Elephants live in Africa or Asia.
(A2) Elephants don't live in Africa.
(A3) Therefore: Elephants live in Asia.

(B) (B1) Coin throws come up heads or tails.
(B2) Coin throws don't come up heads.
(B3) Therefore: Coin throws come up tails.

I will rely on the orthodox account of the logical form of bare plural generics, according to which they have a tripartite structure $\mathrm{GEN}(\Delta; \Phi)$, consisting of a covert generic operator GEN, a restrictor Δ and a scope Φ. But even granting this mainstream view, it is a substantial question what logical form (A) and (B) have under their valid reading. First, sentences with bare plural nouns in subject position can receive non-generic interpretations. Second, since the inferences involve disjunction and negation, this gives rise to different options concerning the scope these operators take with respect to GEN. In particular, it turns out to be a contentious matter whether negation in (A2) and (B2) is best construed as having narrow or wide scope. It will be argued (against Nickel (2010)) that (A) and (B) instantiate the following schema:

(S) (S1) $\mathrm{GEN}(\Delta; \Phi \vee \Psi)$
(S2) $\mathrm{GEN}(\Delta; \neg\Phi)$
(S3) Therefore: $\mathrm{GEN}(\Delta; \Psi)$

Thus, we want a semantic theory that accounts for the validity of schema (S).

Note that (S) is closely related to the inference rule of disjunctive syllogism from propositional logic. Claiming that (S) is valid amounts to claiming that we can apply disjunctive syllogism within the scope of generics (keeping the restrictor constant). I will call the corresponding principle for generics *generic disjunctive syllogism* (GDS):

GDS If $\mathrm{GEN}(\Delta; \Phi \vee \Psi)$ and $\mathrm{GEN}(\Delta; \neg\Phi)$ are true, then $\mathrm{GEN}(\Delta; \Psi)$ is also true.

The main concern of this paper is to investigate several proposals for the semantics of generics with an eye to their ability to validate something along the lines of GDS. As it will turn out, all but one account to be considered *fail* to validate GDS in full generality.

It will be instructive to also consider a *restricted* version of GDS. Note that the inferences considered above are all such that the involved scope predicates concern roughly the same topic. For instance, in the case of (A), 'live in Africa or Asia', 'don't live in Africa', and 'live in Asia' all concern *places to live*, or *habitat*. Many semantic theories of generics employ some kind of mechanism to keep track of this. For instance, many accounts assume that predicates are associated with sets of *relevant alternatives*, where predicates such as 'live in Africa' and 'live in Asia' are associated with the same set of alternatives. The restricted version of GDS concerns only cases in which the set of alternatives is *identical* for the predicates both in the premises and the conclusion of the relevant instance of schema (S). Using f_x to denote the set that is associated with a predicate x, we formulate this restricted principle as follows:

GDS_R For all Φ and Ψ such that $f_{\Phi\vee\Psi} = f_{\neg\Phi} = f_\Psi$:
If $\text{GEN}(\Delta; \Phi \vee \Psi)$ and $\text{GEN}(\Delta; \neg\Phi)$ are true, then $\text{GEN}(\Delta; \Psi)$ is also true.

As will be shown, some semantic theories which fail to validate GDS nevertheless manage to validate GDS_R. This is relevant since inferences such as (A) and (B) not only fall under GDS, but arguably also under GDS_R.[1]

The main goal of this paper is to investigate whether various semantic theories of generics are able to account for instances of generic disjunctive syllogism. Two large families of semantic proposals for generics are, first, accounts which assign a central role to some notion of *normality*, and second, accounts which draw on a notion of *probability*. The paper considers three normality accounts (von Fintel (1997), a modification of von Fintel's account, and the account proposed by Nickel (see Nickel (2009), Nickel (2010), and Nickel (2016)) and one probability-account (the one defended by Ariel Cohen, see Cohen (1999a), Cohen (1999b), and Cohen (2004)). As will be shown, of these four accounts, only von Fintel's original proposal is able to underwrite generic disjunctive syllogism in full generality. However, this account is plagued by clear-cut counterexamples. The paper considers a natural modification of von Fintel's original account which is not subject to these counterexamples. As it turns out, both the modified von Fintel account and the account recently proposed by Bernhard Nickel only validate GDS_R, the restricted version of generic disjunctive syllogism, while the general GDS is predicted to be invalid. Cohen's probabilistic proposal can be shown to be unable to make either principle come out correct.

References

Cohen, Ariel (1999a). "Generics, Frequency Adverbs, and Probability". *Linguistics and Philosophy* 22.3, pp. 221–253.

— (1999b). *Think generic! The meaning and use of generic sentences*. Dissertations in linguistics. Stanford, Calif.: CSLI.

— (2004). "Generics and Mental Representations". *Linguistics and Philosophy* 27.5, pp. 529–556.

Nickel, Bernhard (2009). "Generics and the ways of normality". *Linguistics and Philosophy* 31.6, pp. 629–648.

— (2010). "Generically free choice". *Linguistics and Philosophy* 33.6, pp. 479–512.

— (2016). *Between logic and the world: An integrated theory of generics*. First edition. Oxford: Oxford University Press.

von Fintel, Kai (1997). "Bare plurals, bare conditionals, and only". *Journal of Semantics* 14.1, pp. 1–56.

[1] Generic disjunctive syllogism involves both the interaction of GEN and disjunction as well as the interaction of GEN and negation. For this reason, the paper also investigates candidate principles that we could take to govern these interactions. In particular, the principles of *generic disjunction introduction*, *generic non-contradiction*, and *generic excluded middle* will be considered. For reasons of space, I omit discussion of these principles from this abstract.

Forced Choices: Does accepting its consequences suffice for deeming a deed intentional?

In my talk I want to address a hitherto neglected problem in the discussion of the Knobe paradigm (Knobe 2003/2004/2006). I argue that the study as well as follow-up experiments may be flawed, because they have a forced-choice design. While this testing paradigm may have merits in studies on unconscious recognition in perception (Fechner 1860; Posner 1980), it is controversial in cases where subjects deliberately choose between written options, as discussed in the literature on ipsative personality tests (e.g. Kline 2005; Baron 1996).

In order to test my hypothesis, I conducted five follow-up experiments of the Knobe-paradigm in German, varying the dimensions of forced answers. In each setting, I asked 60 to 80 subjects of native German-speaking non-philosophers and first-year students in philosophy with no prior knowledge of experimental philosophy of action theory.

In the first vignette, I roughly replicated the initial finding on the harm story and the help story. However, in the second vignette, subjects were presented with more options such as "the CEO expected harm/help for the environment" alongside the option "the CEO harmed/helped the environment intentionally". Multiple selections were possible. Here, ascription of "intentionally" dropped to about 25 percent in harm cases, and to about 20 percent in help cases, while ascriptions of "knowingly" or "putting up with the consequences of his/her actions" (in German "billigend in Kauf nehmen", which is the standard term for expected side-effects of actions) were at about 80 percent for help and harm cases.

Further, in a third vignette, subjects were given the forced choice between "intentionally" and "putting up with the consequences of his/her actions". Now, ascription of "intentionally" dropped to levels below 10 percent in both, help and harm cases. Together with further vignettes (testing various attitudes of the CEO towards the consequences as well as testing Knobe's "in order to" cases in German), these results can be taken as evidence that the forced-choice paradigm heavily affects the answers of the subjects. Neither did the results indicate significant differences between help and harm stories for the ascription of "intentionally", nor was the term "intentionally" a common choice given rivaling or merely additional options.

If replicated for other languages, at least two lessons can be drawn from the findings. First, the Knobe effect is neither robust nor is it directly related to attributions of moral blame or praise (Knobe 2006). This is congruent with Machery's trade-off hypothesis (2007) and Nichols's and Ulatowski's (2007) claim that the term "intentionally" may be disambiguated depending on subjects' preferences in the context. Instead of explaining the effect as a case of pragmatics, or more narrowly implicature (Adams/Steadman 2004), it appears to be a semantic effect of the forced-choice paradigm. In a way, it is like voting in the presidential election: given only two improper options, subjects choose the lesser evil. Yet from this, no direct inference can be drawn to the structure of the concept "intentional".

Second, the results raise wider methodological concerns, since they call other forced-choice questionnaires into question that are widely used in experimental philosophy. Further experiments are needed in order to see whether the initial findings can be replicated.

Ironic name reference as echoic mention – A pragmatic analysis based on empirical data

Pure quotation, as in *"Paris" has five letters*, is a device used to point to linguistic shapes; see, among others, Cappelen & Lepore (1997), Ginzburg & Cooper (2014), Recanati (2001) for analyses. Instances are constructions involving a name-mentioning predicate like *call* as in *Blood poisoning is also called "sepsis"* or the prenominal modifier *so-called* as in *The doctor diagnosed a so-called "tennis elbow"*. Constructions of this type inform the addressee about the name of a lexical concept as well as its status as a term used conventionally in a certain speech community. Observe, however, that name-mentioning constructions can also adopt an ironic, modalizing interpretation instead of a name-informing one:

(1) a. *We had to get up early every morning. And that's what you call a "vacation"!*
b. *Guess what, the so-called "hotel" turned out to be a run-down dump.*

In our paper, we focus on prenominal *so-called* and propose a pragmatic analysis to account for its different interpretations. Specifically, we argue that the modalizing interpretation illustrated in (1) results from an echoic use of the mentioned name, which, in turn, produces a comment reflecting the speaker's attitude towards the quoted content. Furthermore, to determine the exact nature of the contents involved in ironic name-mentioning, we report on results from two experimental studies.

The verbal root *call-* of *so-called* involves three thematic arguments, an agent *x*, a theme *y* as well as the name *z* of the theme *y*.

(2) *x call- y z*
∃e [CALL(e) & AGENT(x, e) & THEME(y, e) & NAME(z, y, e)]

The function of the *so* in *so-called* is to bind the name argument *z* of the predicate, analogously to the *so* in *One calls this so. So* is a demonstrative anaphor and, here, operates as a pointer to a name's lexical shape – which, in a *so-called*-construction, is provided by the head nominal. In this respect, *so* fulfills the same function as quotation marks that often accompany the name in a name-mentioning context, as both are means to display a linguistic form through demonstration. A demonstration-based analysis of quotational *so* has a natural fit with Davidson's (somewhat unpopular) Demonstrative Theory of quotation (Davidson 1979), whose central claim is that quotation is an operation through which a linguistic shape is referred to by pointing to something that has this shape.

Viewed from the speaker's perspective, the use of a name-mentioning construction in its default function indicates that the mentioned name is believed not to be established in the addressee's lexicon. With this in mind, we claim a relevance-based implicature (e.g., Horn 1984) to be effective in name-mentioning constructions with nominals that are commonly conventionalized. For example, with a highly conventionalized noun like *hotel* as used in (1b) above, a name-informing *so-called* is in fact irrelevant and, in avoidance of a relevance-maxim violation, an ironic interpretation emerges. The degree of conventionalization of a name can be couched as a function of the corresponding nominal's lexical frequency, which, in turn, can be implemented as a factor determining the different readings. This assumption is corroborated by results from a corpus study we conducted using German data. The results indicate that the interpretation of constructions (n = 600) involving *sogenannt* ('so-called') as either name-informing or modalizing indeed interacts systematically with the head's lexical frequency: The higher the lexical frequency of the head nominal, the higher is the probability for the construction to adopt a modalizing function. From a compositional angle, the two interpretations are coupled with different bindings of the agent-argument variable as well as the event variable of the predicate. While both variables can be assumed to be bound generically with the name-informing use, we claim them to be non-generic in nature in the modalizing use.

With a modalizing *so-called*-construction, (i) a non-literal interpretation of the mentioned name is evoked as the speaker asserts himself/herself to oppose its semantic appropriateness. At the same time, (ii) the speaker expresses a (standardly negative) evaluation of the respective denotatum. We will argue these two contents to result from an echoic use of the mentioned name used in some previous utterance

(e.g., by a travel agent in the case of (1b) above). By explicitly marking the utterance as an echo through the use of *so-called*, the speaker produces a comment, which implicates that he/she says something contrary to what he/she means, and that he/she evaluates the denotatum in a certain way, as reflected in the speaker's attitude. Our approach is in line with echo approaches towards verbal irony (see Jorgensen et al. 1984, Wilson 2006), which implement a (negative) evaluation bias as a characteristic feature of ironic language (Kreuz & Glucksberg 1989), and we propose to analyze the evaluative component to be rooted in the "mockery effect" outlined above.

For a detailed classification of the contents involved in modalizing *so-called*-constructions, we conducted two experimental studies in German. Study 1 (SoSci Survey) tested the pragmatic status of the contents and whether these figure as presuppositions or implicatures. We used a pattern as demonstrated in (3) below, for which participants (n = 37) were asked to rate (5-point scale) the coherence of small dialogues and, specifically, the suitability of dissents that follow discourse-interrupting utterances like *Wait a second!* […] targeting the different content types:

(3) A: *We had booked our accommodation before the trip. When we finally arrived at the so-called hotel, we flopped dead tired into bed.*
B: *Wait a second! Are you saying that …*
(a) *your accommodation wasn't a real hotel* / (b) *you didn't like the hotel* / (c) *somebody had referred to your accommodation as a hotel?*
A: *I haven't said that!*

The results indicate that the non-literalness of the name's meaning (= (a) in example (3)) as well as the denotatum's evaluation (= (b)) can be dissented with more easily as compared to the previous use of the name (= (c)); which we take to reflect the latter's status as a presupposition, in contrast to the former two, which we conclude to be implicatures. Study 2 aimed at determining the at-issueness (see, e.g., Gutzmann 2015, Tonhauser 2012, Potts 2015) of the different contents. Crucially, because of the nature of attributive modification, we assume all three contents to represent not-at-issue information, but to different degrees. Participants (n = 56) were asked to choose (5-point scaled choice) between two responses to modalizing constructions of the type in (3), targeting the different contents: an at-issue rejection form (e.g., *That is not true, it is a real hotel*) vs. a not-at-issue rejection (*Wait a second, it is a real hotel*). The results indicate graded at-issueness for the construction's contents such that the negative evaluation component exhibits the lowest degree of at-issueness, followed by the previous name use and then non-literalness, which shows the highest degree of at-issueness. We interpret this result to reflect a central property of verbal irony, i.e., to give rise to non-literal meanings of expressions (see, e.g., Wilson & Sperber 1992).

To conclude, we assume ironic name reference to involve mentioning a name echoically, which produces a non-literal interpretation of the name and signals the speaker's evaluation of the denotatum. Our analysis uses a single underspecified semantic format for *so-called*, with pragmatic factors determining the different interpretations. Viewed globally, our findings contribute to the demarcation between primary and secondary content as well as to a classification of verbal irony in this continuum.

Cappelen, Herman & Ernest Lepore. 1997. Varieties of Quotation. *Mind* 106(423). 429–450. • Davidson, Donald. 1979. Quotation. *Theory and Decision* 11(1). 27–40. • Ginzburg, Jonathan & Robin Cooper. 2014. Quotation via Dialogical Interaction. *Journal of Logic, Language and Information* 23(3). 287–311. • Gutzmann, Daniel. 2015. *Use-Conditional Meaning*: *Studies in Multi-Dimensional Semantics*. Oxford: Oxford University Press. • Horn, Laurence R. 1984. Toward a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature. In Debora Schiffrin (ed.), *Meaning, Form and Use in Context*. (GURT '84), 11–42. Washington: Georgetown University Press. • Jorgensen, Julia, George Miller & Dan Sperber. 1984. Test of the Mention Theory of Irony. *Journal of Experimental Psychology: General* 113(1). 112–120. • Kreuz, Roger & Sam Glucksberg. 1989. How to be Sarcastic: The Echoic Reminder Theory of Verbal Irony. *Journal of Experimental Psychology: General* 118(4). 364–386. • Potts, Christopher. 2015. Presupposition and Implicature. In Shalom Lappin & Chris Fox (eds.), *The Handbook of Contemporary Semantic Theory*, 2nd edn. 168–202: Wiley-Blackwell. • Recanati, Francois. 2001. Open Quotation. *Mind* 110(439). 637–687. • Tonhauser, Judith. 2012. Diagnosing (not-)at-issue content. In: *Under-represented Languages of the Americas (SULA)*, 239–254. Amherst. • Wilson, Deirdre. 2006. The Pragmatics of Verbal Irony: Echo or Pretence? *Lingua* 116(10). 1722–1743. • Wilson, Deirdre & Dan Sperber. 1992. On Verbal Irony. *Lingua* 87(1). 53–76.

The semantic content of imperatives

How, in terms of meaning, should imperatives be distinguished from declaratives? Hanks (2015) argues that imperatives express a different type of proposition to that expressed by declaratives: declaratives express a predication, but imperatives do not. Hanks finds support for this claim in the observation that complement clauses used when reporting assertions are distinct from those used when reporting report directives (declarative vs. infinitive, in English). But, if Hanks were correct that different types of propositions are expressed in each case, then one would expect to find the same patterns in reports of the attitudes associated with each speech act. The data, however, are not as neat as Hanks's picture would lead us to expect. For example, in English 'believe' can take either a declarative or a infinitive complement; in German, 'want' takes a declarative complement; and in Spanish, while the subjunctive appears in reports of directive speech acts, it also appears in negative belief ascriptions, certain factive constructions and after some uses of 'el hecho de que' ('the fact that').

Furthermore, imperatives have a use which is paraphrasable by a conditional sentence (and there is no doubt that these are imperatives, see Jary & Kissine, 2014: ch.3). This suggests that the content of an imperative can fulfil the role of the antecedent of a conditional, and thereby undermines somewhat the claim that imperatives and declaratives express distinct propositional types.

(1) Catch a cold and you'll be off work for a week.
(2) If you catch a cold, you'll be off work for a week.

Despite these issues, we agree with Hanks that an absence of predication in the imperative underlies its fundamental semantic character. That absence, however, we see as resulting from the how the imperative is interpreted *by the addressee*. Interpretations of utterances of imperative sentences by non-addressees, by contrast, consist in the description of a speech act and/or the attribution of an attitude. Such interpretations of the utterance amount to descriptions of actions and ascriptions of attitudes in propositional (i.e. predicative) terms.

Underlying our proposal is a distinction we draw between utterance interpretation in general and a sub-type of interpretation we call 'consumption'. To consume an utterance is to token a mental representation that has both the format specified by the utterance and the same satisfaction conditions as the utterance. Consuming a declarative requires tokening a mental representation in the belief format, while consuming an imperative requires tokening a mental representation in a format whose role in cognition is to trigger and guide action: call these 'action representations'. (That two formats of representation are needed for any system capable of action planning emerges strongly from the work of Millikan (2004)).

Due to their role in cognition, action representations have a number of characteristics that set them apart from belief-format representations. For one thing, an action representation requires no explicit representation of the agent: the individual for whom it serves as a representation will be the agent (by virtue of fact that it is an action representation for that individual). To be sure, an objective statement of the satisfaction conditions of an action representation requires that the agent be

specified, but the representation itself needs no constituent that stands for the agent: the fact that @ is a representation of Mary's makes it that case that Mary is the agent of the act denoted.

One consequence of this is that an action representation need only express a property (rather than a predication), and the agent will judge the representation satisfied only if she attributes that property to herself. Another consequence is that an action representation belonging to one individual cannot have the same satisfaction conditions as an action representation belonging to another: John's action representation cannot have the same satisfaction conditions as Mary's, for John's will require that he ascribe the property denoted/expressed to himself, while Mary's will require that the she ascribe it to herself. Thus action representations are necessarily *de se*: a non-*de se* representation could not function as an action representation.

On this view, then, an imperative will not express a predication. This is a welcome result, given that it has been argued in the syntax literature (e.g. Platzack & Rosengren, 1997, Jensen, 2003, Zanuttini, 2008) that the relationship between subjects and verbs in imperative sentences is such that, unlike in declaratives, the subject does not form a predication with the verb, but serves to specify the addressee. Another consequence is that the contents of action representations cannot be shared. Now, shareability is a crucial characteristic of Fregean thoughts. In our framework, we can say that two individuals share the same thought when each tokens a representation in the belief format with the same satisfaction conditions. But this is not possible in the case of action representations: as we have seen, no two individuals can token action representations with the same satisfaction conditions. We thus uncover a fundamental distinction between the content of belief representations and that of action representations (and hence of declaratives and imperatives): the former is shareable, the latter not.

Another way of putting this last point is to say that, while declarative utterances are for general consumption (in that anyone can, in principle, token a belief-format representation which has the satisfaction conditions of the utterance), imperatives can only be consumed by the addressee. Interpretations of the utterance by non-addressees must take the form of a description of an act directed at a proposition and/or the expression of an attitude towards that proposition. Hence, from the point of view of the addressee, imperatives do not express predications, but non-addressees have no choice but to interpret the utterance of an imperative as directed towards a proposition, one denoting the state of affairs in which the addressee assigns himself the property expressed by the imperative sentence.

Hanks, Peter. 2015. *Propositional Content,* Oxford: OUP.
Jary, Mark & Mikhail Kissine. 2014. *Imperatives,* Cambridge: CUP.
Jensen, Britta. 2003. Syntax and semantics of imperative subjects. *Nordlyd,* 31, 150-164

Millikan, Ruth. 2004. *The varieties of meaning: The 2002 Jean Nicod lectures,* Cambridge, MA: MIT Press.

Platzack, Christer & Inger Rosengren. 1997. On the subject of imperatives: A minimalist account of the imperative clause. *Journal of Comparative Germanic Linguistics,* 1, 177-224.
Zanuttini, Raffaella. 2008. Encoding the addressee in the syntax: evidence from English imperative subjects. *Natural Language and Linguistic Theory,* 26, 185-218.

Confabulation and Rational Obligations for Self-Knowledge

Confabulation as concerns me here is, roughly, the phenomenon whereby subjects incorrectly explain their attitudes when asked why they have them. I argue that it is motivated by the desire to believe we have fulfilled rational obligations to be in a position to know and report certain mental facts about ourselves. My proposed account of confabulation, I suggest, explains it particularly well. This paper's conclusion is important on two fronts. It tells us something about confabulation – how it is brought about – and also self-knowledge – namely, that we bear rational obligations to acquire it.

In §1, I introduce confabulation and set out my explananda for an account of it. The paradigm example of what interests me here is Nisbett and Wilson's 'panty hose experiment'. The experimenters arranged four pairs of identical tights on a table and asked individuals which one they preferred. The majority picked those placed towards the right of the table but when asked why, the participants did not say that they chose it because of the position, or for no reason at all. Instead, they offered reasons like 'its superior knit, sheerness, or elasticity' (Wilson 2002: 103). This paper, then, seeks to explain cases in which subjects express mistaken beliefs about why they have their attitudes when prompted to explain them.

I then introduce and motivate two explananda:

EXPLANANDUM ONE: Confabulators tend to mistakenly self-ascribe putative motivating reasons in particular, that is, they believe there is a reason for which they have the attitude. Such reasons can be contrasted with *purely causal explanatory* reasons which just explain the attitude by reference to causal factors such as the subject's character traits and biases.

EXPLANANDUM TWO: Not only do we tend to mistakenly ascribe motivating reasons to ourselves, but we do so more readily than to others.

§2 considers an influential explanation that we can find in Cassam (2014), Carruthers (2013), Wilson and Dunn (2004), Wilson (2002), and Nisbett and Wilson (1977). It accepts the following: we standardly form self-ascriptions using inferences that are alike those we use with other people apart from some additional evidence. We confabulate, it is thought, because this inference goes wrong in some way. This, though, has trouble accommodating EXPLANANDUM THREE in a simple manner.

In §3, I introduce my own proposal. It supposes that we bear the following rational obligations:

- *Providing-reasons* obligation: the obligation to be in a position to *provide* the reasons for my attitude where these are specifically motivating ones (Anscombe 2000; Boyle 2011a, 2011b; Moran 2001).

- *Knowing-reasons* obligation: the obligation to be in a position to *know* the reasons for my attitude where these are specifically motivating ones (Boyle, 2011a).

I use these to explain confabulation in the following way:

> *We confabulate, and indeed confabulate with the content we do, because we are motivated by our desire to believe that we have fulfilled the* providing- *and* knowing-reasons *obligations*.

We therefore confabulate when we lack an accessible explanation that would enable us to fulfil the obligations, i.e. when we lack a motivating reason. This explanation is compatible with the inferentialist account proposed by Cassam, Carruthers and so on. They could say that we infer the way we do because of these desires. Yet, my explanation still differs from theirs in that the motivational factors do the main explanatory work.

With my proposal in hand, §4 illustrates its explanatory benefits. It explains all three explananda. And further, it does so in a non-adhoc way. That we have these obligations, and that we would desire to believe that we have fulfilled them, I suggest, is already independently plausible.

The paper concludes on the basis of explanatory merit that we bear the *providing*- and *knowing-reasons* obligations and that confabulation is motivated by the desire to believe we have fulfilled them.

Polysemous names

Proper names are usually considered as devices of singular reference. Considered as word-types, however, they also exhibit other kinds of uses (Burge 1973, Fara 2015a,b, Jeshion 2015, 2015a, Elbourne 2005, Leckie 2013, Schoubye 2016, Kijania-Placek 2016). Leckie in (2013) proposed a polysemous account of proper names that accounted for referential and predicative uses of such expressions. In this talk I will propose an account of proper names in terms of a generalized conception of polysemy, which I call rule-based systematic polysemy. The analysis encompasses not just referential and predicative uses, but also anaphoric, bound, deferred and descriptive ones.

The traditional theories of polysemy attempt to account for the multiplicity of stable senses for one linguistic unit, where the sense of a word determines its propositional contribution. Yet the insight that we may glean from the works of Kaplan (1989a,b) concerning the concept of linguistic meaning calls for a generalization of the understanding of the phenomena of polysemous meaning. According to Kaplan, expressions do not necessarily exhibit a meaning that provides content, i.e. propositional contribution, directly, but may instead rely on a rule that for the same word gives (possibly) different contents in different contexts. Combining the ideas of Kaplan with the traditional accounts of polysemy (Apresjan 1973; Pusteyovsky 1995; Pethö 2001), I will propose a two dimensional account of the latter that allows for connecting words not just with sets of stable senses, but also with sets of content generating rules.

I intend to show that the multiplicity of kinds of uses of proper names considered as word-types can be accounted for by the rule-based systematic polysemy. The respective uses can be exemplified by the following sentences:

(1) Referential uses: To know Cecil is to love him.

(2) Predicative uses: There were three Joannas in my class in high school.

(3) Anaphoric uses: If John insists on calling his next son Gerontius, then his wife will be annoyed and Gerontius will get made fun of because of his name.

(4) Bound uses: If a child is christened ‘Bambi’, and Disney Inc. hear about it, then they will sue Bambi’s parents.

(5) Deferred uses: Al Pacino is going to order another Martini but should go home instead. [uttered while pointing at a man resembling Al Pacino]

(6) Descriptive uses: He gave me a Picasso for my birthday.

In the case of generalised, rule-based polysemy we do not expect a set of stable senses determining concrete contents but rather a set of rules that generate contents in contexts. Basing the linguistic meaning of a name on a set of rules - one for each kind of use - will allow for an explanation of both the productivity as well as the systematicity of their uses. Each proper name may thus be used to express a virtually unlimited number of contents but, due to the systematic nature of the underlying mechanisms, the contents are cross-linguistically uniform and predictive. A rules based approach to polysemy thus allows us to account for both its conventional and generative aspects (Recanati, forthcoming).

References:

Apresjan, J. (1973) Regular polysemy. Linguistics 142:5-32.

Burge, T. (1973) Reference and Proper Names. J. of Philos. 70:425–39.

Elbourne, P. (2005) Situations and individuals. MIT Press.

Fara, D. (2015a) Names are predicates. Philos. Review 124:59-117.

Fara, D. (2015b) 'Literal' uses of proper names. In A. Bianchi, (Ed.), On Reference. OUP, 249–277.

Jeshion, R. (2015) A Rejoinder to Fara's "'Literal' Uses of Proper Names", In A. Bianchi, (Ed.), On Reference. OUP.

Jeshion, R. (2015a) Referentialism and Predicativism About Proper Names. Erkenntnis 80:363–404.

Kaplan, D. (1989a) Demonstratives. In: Almog J., Perry J., Wettstein H. (eds.) Themes from Kaplan, OUP, 481–563.

Kaplan, D. (1989b) Afterthoughts. In: Almog J., Perry J., Wettstein H. (eds.) Themes from Kaplan, OUP, 565–614.

Kijania-Placek, K. (2016) Can Minimalism About Truth Embrace Polysemy? Synthese, online first, DOI 10.1007/s11229-016-1228-4

Leckie, G. (2013) The Double Life of Names. Philos. Studies 165:1139–60.

Matushansky, O. (2008) On the Linguistic Complexity of Proper Names. Ling. and Philos. 21:573–627.

Nunberg G. (1993) Indexicality and deixis; Ling. and Philos. 16: 1-43.

Pethö, G. (2001) What is polysemy? In Eniko N., Karoly B. eds. Pragmatics and Flexibility of Word Meaning, Elsevier, 175-224.

Pustejovsky J. (1995) The Generative Lexicon. MIT Press.

Recanati F. (forthcoming) Contextualism and Polysemy.

Schoubye, A. (2016) Type-Ambiguous Names, Mind.

Too-Shifty Indexicals in Answering Machine Puzzle

In Kaplanian semantics, character and circumstances of evaluation determine the truth-value of a sentence. Character is a fixed-rule that determines the meaning of a sentence, ⟦α⟧, by taking context *c* to contents, yielding $⟦α⟧^{c}$. The character of 'I', 'here', and 'now' is a function from the context c to its content – the speaker, location, and time. A world and time <w, t> pair, the circumstances of evaluation which yields $⟦α⟧^{c, <w, t>}$ when combined with $⟦α⟧^{c}$, gives $⟦α⟧^{c, <w, t>}$ a truth-value, depending on which world and time $⟦α⟧^{c}$ was tokened.

According to Kaplan(1989: 492), the content of indexicals – 'I', 'here', 'now' – are context-sensitive: the character determines the content – the direct referents – in different contexts. The character of indexical is preserved within any context, but the content always varies depending on the context.

(1) I am here now.

seems always true because the utterer of (2) is always at the location of utterance at the time of utterance. Then

(2) I am not here now.

is always false. However, if played by an answering machine, (2) sounds intuitively true(Sidelle 1991). Such an intuition raises a problem called Answering Machine Puzzle (AMP).

In this paper, I argue that Eliot Michaelson(2014)'s solution to AMP is uneconomical, possibly ad hoc. Michaelson's reason to motivate Shifty Character View(SCV) to AMP arises from refuting Stevens(2009). Stevens doesn't think that AMP is a problem because "it falls outside the confines of formal semantics"(221): (2) is either plainly false or not an utterance under the context of answering machine. However, Michaelson defends Kaplan(1989)'s aim to provide a theory of sentences-in-context rather than a theory of utterances only confined face-to-face context; a semantic theorist should explain sentences-in-context where the recorded use of indexicals shows observable regularities.

According to Michaelson(2014), our ability to adapt our language to new circumstances gave rise to recorded utterances. Describing indexicals as shifty in different contexts, Michaelson asserts that the references of indexicals should be evaluated relative to the different character-rules. He agrees with Kaplan that the character-rules, which determine the content of indexicals, should remain constant: 'I' refers to the agent, 'here' the location, 'now' the time of the context. Where he differs from Kaplan is *how* to interpret agent, location, and time: "indexicals are to be evaluated at a context of interpretation according to a rule that is itself determined by what sort of communicative channel is used to deliver the message at that context"(528). He suggests three character-rules of indexicals in different contexts:

(3) *Face-to-Face*: 'I' refers to the speaker *s*, 'here' the location of production, 'now' the time of production *t.*

(4) *Recorded(Answering-Machine)*: 'I' refers to the owner of the line, 'here' the location of playback, 'now' the time of playback.

(5) *Recorded(Postcard):* 'I' refers to the author, 'here' the location of production, 'now' time of production *t.*

With these rules, SCV appears to provide the solution to AMP as well as the explanation of a tokened instance of (1) on a postcard. Even though the typology of SCV seems ad hoc, Michaelson replies that ad hockery is virtuous because a semantic theory should responsibly explain our best data, including (a) the intuitions of competent language users on the truth and falsity of sentences and (b) observed regularities of tokened sentences in various contexts.

I present three arguments against SCV. First, I present Multi-Layered Gettier Cases to argue against (a): SCV doesn't reflect the competent language users' intuitions. Second, I present New-Technology Cases to argue against (b): SCV faces individuation problem if it explains every observed regularities in new contexts. Lastly, I present Combined-Context Cases to show that SCV itself is futile.

Firstly, I argue that SCV doesn't reflect the competent language users' intuitions by presenting Multi-Layered Gettier cases. Adrian, who couldn't go to work today from sickness, asks his officemate Bastian to write "I am not here now" on a paper and put it up on his door. Looking for Bastian who happens to be out smoking at the moment, Cathy visits Bastian's office and sees "I am not here now" written on the paper. After Bastian puts the paper up on Adrian's door, it accidently falls. The janitor finds the paper on the floor. Remembering that Derek is out today, she sticks the paper on Derek's door.

Michaelson thinks Multi-Layered Gettier cases shouldn't be dealt with semantically because Bastian's writing (2) is not uttered or played-back. However, I object that the instances of Cathy and Janitor's reading (2) are cases of playback. Applying the rule from SCV for the context for Answering-Machine, 'I' refers Adrian, 'here' the location of Bastian's office, 'now' the time of playback: Adrian is not at Bastian's office when Cathy is reading the note. However, it conflicts with ordinary intuitions as Cathy thinks that Bastian is not at Bastian's office. Since SCV has only provided three context-rules(Face-to-Face, Answering Machine, Postcard) which does not reflect such an instance, a new rule –call it *proxy*– needs to be introduced to reflect Cathy's intuition:

> (6) *Recorded(Proxy):* 'I' refers to the requestor who asked to write the sentence, 'here' the location that requestor should be, 'now' the time of playback.

However, applying (6) to Janitor's intuition (6) means that Adrian is not at Adrian's office when the janitor reads the note. However, the janitor believes that Derek is out when she reads (3). Even after introducing (6), SCV doesn't appear to capture the competent language users' intuitions in Multi-Layered Gettier Cases.

Secondly, I argue that SCV faces individuation problem if it explains every observed regularities in new contexts by presenting three New-Technology Cases. First, I introduce the cases of artificial bot, Siri and Watson, answering (1) to the users. I argue that applying (3), (4), or (5) does not reflect the multi-reliability of artificial bots as referred as "I" in different places, which demands a need for introducing new character-rules for these artificial bots answering (1) or (2). Second, I introduce Skype meeting cases in order to argue that uttering (1) and typing (1) bring out a need for different character-rules. In the context of chatting, there is a gap between when "now" is typed and when it is read, which is not reflected in the character-rules suggested by Michaelson. Third, I introduce avatar cases in the context of game in order to bring out the problem that "here" refers to two different places at the same time due to differences in servers of game players and the differences in location of the soul and the body in case of the death of the character.

Thirdly, I argue that the Combined-Contexts cases present SCV as futile. The first case of Combined-Contexts presents a case where Gina left a voicemail, reading a postcard from her partner which says "I am here now." When the contexts are combined, the purpose of differentiating contexts to typify the unique shifty-nature of indexicals is defeated because (4) can be applied in this different context. The fact that different contexts have the same rules undermines SCV's claim that different contexts reflect different character-rules. Another problem arises in the second case of Combined-Context where "I am not here now" from another answering machine is recorded in an answering machine. In order to reflect such a situation, a complicated, ad hoc character-rule needs to be introduced in order to map "the full extent of [indexicals'] various journeys" as it will prove to be an interesting empirical project as Michaelson(2014: 539) foresees.

By introducing counter-examples to SCV, I have argued that SCV doesn't reflect the competent language users' intuitions, faces individuation problem if it explains every observed regularities in new contexts, defeats the purpose of SCV itself, and introduces uneconomical character-rules.

On the Availability of Presuppositions in Conversation

Abstract

An influential conception of the conversational context models it in terms of the information mutually taken for granted by all interlocutors. Stalnaker's notion of the conversation's *common ground* has been particularly productive for many approaches to the pragmatics of communication (Stalnaker, 1978, 2002, 2014). Yet Stalnaker's account has two deplorable features:

1. Like any account of pooled information – Stalnaker's notion of context suffers from the problem of logical omniscience. For Stalnaker's possible-worlds account, this is the problem of idealizing interlocutors' rational abilities: the total set of interlocutors' joint presuppositions is consistent (if it is to carry any information) and it is deductively closed – that is, any information entailed by the presuppositions is itself presupposed.

2. On Stalnaker's account, all information in the common ground is equally *available*, or *accessible*, to all participants. But while some information – whether it was contributed as at-issue content or non-at-issue content – is available at a given moment in conversation for conversational moves and/or interpretation, other information in the common ground is not. Or so I will try to show.

This paper develops an account of the conversational common ground that improves on these two problems. The basic idea is that the common ground is 'fragmented': the common ground does not form a single consistent, deductively closed set of presuppositions, but is rather organized into a number of fragments, each of which contains presuppositions that together are consistent and deductively closed, but which need not be consistent with beliefs in other fragments[1] In this short talk, I mention the first feature only in passing and focus on the second feature.

A. In the first part of the paper, I outline the second problem for the standard Stalnakerian notion of context. I argue that introducing the notion of *available information* in the common ground helps explain conversational phenomena.

(a) For instance, some information (e.g. that Martin Scorsese is the director of 'Taxi Driver') may be available for answering one question (*Did Martin Scorsese direct 'Taxi Driver'?*) but not for answering another (*Who directed Taxi Driver?*).

(b) A second phenomenon concerns presupposition accommodation in disagreements. In disagreements – say, over the existence of god – speakers are often willing to temporarily accept presuppositions of their opponents. In such cases, these presuppositions are only 'active' and *available*, however, as long as the parties jointly accept one view. Different presuppositions become *available* with the temporary joint acceptance of the other view. Fragmentation helps explain why some presuppositions are available jointly (e.g. all those pertaining to the view that god exists/doesn't exist) at various stages in the conversation.

(c) Asserting information that is already in the common ground can be a felicitous move and can serve a conversational purpose. Stalnaker's model follows Gricean maxims and predicts that it is infelicitous to contribute information that is already mutually accepted. Fragmentation makes room for the fact that it can be felicitous to make information available that is in the common ground but prior to the assertion unavailable.

[1] Cf. Stalnaker (1984, 1991); Lewis (1982); Cherniak (1986); Egan (2008); Elga & Rayo (2014); Rayo (2013) for mental fragmentation.

B. In the second part of the paper, I develop a model of the common ground that makes room for fragmentation. An information state can be called fragmented in case the following theses hold:

F1. Information in an overall fragmented information state is stored in several distinct fragments.

F2. Each information-fragment is consistent and closed under multiple-premise closure.

F3. Fragments of the same overall information state are (somewhat) independent of each other. They may not be consistent with each other, and an entailment of two or more fragments taken together may not be information provided by the overall information state.

F4. Different fragments of the same information state are available for different purposes.

An important and difficult question for fragmentation views concerns the individuation of fragments. In order to make progress on this question for the application to the common ground, I propose to tie the individuation of fragments to questions-under-discussion (Roberts, 1996, 2004; Büring, 2003): pieces of information in the common ground belong to the same fragment in case they provide (partial) answers to a 'strategy' – a complex of questions (related by the *sub-/super-question* relation (cf. Roberts (1996, 2004))). Information in the common ground is available at a given moment in conversation in case it is in what I call the 'live' fragment. It is unavailable at that moment if it is in another fragment of the common ground. I develop this fragmented model of context in formal detail in the talk.

Keywords

Context, Conversation, Common Ground, Presupposition, Question-under-Discussion, Fragmentation

References

Büring, D. (2003). On d-trees, beans, and b-accents. *Linguistics and Philosophy*, *26*(5), 511–545.
Cherniak, C. (1986). *Minimal Rationality*. Cambridge, MA: MIT Press.
Egan, A. (2008). Seeing and believing: perception, belief formation and the divided mind. *Philosophical Studies*, *140*(1), 47–63.
Elga, A. & Rayo, A. (2014). Fragmentation and information access. Unpublished Ms.
Lewis, D. K. (1982). Logic for equivocators. *Noûs*, *16*(3), 431–41.
Rayo, A. (2013). *The Construction of Logical Space*. Oxford: Oxford University Press.
Roberts, C. (1996). Information structure in discourse: Towards an integrates formal theory of pragmatics. In J.-H. Yoon & A. Kathol (Eds.), *Working Papers in Linguistics No. 49: Papers in Semantics* (pp. 91–136). Ohio State University.
Roberts, C. (2004). Context in dynamic interpretation. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 197–220). Oxford: Blackwell.
Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and Semantics 9: Pragmatics* (pp. 315–32; page references are to the reprint in Stalnaker (1999)). New York: Academic Press.
Stalnaker, R. C. (1984). *Inquiry*. Cambridge, MA: MIT Press.
Stalnaker, R. C. (1991). The problem of logical omniscience I. *Synthése*, *83*(3), 425–40.
Stalnaker, R. C. (1999). *Context and Content*. Oxford: Oxford University Press.
Stalnaker, R. C. (2002). Common ground. *Linguistics and Philosophy*, *25*(5), 701–721.
Stalnaker, R. C. (2014). *Context*. Oxford: Oxford University Press.

BILDUNG, SECOND NATURE, AND CONCEPTUAL CONTENT IN MCDOWELL'S MIND AND WORLD

In *Mind and World* McDowell aims to answer the question of how both our experience of the world and the world our experience is about can be conceptually framed through and through, and yet the world can be understood as independent from our empirical judgements. McDowell's approach is that, to answer this question, we need to steer our way between what he calls "bald naturalism" and "rampant Platonism" as two extreme ways to explain our capacity to use concepts and respond to reasons. On the one hand, bald naturalism tells us that the only way to save the elements that are essential to our conception of ourselves is to ground our self-explanation in phenomena which only science can observe and verify. On the other hand, rampant Platonism is a position that construes our reasons as forming an autonomous structure independent from anything specifically human, as far as what is specifically human is also considered as natural.

As a result, McDowell arrives at the ideas of second nature and Bildung. He believes that these ideas, on the one hand, render rationality autonomous without presenting it as something rampantly supernatural, and, on the other hand, keep rationality connected to its natural condition. McDowell needs to reconcile the rational and the natural in order to explain how the answerability of our beliefs to the world that is independent from them is compatible with the notion that the world affects our beliefs and judgments not just causally, but in terms of their *conceptual content*.

In this paper, I claim that the idea of conceptual content, as it is introduced in *Mind and World*, allows two conflicting interpretations which cannot coexist within the framework of McDowell's account. From one point of view, according to McDowell, conceptual contents open us to something that is thus-and-so. The conceptual thus-and-so, as strictly distinct from an *act* of thinking, becomes available in our linguistic exchanges and constitutes a point of agreement between us, no matter how different we might be in terms of our perception and our cultural background. All we need to do, as McDowell suggests, is to go back in our reasoning far enough to reach the contents that sit closest to the impacts of the world on our sensibility. On this account, there is a Wittgensteinian structural analogy between thoughts, experiences and judgments: one may *think that* spring has begun and *report that* spring has begun, and it also may *be that* spring has begun. Given this, McDowell's assumption is that we can always go back to the point where our empirically grounded thinking radically opens us to the world. Although our experiential openness to the world is one of the central metaphors of *Mind and World*, McDowell does not give us too much detail on what the openness actually consists in. His overall strategy is to reject any sort of systematic *theory* of the relationship between mind, world, and conceptual content, and instead to propose a *therapy* capable of disavowing theories that are, for a variety of reasons, unsatisfactory.

From another point of view, as a good Sellarsian, McDowell claims that conceptual contents always come in bundles and that they are something we perpetually reflect upon and change our minds about. We change our minds in the course of *Bildung* – a process defined by McDowell as the transformation of us human beings who are born mere animals into thinkers and intentional agents who partake in second nature. As second-natural creatures, we are initiated into our conceptual capacities within the framework of a particular communal tradition, where conceptual material undergoes perpetual

reassessment. From this perspective, meaning is not so much about being open to factuality, as it is about *doing* something. In this case, what is properly external to individual acts of empirical thinking is not objective factuality, but the social understood as patterns of interrelation between communal practices and environment. Thus understood, conceptual contents form inferential networks which we can never be open to in the way we are, according to McDowell, open to the factual. The reason is that what particular instances of using language reveal within such networks is based on the interplay between explicit statements and norms implicit in discursive practices. We can go back and forth changing the configuration of our commitments to the norms, but we cannot play it all back to the contents that sit closest to the impacts of the world on our sensibility. On this account, as there is nothing between discursive practices and the world that might be called "experience", there is nothing that can relate to the impacts of the world in such way.

I claim that, from the factual perspective, given that institutional, cultural, and normative presuppositions framing our language use vary, McDowell owes us an account of how we can come to an agreement with each other based on contents which are constitutive of second nature and, as McDowell seems to imply, appeal to *any* natural language. If an agreement is based on our openness to matters of facts, we do need to know how we confirm that we are open to the same regions of the conceptual. At the same time, from the second, social-historical perspective, this sort of account is not needed at all, since the normative force pushing us towards an agreement originates from pragmatic know-hows implicit in discourse, which do not presuppose the strong realist commitments that McDowell's factual perspective on conceptual contents seems to require. In this case, the necessary conditions that secure the agreement are wired in the recognition-based discursive practices that we share.

Given the conflict between the two interpretations of the idea of conceptual content in McDowell's *Mind and World*, I suggest that, in order to accommodate both of them within his theory, it makes little sense for McDowell to try to give an account of how Bildung, a process that is essentially social, can possibly open us to the sphere of the conceptual that, according to McDowell, is self-standing and expressly independent of social matters. Instead, McDowell needs to offer an account of how norms and regularities of second nature *emerge* from first-natural laws. In other words, we need a theory that would explain the relationship between language use as a set of social acts and objective conceptual contents and, *in so doing*, would provide some sort of a *genealogy of the normative*. Until such a theory is in place, McDowell's appeal to *second* nature can only restate the problems he ascribed to the idea of the *first* one in new terms.

NORMS OF ASSERTION: EMPIRICAL DATA

1. Introduction

Are assertions governed by *norms*? Differently put, are there rules of linguistic communication that determine whether or not an assertion should be made? Inquiries of this sort presuppose that (i) assertions of different sorts form a unified type of linguistic move or speech act, and (ii) that 'the speech act [of assertion], like a game and unlike the act of jumping, is constituted by rules.' (Williamson, 1996, 489). If we accept these premises, we can ask ourselves what requirements a speaker must fulfill in order to satisfy the norm of assertion. Four proposals dominate the literature (for recent reviews, cf. e.g. Weiner 2007; Pagin 2015; Goldberg 2015):

> *Belief*: Assert that *p* only if you believe that p. (Bach 2008; Bach and Harnish 1979)
> *Justified belief*: Assert that *p* only if you justifiedly believe that p. (Douven 2006; Lackey 2007)
> *Truth:* Assert that *p* only if *p* is true. (Weiner 2005; cf. also Dummett 1959)
> *Knowledge*: Assert that *p* only if you know that p. (Williamson 1996, 2002; Brandom 1998; DeRose 1996, 2002; Adler 2002; Hawthorne 2003; Turri 2011; Benton 2011; for early accounts, cf. Black 1952; Unger 1975; Searle 1976)

The epistemic requirements of the four accounts differ considerably. Whereas on the last two views, the norm of assertion is factive, it is acceptable on the first two accounts to make an assertion that *p* even if *p* is false. Which of the above rules, if any, constitutively defines the speech act of assertion is principally a matter of empirical inquiry – an inquiry that 'must face the linguistic data' (Douven, 2006, p. 450, cf. also Turri 2013, Pagin 2015).

Turri (2013) has presented evidence according to which the norm of assertion is factive; in further works (Turri 2015, 2016) he reports findings in favour of the knowledge account. In my talk, I will demonstrate that most of these experiments are misconceived, and that Gettier cases can play a helpful role in empirically exploring the norm of assertion. I will report a series of studies with a total of about 1500 participants, which suggest that the norm of assertion is neither truth nor knowledge, but justified belief. The following section briefly summarizes one experiment.

2. An Experiment

2.1 Participants, materials and procedure

In an online experiment with 379 participants (164 female, age M=36 years, SD=12 years), participants were randomly assigned to one of three conditions of the well-worn AMERICAN CAR vignette (Nichols et al. 2003): Bob's friend Jill has driven a Buick (an American car) for years. In the knowledge condition K, Jill still drives a Buick. In the Gettier condition TJB, Jill has replaced it with a Pontiac (also an American car), of which Bob is unaware. In the false belief condition FB, Jill now drives – unbeknownst to Bob – a Mercedes. Bob's wife wants to know what car Jill drives. Participants were asked forced-choice questions whether Bob should say *p*, whether Bob really knows that *p*, whether *p* is true and whether Bob is justified in believing that *p* (in that order), where *p* stands for 'Jill drives an American car'.

2.2 Results and Discussion

A logistic regression was run to explore the effects of knowledge, truth, and justification on assertability (Figure 1). The logistic regression model was significant, $\chi^2(3)=33.08$, $p<.001$, explained 21% (Nagelkerke R^2) of the variance in assertability and correctly classified 93.9% of cases. Knowledge proved insignificant (Wald $\chi^2(1)=.006$, p=.940). Justification had a significant impact on assertability (Wald $\chi^2(1)=9.58$, p=.002), and so did truth Wald $\chi^2(1)=13.77$, p<.001). While advocates of a factive account of assertion might consider this a welcome result, a closer look at the FB condition frustrates further expectations. Though only 4% of participants deem *p* true, 84% consider it nonetheless assertable. A second logistic regression explored the effects of knowledge, truth, and justification on assertability in the FB condition. The model was significant, $\chi^2(3)=11.46$, $p=.009$, explained 14% (Nagelkerke R^2) of the variance in assertability and correctly classified 86.2% of cases. The impact of truth on assertability proved insignificant (Wald $\chi^2(1)=.146$, p=.702), and so did the impact of knowledge (Wald $\chi^2(1)=.545$, p=.460). The impact of justification on assertability was significant (Wald $\chi^2(1)=.8.38$, p=.004).

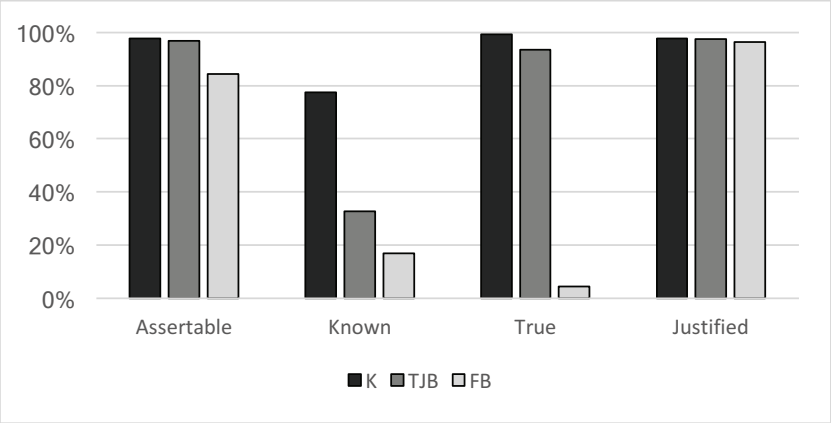


Figure 1: Percentages of participants who deem *p* assertable, known, true, and justified across the three conditions of the AMERICAN CAR scenario.

The experiment suggests that knowledge is not the norm of assertion, and that it isn't factive either. Justification, by contrast, seems to be a strong predictor of assertability. I will present three more experiments using different scenarios, experimental designs and analyses to demonstrate that these findings are robust.

Bibliography

Adler, J. (2002). Belief's own ethics. Cambridge, Mass.: MIT Press.
Bach, K. (2008). Applying pragmatics to epistemology. Philosophical Issues, 18(1), 68-88.
Bach, K., & Harnish, R. (1979). Linguistic communication and speech acts. Cambridge, Mass.: MIT Press.
Benton, M. A. (2011). Two more for the knowledge account of assertion. Analysis, 71(4), 684-687.
Black, M. (1952). Saying and disbelieving. Analysis, 13(2), 25-33.
Brandom, R. (1998). Making it explicit: Reasoning, representing, and discursive commitment. Cambridge, Mass.: Harvard University Press.
DeRose, K. (1996). Knowledge, assertion and lotteries. Australasian Journal of Philosophy, 74(4), 568-580.
DeRose, K. (2002). Assertion, knowledge, and context. The Philosophical Review, 111(2), 167-203.
Douven, I. (2006). Assertion, knowledge, and rational credibility. The Philosophical Review, 449-485.
Dummett, M. (1959). Truth. Proceedings of the Aristotelian Society, 59(1), 141-162.
Goldberg, S. C. (2015). Recent Work on Assertion. American Philosophical Quarterly, 52(4), 365-380.
Hawthorne, J. (2003). Knowledge and Lotteries (Vol. 1). Oxford: Oxford University Press.
Lackey, J. (2007). Norms of Assertion. Nous, 41(4), 594-626.
Nichols, S., Stich, S., & Weinberg, J. (2003). Metaskepticism: Meditations in ethno-epistemology. The skeptics, 227-247.
Pagin, P. (2015). Assertion. In Stanford Encyclopedia of Philosophy.
Searle, J. R. (1976). A taxonomy of illocutionary acts. In K. Gunderson (Ed.), Language, mind and knowledge (pp. 344-369). Minneapolis: University of Minnesota Press.
Turri, J. (2011). The Express Knowledge Account of Assertion. Australasian Journal of Philosophy, 89(1), 37-45, doi:10.1080/00048401003660333.
Turri, J. (2013). The test of truth: An experimental investigation of the norm of assertion. Cognition, 129(2), 279-291.
Turri, J. (2015). Knowledge and the norm of assertion: a simple test. Synthese, 192(2), 385-392.
Turri, J. (2016). Knowledge and assertion in "Gettier" cases. Philosophical psychology, 29(5), 759-775.
Unger, P. (1975). Ignorance: A defense of skepticism. Oxford: Clarendon Press.
Weiner, M. (2005). Must We Know What We Say? The Philosophical Review, 114(2), 227-251.
Weiner, M. (2007). Norms of assertion. Philosophy compass, 2(2), 187-195.
Williamson, T. (1996). Knowing and asserting. The Philosophical Review, 105(4), 489-523.
Williamson, T. (2002). Knowledge and its Limits. Oxford: Oxford University Press.

Mechanisms and representations do not mix

Abstract: In this paper I argue that from a neo-mechanistic standpoint the appeal to representations to explain cognition is vacuous, and consequently, that any mechanistic account of cognition will be non-representational. I show that the causal and reductive nature of mechanistic explanations presents a challenge to advocates of representationalism: to identify the causal relations from which the activity of representing is constituted. In order for representations to do explanatory work in a mechanistic explanation of any cognitive phenomenon *qua* representations and not just as purely formal tokens, or in any other capacity, their activity of representing must enter the causal chain leading from mechanism set-up conditions to mechanism termination conditions. Furthermore, the activity of representing must be reducible to the interaction of some lower-level entities, so that the principle of hierarchical nesting of mechanisms is respected. I analyse 4 options available to the representationalists: (1) representing is constituted by some specific pattern of interaction between the representations and the environment outside of the cognitive system; (2) representing is constituted by some specific pattern of interaction between the representations themselves; (3) representing is constituted by some specific pattern of interactions by the entities making up the representational vehicle; and (4) representing is constituted by some specific pattern of interaction between the representations and the environment outside of the cognitive system. I argue that attempts to account for representing using causal interactions with the environment fail, because they can provide at most a distal causal mechanism for representing, but not a constitutive one which is needed. Accounts appealing to the structure of the representational system are shown deficient because they ground representing in relations of similarity, not in causal interactions. Accounts appealing to the internal workings of the representational vehicle either ground representation in similarity, or imply that representational content is fully determined by the properties of the representational vehicle, without any contribution by the object represented. Finally, in accounts which appeal to the interactions between the representation and other parts of the cognitive system, the activity of representing is explanatorily superfluous, since it is constituted by other activities performed by the representation, and does not reduce into a lower level. I analyse existing accounts of representational mechanisms by Grush (1997, 2004), Bechtel (2008), Miłkowski (2013) and Gładziejewski (2015) and show that they exhibit the abovementioned flaws. I therefore conclude that none of the available options is satisfactory, and representing cannot be an explanatorily valuable part of mechanistic accounts of cognition.

Bibliography

BECHTEL, W. (2008). *Mental Mechanisms.* New York: Lawrence Erlbaum.

GŁADZIEJEWSKI, P. (2015). Explaining Cognitive Phenomena with Internal Representations: A Mechanistic Perspective. *Studies in Logic, Grammar and Rhetoric* 40(53), 63-90.

GRUSH, R. (1997). The Architecture of Representation. *Philosophical Psychology* 18(1), 5-24.

GRUSH, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences* 27(3), 377–396.

MIŁKOWSKI, M. (2013). *Explaining the Computational Mind*. Cambridge, MA: MIT Press.

Are the Mental States Underlying Implicit Biases Unconscious? – A Neo-Freudian Answer

ABSTRACT Many philosophers and psychologists hold that implicit biases are due to *unconscious* mental states. The underlying argument seems to be an inference to the best explanation of the mismatch between a subject's explicit report (the "explicit attitude") and her scores in the respective implicit test, such as the Implicit Association Test (the "implicit attitude"). This inference is rejected by defenders of the so-called *Associative-Propositional Evaluation* (APE) model. According to this model, implicit biases are due to mental associations between concepts that induce affective reactions that can be introspected and that are, thus, conscious. The mismatch between the implicit and the explicit attitude is explained by the fact that the subject disqualifies the affective reaction as a basis for her explicit judgements. I will show that while the APE model is promising, the explanation of the mismatch between implicit and explicit attitudes is implausible. The main problem is that the APE-model neglects the fact that the metal states underlying implicit biases usually *are* unconscious in some sense. I will introduce what I will call the *Repression Model* of implicit bias, and I will show how this model can maintain the merits of the APE model while avoiding its problems.

1. Introduction

The focus of this paper will be cases of implicit bias that can be characterized as follows:

(*implicit bias*) An implicit bias is a psychological disposition of a subject S that:
- i) is induced by a mental state M of S (its categorical basis),
- ii) manifests in S's behavior B that "involves a deviation from norms of fairness" (Frankish, 2016),
- iii) B looks like M is a belief that p,
- iv) S reports that she does not believe that p.

Examples of implicit bias in this sense are cases in which a person reports non-racist attitudes but still behaves in a racist way (Haslanger, 2015). Many philosophers as well as psychologists hold that the mental states underlying implicit bias ("M" in condition i)) are unconscious (Banaji, Bhaskar, & Brownstein, 2015, p. 183; Gawronski, Hofmann, & Wilbur, 2006, p. 487; Mandelbaum, 2016, p. 629). The implicit argument for this claim seems to be an inference to the best explanation of condition (iii) in light of condition (iv).

2. The Mental States Underlying Implicit Biases Are Not Unconscious – The APE-Model

Despite its wide acceptance and despite the fact that it provides a straightforward explanation of the mismatch between conditions (iii) and (iv), the claim that the states underlying implicit biases are unconscious has recently come under attack. Most importantly, (Hahn, Judd, Hirsh, & Blair, 2014) show that subjects can correctly predict their *implicit association test* (IAT) scores if asked to focus on their gut feelings. This suggests, according to Hahn et al. and other authors, that the states underlying implicit biases are introspectable, and thus, not unconscious. (Hahn et al., 2014) explain their results on the basis of the so-called 'APE-model' (*Associative–Propositional Evaluation*) (Gawronski & Bodenhausen, 2006). According to this model, implicit biases are mental associations between concepts such as AFRICAN AMERICAN and HOSTILE that induce spontaneous affective reactions which influence a person's behavior. The affective reactions can, due to their phenomenal character, be introspectively accessed.

If implicit biases are not unconscious, the question arises of how the mismatch between conditions (iii) and (iv) is to be explained. Gawronski uses the following example to illustrate how the APE-model explains the mismatch (Gawronski, 2012, p. 662): Imagine a person who has negative affective reactions towards African Americans. Based on these negative affective reactions she will come up with a corresponding belief (statement (1)). At the same time, this person has several further beliefs, which results in the following belief set:

(1) I dislike African Americans.
(2) African Americans are a disadvantaged group.
(3) Negative evaluations of disadvantaged groups are wrong.

This belief set is *cognitively inconsistent* (Gawronski, 2012). An implicit bias arises when the subject resolves the inconsistency by rejecting (1) while the mental association and the corresponding negative affective reaction that led to first holding (1) persists. This explains the mismatch between conditions (iii) and (iv). It is important to note that the reestablishment of cognitive consistency is supposed to be an entirely conscious process (Gawronski & Bodenhausen, 2014, p. 194).

The APE-model is promising because it accounts for the fact that subjects are introspectively aware of their biases, while being able to explain the mismatch between conditions (iii) and (iv). Alas, the explanation is problematic. First, it is assumed that the reestablishment of cognitive consistency that leads to implicit biases is a conscious process. This implies that everyone who has an implicit bias should be aware of this fact. This contradicts the findings that not all subjects in Hahn et al.'s study could become aware of their biases, and that even successful subjects were surprised and felt guilty when realizing that they are biased. This shows that implicit biases *are* unconscious in some sense. A second problem is that it is implausible to think that people can decide to like someone if they *know* that they do *not* like that person.

3. The Repression Model of Implicit Bias

The problems of the explanation of the mismatch based on the APE-model can be avoided if one thinks of implicit biases in terms of repression. According to the Repression Model of implicit bias, the mental states underlying implicit biases are repressed in the sense that they are, first, not attended to as the result of an impulsive act that is induced by an inner conflict that results from the awareness of having a certain feeling towards P in combination with the desire not to have P, where the desire is a consequence of the self-image and/or the internalized social norms. Second, the fact that attention is always shifted away from the feeling has the consequence that the feeling is prone to be miscategorized (e.g., a feeling of fear is miscategorized as a bad stomach). Third, the repression behavior becomes automatized such that the trigger of the repressed feeling will automatically induce the shift of attention and the miscategorization without inducing the conflict. At this stage, repression is complete in the sense that the feeling is unconscious (= disposed not to be in the focus of attention; disposed to be miscategorized) and the inner conflict is not induced anymore. Still, the feeling will influence the subject's behavior in a way that is not controllable for the subject because she is not aware of having the feeling.

The Repression Model accounts for the advantages of the APE-model: it assumes that implicit biases are due to feelings that can in principle be introspected; it is compatible with the idea that these feelings are due to mental associations; and these feelings plausibly cause behavior in the same way the affective reactions in the APE-model do. Still, the Repression Model avoids the problems afflicting the APE-model. First, we can explain why only some subjects were able to introspect their biases: depending on the strength of the automatization of the repression, one might not be able to detect the bias-feeling and/or categorize it correctly. Similarly, we can explain why subjects were surprised and felt guilty when realizing that they are biased: so far, they have been repressing the bias-feeling—the experimental setup made them pay attention to the feeling and it provided the subject with the right concepts to label the bias-feeling. This new awareness, then, triggered the conflict-feeling again. Second, we are not committed to the implausible assumption that people can simply decide to like someone while knowing that they do not like that person. According to the Repression Model, the subject does not have to make the evaluation "I don't like African-Americans" in order to, then, reject it. Implicit biases do not result from cognitive inconsistency that has to be resolved by rejecting a belief. Rather, implicit biases result from the subject's attempt to avoid the violation of a desire. Finally, based on the Repression Model we can make several interesting empirical predictions. For example, if the model suggested here is adequate, subjects with a repressive personality style should be less likely to accurately predict their IAT-scores, and people with implicit biases should have higher skin conductance responses when confronted with bias-related stimuli (Kehyayan, Best, Schmeing, Axmacher, & Kessler, 2013).

Making Sense of the Emergent Property: Association or Inference?

Hsiulin Ku

The issue of the emergent property, or the *ad hoc* concept, has been quite come to the front in truth conditional pragmatics. It seems that the contextual influence not only threats the status of minimal content, but also forces us to reconsider how it is possible to have the novel application of a word in sentences, except some local pragmatic modulation of the word's contribution to sentence, such as loose use or enrichment. The ability to make sense of an utterance, or the ability to grasp the primary content (explicature), with novel application or not, has been given two kinds of model, the associative model proposed by Francois Recanati and the relevance-inferential ones by Relevance Theory (RT). In this paper, I advocate for the former stance and suggest a new way to accommodate the emergent property. The argumentation goes in following points:

(i) There is need to indicate that the notion of inference and relevancy in RT's model is not that kind of clear to deal with primary content or explicature. There are cases to show that the routes of inference can be multiple and equally relevant to the context, but the one that jumps into or stays in interpreter's mind is the one that most fit with the occurrent interpretation schema. There are also cases showing that even being corrected, one still follows the wrong interpreting method unconsciously—simply cannot transform the interpretational policy easily. Moreover, there are cases where the interpreter cannot reflect the way of interpreting or modulating, let alone inference, but the communication goes well. There are even cases that, facing the exactly the same utterance, under the same context and conditions, one may not have the same interpretation. I suggest that schema can have better explanatory efficacy to these cases, and both notions of inference and relevance can be accommodated by the notion of schema.

(ii) RT maintains that the notions of meta-representation and schema-shift in Recanati's associative model give the way to the relevance-inferential model, since any move to meta-level or shift depends on the interpreter's finding something and reasoning to the new way of interpretation. To compose and modulate the words' meanings in an utterance, one needs other information to decide the target and direction. And this is where inference comes in.

However, I indicate that any seemingly intermediate step to meta-level or shift to other schema is not exactly bridged by inference, but guided by the semantic sentiment of the schema--Frege calls such sentiment "Färbung", Dummett calls it "tone", and Picardi calls it "aura". Furthermore, I show that this sentimental function of the schema makes a better account for the emergent property.

Key words: emergent property, association, inference, Relevance Theory, Recanati

A Unified, Kaplanian Metaphysics for Linguistic Types

In the sentence "Narcissus loves Narcissus" there are two word types: the proper name type "Narcissus" (which occurs two times) and the predicate type "loves". In the fragment of discourse "Narcissus loves Narcissus. Let me repeat it. Narcissus loves Narcissus" there are two sentence types: the sentence type "Narcissus loves Narcissus" (which occurs two times) and the sentence type "Let me repeat it".

In the debate on the metaphysics of words, and in particular in a tradition stemming from Kaplan's seminal paper "Words" (1990), word types – and in particular the types of proper names such as "Narcissus" – are often deemed to be radically different from other kinds of linguistic types (such as sentence types). They would belong to a different metaphysical category. The purpose of my talk is twofold. First, I show that it is unacceptable to think that word types or name types are so special among linguistic types. Second, I suggest an amendment to the Kaplanian metaphysics of word types that aims to make it extendable to other linguistic types (including sentence types).

Indeed, according to the Kaplanian tradition, word types are temporally persistent, concrete entities that are created at a certain time, and then intentionally repeated or stored in the memory or in an external support. When no repetition occurs and there is no storage of the word, the word ceases to exist. Word types are individuated by their origin (as Sainsbury & Tye 2013, § 4.1, and Sainsbury 2015 argue for proper names), and not by their form or function.

By contrast, sentence types are abstract entities, which exist also if and when nobody utters them; they exist in force of the syntactic rules that make them admissible. As Kaplan puts it:

> My creationism about words does not extend to sentences. The world in which sentences and other compounds live is brimming with untokened types. Put roughly, the basic elements of the language are earthly creations, but the compounds generated by syntactical rules [...] are structures – types – which may or may not have tokens. (Kaplan 2011, p. 511)

There are two *prima facie* sensible – yet defeasible – reasons why Kaplan and his followers opt for a differential treatment of word types and sentence types.

a) Some of the arguments in favour of the Kaplanian view of word types cannot be directly applied to sentence types. Kaplan (1990) suggests that his approach draws the distinction between informative and non-informative identity statements in an exact way; allows for the distinction between specific (or *common currency*) and generic names; helps to make sense of the so-called Paderewski puzzle about belief ascriptions. All these problems primarily concern words, and more specifically proper names. This is also a reason why Sainsbury (2015) limits his originalist metaphysics to the types of proper names. However, this further limitation to names of the Kaplanian approach could suggest that there is a (quite unexplainable) metaphysical divide not only between words and sentences, but even between proper names and other kinds of words (predicates, adverbs, articles and so on).

b) The so-called *productivity objection* to the Kaplanian view (Hawthorne & Lepore 2011, Bromberger 2011) can perhaps be answered in the case of word types (and in particular when the words at stake are proper names), but turns out fatal if the Kaplanian view (in its original form) is applied to sentence types. The gist of the productivity objection is that, in many cases, linguistic types are not introduced through a sort of baptism, but are *produced* from other linguistic types, in accordance with some rules. Consider sentences: it is simply false that someone at a certain point in time uttered the sentence "the snow is white" for the first time, and that this sentence was later repeated and stored, thereby determining a historical tree to which all the tokens of "the snow is white" belong. More plausibly, many tokens of "the snow is white" are produced independently, by combining the words in it in accordance with the syntactic rules of English. The productivity objection concerns also words (insofar as many words – such as "unhappy" or "toothbrush" – are also produced from other words and morphemes, in accordance with morphological rules), but it is possible to avoid it by restricting the Kaplanian approach to simple, non-compound words. This is another reason why plausibly Sainsbury (2015) restricts his attention to proper names, for which morphological composition is less common and does not follow clear rules.

The heterogeneity of the resulting metaphysics of linguistic types risks growing when other linguistic types are considered. It is not clear how – in the Kaplanian approach – the line between concrete,

persistent, history-identified linguistic entities on one hand and abstract, eternal and structure-identified on the other should be drawn for other linguistic types: what about types of sub-sentential phrases, or types of multi-sentence discourses?

My first aim is to show that the resulting, heterogeneous metaphysics of linguistic types is unacceptable. At the level of tokens, if a word token *is part* of a sentence token, then they belong to the same category of entities (e.g. they are both sequences of sounds, if the sentence is pronounced; or both inscriptions, if the sentence is written). But a parthood relation seems to hold also at the level of types, and is not clear why at this level it should happen that a concrete entity (a word type) is part of an abstract one (a sentence type). Parthood relations are often thought not to trespass the abstract/concrete divide (Simons 2006). Even less plausible is the hypothesis that only the types of proper names are concrete, while the other word types are abstract, so that a sentence type would be an abstract/concrete hybrid whole.

Moreover, the distinction between words and sentences is too porous to be a categorical distinction: many sentences consist of a single word. In the case of the Italian sentence "piove", the Kaplanian would be forced to say – quite implausibly – that the sentence type "piove" is abstract and structurally/functionally individuated, while the word type "piove" is concrete, in spite of the fact that all the tokens of the sentence type "piove" are also tokens of the word type "piove".

My second aim is to suggest that Kaplan's approach can indeed be extended to a unified metaphysics of linguistic types, once we withdraw the claim that linguistic types have a single origin and are individuated by this origin. Instead, linguistic types should be individuated by their whole history, and a full account of their history should countenance not only the chain of repetitions and storages, but also the syntactical and morphological rules according to which words, sentences, sub-sentential phrases etc. are produced. The resulting identity criterion for linguistic types would be the following:

> *Identity criterion for linguistic types.* Two tokens are tokens of the same linguistic type if and only if *either* they belong to a single (potentially multi-origin) history of repetitions and storages *or* if they belong to histories whose initial episodes consist in the application of the same morphological or syntactical rules to the same types.

Suppose that a child utters mechanically and unreflectively "the snow is white" after hearing her mother uttering "the snow is white": in this case, the utterances would belong to the same sentence type because they belong to a same history of repetitions and storages. On the other hand, two utterances of "the snow is white" which *do not* belong to a same history could still be tokens of the same type, inasmuch as they are produced by applying the same syntactic rules to the same word types. A similar analysis could be applied to compound words (such as "unhappy") and sub-sentential phrases.

Admittedly, words are *more often* historically transmitted than independently produced by applying morphological rules. By contrast, sentences are more often produced by applying syntactical rules than simply repeated. However, this statistical difference does not justify a sharp metaphysical distinction. Analogously, the fact that a limited class of words – such as simple proper names – are never the outcome of syntactical or morphological productivity is no reason to think that they are metaphysically special.

References

S. Bromberger, "What Are Words? Comments on Kaplan (1990), on Hawthorne and Lepore, and on the Issue", The Journal of Philosophy 108 (9), 2011, 486-503
J. Hawthorne, E. Lepore, "On Words", The Journal of Philosophy 108 (9), 2011, 447-485
D. Kaplan, "Words", Aristotelian Society Supplementary Volume 64 (1), 1990, 93-119
D. Kaplan, "Words on Words", The Journal of Philosophy 108 (9), 2011, 504-529
R.M. Sainsbury, "The Same Name", Erkenntnis 80 (S2), 2015, 195-214
R.M. Sainsbury, M. Tye, *Seven Puzzles of Thought and How to Solve Them*, OUP, Oxford 2012
P. Simons, "Real Parts, Real Wholes: Mereology Without Algebra", The Journal of Philosophy 103 (12), 2006, 597-613

Multipropositionalism and Necessary A Posteriori Identity Statements

Abstract

In contrast to the prevailing positivist views of the time, Kripke (1971/1980) famously argued that "an identity statement between names, when true at all, is necessarily true, even though one may not know it apriori" (1980, 108). Kripke's claim was revolutionary, and, though the intuitions motivating Kripke's claim are widely accepted, contemporary analytic philosophers continue to be puzzled by it. How can empirical evidence be required to know whether or not something is true, if it is necessarily true, i.e. true in every metaphysically possible world? Empirical evidence is typically required for knowledge only when something is true in some possible worlds, yet false in others; one needs empirical evidence to determine to which group the actual world happens to belong. So why would one need empirical evidence to know something that is true in every possible world? Why would one need empirical evidence that something *is* the case, if it *cannot but be* the case? Here we will be concerned to solve this puzzle. More precisely, here we will be concerned to explain why utterances of the all-too familiar sentence

(1) Hesperus is Phosphorus.

can be correctly categorized as semantically expressing both something that is (weakly) necessary, and something that is aposteriori. The explanation we endorse relies on a version of Perry's (2001/2011) multipropositionalism: A single utterance of a declarative sentence semantically expresses multiple propositions. On our view, what explains Kripke's revolutionary claim with regard to an utterance of (1) is that such an utterance *semantically expresses* several propositions, one of which is necessary (and apriori), and another of which is aposteriori (and contingent).

This prima facie plausible explanation of Kripke's claim, however, seems to be threatened by Soames (2005) *nesting argument* against *two-dimensionalism*. Two-dimensionalism (Soames treats Chalmers (1996) and Jackson (1998) as paradigmatic proponents of two-dimensionalism) also associates multiple propositions with particular utterances of (1), and moreover it also claims that one of these propositions is necessary (and apriori), while another is aposteriori (and contingent). Soames' nesting argument, succinctly put, is that views which propose explaining seemingly incompatible properties of what is semantically expressed by an utterance by associating multiple propositions with the utterance cannot provide an adequate account of utterances of complex sentences in which an epistemic operator is nested inside the scope of metaphysical operator. If Soames' nesting argument refutes two-dimensionalism –and we grant that it does– then, given the similarities between two-dimensionalism and Perry's multipropositionalism, does Soames' argument not also undermine our multipropositionalist explanation and interpretation of Kripke's claim? We argue that it does not. Though both explanations rely on the plausible idea that a single utterance is in some way associated with multiple propositions, two-dimensionalism differs from multipropositionalism in that it

is further committed to the traditional monopropositional framework according to which utterances semantically express (at most) one proposition. Soames' argument succeeds against two-dimensionalism only because two-dimensionalism is committed to monopropositionalism (i.e., the idea that the utterance of a sentence semantically expresses at most one proposition). Our multipropositionalist theory rejects the traditional monopropositional framework, and consequently it is not refuted by Soames' nesting argument.

Our multipropositionalist explanation of Kripke's claim is presented and defended in four sections. In section 1 we present some of the main features of Perry's multipropositionalist theory and illustrate why the theory offers a compelling explanation of Kripke's claim. In section 2 we present two-dimensionalism's similar explanation, and in section 3 we formulate a version of Soames' (2005) nesting argument against two-dimensionalism. Finally, in section 4 we consider Soames' nesting argument as an objection against the multipropositionalist explanation and interpretation proffered in section 1, and we demonstrate that, despite the similarities between two-dimensionalism and multipropositionalism, Soames' argument does not apply against multipropositionalism.

Keywords: identity statements; multipropositionalism; necessary aposteriori.

Against Predicativism of Names (Abstract)

According to predicativism of names, even names that occur in argument positions have the same type of semantic contents as predicates. (Sloat 1969; Burge 1973; Bach 1987, 2002; Elbourne 2005; Matushansky 2008; Sawyer 2010; Fara 2011, 2015a 2015b). I first present three objections against the description theory which holds that a bare singular name "*N*" has the same semantic content as the definite description "the object (appropriately) named '*N*'" as the view shared by most predicatvists. These are the modal, the epistemic, and the translation objections.

Let us first briefly consider the Kripkean modal objection (Kripke 1980). Compare the following two modal sentences:

(1) Trump might not have been Trump.

(2) Trump might not have been the man named "Trump".

There seems no true reading of (1), whereas there is an intuitive true reading of (2). Names are rigid designators, while definite descriptions are not.

The second objection is the Kripkean epistemic argument (Kripke 1980). It is based on the fact that the following two sentences have the different epistemic statuses:

(3) If Trump exists, Trump is the man named "Trump".

(4) If the man named "Trump" exists, the man named "Trump" is the man named "Trump".

(4) is clearly knowable *a priori*, while (3) is only knowable *a posteriori*. In order to know that Trump has the name "Trump", I have to see what name linguistic community members commonly use to refer to Trump, or they have to inform me of what Trump's name is. Before having the relevant experiences about it, there is no way to know what name Trump has.

Let us finally consider the translation objection. Compare the following two sentences:

(5) Trump is the present U.S president.

(6) The man named "Trump" is the present U.S. president.

This objection hinges on the intuition that (5) is simply about the person Trump, while (6) is about the particular name "Trump". It becomes clearer if we translate them into another language:

(7) 트럼프는 현재 미국의 대통령이다.

(8) "Trump"라 이름 붙여진 사람은 현재 미국의 대통령이다.

They seem like correct translations, which preserve the literal meanings of (5) and (6). However, it seems possible that a certain Korean speaker who knows the proposition expressed by (7), i.e., that Trump is the present U.S. president, does not know the proposition expressed by (8) if she does not know any letter of English.

All of these three objections strongly suggest that the bare singular name "Trump" cannot have the same semantic content as the definite description "the man named 'Trump'".

However, Fara (2015a) attempts to resolve the modal objection with her "the"-predicativism, the view that whenever a name "*N*" occurs as an unmodified bare singular in an argument position, the unpronounced definite article "the" always co-occurs in front of "*N*" and they constitute the incomplete definite description "$\emptyset_{\text{the}}$ *N*" (Sloat 1969; Matushansky 2008; Fara 2015a, 2015b). According to these "the"-predicativists, there is actually the definite article "$\emptyset_{\text{the}}$" ahead of the bare

singular name "Trump", and the incomplete definite description "the Trump" works the same way as other incomplete definite descriptions, say, "the table".

Here Fara argues that "the"-predicativism immune to the modal objection because incomplete definite descriptions are rigid. Consider the following sentence:

(9) Olga enjoyed the party.

Fara claims "the party" in (9) is a rigid designator. She says, "If you were to sincerely utter [(9)], you would be talking about some one particular party – call it BASH – and saying of it that Olga enjoyed it." (Fara 2015a: 97).

Fara's argument for it is based on the observation that incomplete descriptions do not seem to take narrow scope like names (Fara 2015a: 99-100). However, it seems that incomplete descriptions can take narrow scope. Suppose that my friend and I know that Olga enjoys the party when and only when I am with her, and Olga really wanted to go to the party with me last night, but she could not. If I utter "Olga might have enjoyed the party" in this context, my utterance seems to express the proposition that this might have been: Olga enjoyed the party that I went to last night, not the propositions that the party that I went to last night is an x such that this might have been: Olga enjoyed x.

Moreover, the semantic content of an incomplete description does not seem to be the denoted object unlike a name. Consider following two sentences:

(10) The party was a social gathering.

(11) BASH was a social gathering.

If the semantic content of (10) is the same as that of (11), which is a singular proposition, then it is puzzling why (10) seems *a priori* (and analytic), whereas (11) seems not. If I say, "I participated in BASH yesterday", then the hearer comes to know that BASH is some kind of event and can start to competently use the name "BASH". But she can still wonder what kind of event it is: knowing the proposition expressed by (11) is surely *a posteriori*.

Bach's response to the objections is completely different from Fara's. He agrees with the idea that descriptions, whether they are complete or incomplete, are not rigid. He claims rather that a name "*N*" is not rigid, because it is semantically equivalent to the description "the bearer of '*N*'", and this description is not rigid (Bach 1987: 167-174; 2002: 85-88).

Bach provides some cases where we use bare singular names non-rigidly. He says, "[S]uppose that Aristotle's parents decided to name their first two sons "Aristotle" and "Aristocrates" but hadn't decided in which order. Then, when their first son was born, they made up their minds and named him "Aristocrates", saving "Aristotle" for their second son, the future student of Plato. They could have made the reverse decision" (Bach 2002: 88). In this circumstance, according to Bach, we have a true reading of the sentence "Aristotle might not have been Aristotle", which actually means that Aristotle might not have been the bearer of "Aristotle".

However, the intuition of the suggested true reading seems to arise from the use-mention confusion. This becomes clearer if we devise the exactly same kind of example for ordinary predicates: Suppose that the inventor of the laptop hadn't decided which word she used to name it. She was torn between two words "laptop" and "lapdown". Finally, she made up her mind and named it "laptop". In this circumstance, we can pragmatically covey something true by uttering the sentence "Laptops might not have been laptops." But it seems clear to me that this sloppy, colloquial speech does not suggest any view that the semantic content of "laptop" is the property of being in the extension of the word "laptop". The semantic content of "laptop" is surely the property of being a small portable computer.

2

On the Uniformity of Proper Names

In the semantics for proper names, predicativism is the view that proper names (a) have the same semantic type as count nouns (e.g. <e,t>-type in extensional semantics), and (b) express properties that are applicable to multiple objects. Predicativism is mainly aimed at providing a uniform account (that is, one without appealing to lexical ambiguity) of the following two types of uses:

(1) Stella is in the museum. (Referential Uses)

(2) Two Stellas are in the museum. (Count Noun Uses)

Predicativists (e.g. Burge and Fara) generally claims that proper names express name-bearing properties. For example, "Stella" expresses the property of *being called "Stella"* or *being named "Stella"*. Referential uses like (1) are explained by postulating a covert determiner that functions like "the" (e.g. Fara) or "that" (e.g. Burge). Using "Stella" to refer to a particular Stella is just like using "the cat" or "that cat" to refer to a particular cat. In this paper, I argue that this predicativist account gives wrong predictions about the meanings of bare proper names (that is, names without any determiners) and then develop a positive account of referential and count noun uses of names, according to which proper names are semantically referential but there is a null derivation that transforms names to count nouns, working semantically like the affix "-ian".

According to predicativism, a proper name appears as a count noun in every occurrence and thus it always expresses a property which can apply to multiple objects. Predicativism predicts that bare proper names do not have referential readings. But this prediction is wrong. I appeal to two main kinds of data: incorporated names and modified names. Incorporated nouns and modified nouns as the arguments of non-intersective adjectives are known as bare in the sense that they do not carry any determiners.

(3) John is a [bike owner].

(4) *John is a [[the/that/this bike] owner].

(5) There is a [fake princess] in the party.

(6) *There is a [fake [the/that/this princess]] in the party.

(7) John is a Donald Trump supporter.

(8) Since the breakdown of Imperial Russia, there have been many fake Anastasias.

If names are a kind of count noun, it is expected that incorporated names and modified names as the arguments of non-intersective adjectives should be bare and so lack any referential readings. But they have referential readings. That is, "Donald Trump" in (7) can refer to a particular Trump and "Anastasia" in (8) can refer to a particular Anastasia.

Predicativists might respond by claiming that predicativism with a certain version of a domain restriction theory (e.g. Stanley and Szabo's Nominal Restriction Theory) can explain the referential readings of (7) and (8). According to the nominal restriction theory, there is an unpronounced variable which comes with a noun. The linguistic contribution of the variable is to provide a contextually salient property, and the meaning of the whole noun phrase including both an overt noun and the covert variable is the intersection of the set contributed by the noun and the set contributed by the hidden variable. Given this picture, predicativists can say that the referential readings of (7) and (8) are derived when the intersection of the two sets contributed by a noun and the variable is a singleton set.

However, there are two problems with this suggestion. First, predicativism with the nominal restriction theory might be able to explain positive evidence, but it still fails to explain negative evidence that incorporated count nouns do not have referential readings. If predicativists accept the nominal restriction theory, then it seems that a count noun as well as a proper name should carry a domain variable and so incorporated count nouns should be able to have referential readings. But they do not have such readings. That is, "bike" in (3) cannot refer to a particular bike.

The second problem of the combination of predicativism and the nominal restriction theory is that it produces wrong truth conditions regarding the referential readings of modified proper names. If "Anastasia" carries a domain variable, both "Anastasia" and the domain variable will be in the scope of "fake" in (8). So the logical form of "fake Anastasia" is like "fake F G". In order to be a fake F G, one should 1) present oneself, or be presented, as F and G and 2) be not (F & G). For example, a fake Russian princess is a person who presents herself (or is presented) as Russian and a princess but is either not Russian or not a princess. Thus, if predicativism is correct, in order to be a fake Anastasia, one should present herself (or be presented) as being called Anastasia. However, this is not correct because

a fake Anastasia can present herself (or be presented) as an Anastasia who lost any memory of her name for some reason.

As Jeshion (2015) pointed out, count noun uses of proper names are not limited to meta-linguistic examples.

(9) a. Two Stellas are inside the museum. (Producer Example)
b. There are two paintings by the artist Frank Stella inside the museum.
(10) a. Two Obamas came to the Halloween party. (Representation Example)
b. Two people dressed as Obama came to the Halloween party.
(11) a. Two little Lenas just arrived. (Resemblance Example)
b. Two children (of Lena) who resemble Lena just arrived.

The a-sentences in (9) - (11) can be used to express roughly the same propositions as the b sentences. That is, proper names can be used not only as count nouns that express meta-linguistic properties (e.g. being called Stella) but also as count nouns that express individual related properties (e.g. paintings by Stella, people dressed as Obama, children of Lena who resemble Lena).

My suggestion is that there is a null-morpheme, the function of which is similar to that of "-ian". When the affix "-ian" combines with a name, it plays at least two roles. One is to change the semantic type of proper names (e-type) into the predicative semantic type (<e,t>-type). The other is to introduce a relation into the meaning of the resulting expression. For example, depending on context, "Kripkean idea" can have a variety of meanings: ideas claimed by Kripke, ideas which conform to the philosophy of Kripke, ideas originated with Kripke and so on. So, the lexical entries of "-ian" and "Kripke" can be given as follows:

(12) $[\![\text{-ian}]\!] = \lambda.R \in D_{<e<e,t>>}.\ \lambda.x \in D_e.\ \lambda.y \in D_e.\ Ryx.$
(13) $[\![\text{Kripke}]\!] = \text{Kripke}.$

According to (12), "-ian" takes a relation R and an individual x, and then produces the property of being in the relationship R with x. Since "-ian" has to take a relation before combining with a proper name, it should be understood as carrying the relation variable R, the value of which should be assigned by a context. Then, the compositional meaning of "Kripkean" can be given as follows:

(14) $[\![\text{Kripke [R -ian]}]\!] = \lambda.y \in D_e.$ y is in the relationship R with Kripke.

According to (14), "Kripkean" expresses a property of being in some contextually provided relationship R with Kripke. Since the value of the relation variable R is determined by a context, the meaning of "Kripkean" is context-dependent.

According to my suggestion, the null morpheme "Ø" is present in the logical forms of the proper names in (9)-(11).

(15) $[\![\text{-Ø}]\!] = \lambda.R \in D_{<e<e,t>>}.\ \lambda.x \in D_e.\ \lambda.y \in D_e.\ Ryx.$

When the relations of *Being painted by*, *Being dressed as*, and *Resembling* are assigned as the values of the relation variable, the combinations of the proper names in (9) - (11) and the null-morpheme "Ø" can express the properties of *Being painted by Frank Stella*, *Being dressed as Obama*, and *Resembling Lena*.

One might object to my claim that all count noun uses of proper names can be explained by the referential meanings of proper names and the meaning of the null-morpheme "Ø" on the ground that it does not explain the meta-linguistic meaning of a proper name that predicativists have assumed. That is, in (2), the semantic contribution of the name "Stella" is not the object Steall in the world, but the meta-linguistic object "Stella". It seems that in this context, the property expressed by "Stella-Ø" is *Bearing the name "Stella"*, rather than *being in some relationship with a particular person Stella*.

My answer to this objection is that linguistic expressions are sometimes used to refer to themselves, or that quotation marks are sometimes omitted from the names of linguistic expressions (Jeshion 2015c; Leckie 2013). Consider the following examples:

(16) a. My name is Stella.
b. Stella is quite popular these days (with expectant parents). (Jeshion 2015).

Even though proper names are not placed in quotes, the sentences in (16) appear to be very natural in both spoken and written language. If "Stella" in (2) has been used in the same way as the names in (16), then the property of *Bearing the name "Alfred"* can be derived from the semantic contributions of "Stella" and "Ø" when the relation *Bearing* is assigned as the value of the relation variable.

Towards a Unified Semantics for Extensional and Cognitive Verbs

1. Introduction. The principle of compositionality is one of the central tenets of contemporary semantics. This principle states that the content (or 'meaning') of a complex linguistic expression is determined by the contents of its syntactic constituents and their mode of combination (cf. Hodges 2001; Pagin and Westerståhl 2010). The compositionality of linguistic contents explains the productivity and systematicity of linguistic understanding[1] and is supported by the preferability, *ceteris paribus*, of semantic theories with minimal complexity (cf. Szabó 2017; Pagin 2012).

The past decades have witnessed the remarkable success of theories assuming compositionality. However, recent work in semantics and the philosophy of language has identified a number of challenges for compositionality. These challenges prominently include the contexts provided by *de dicto*-readings of cognitive attitude verbs like believe (cf. Cresswell 1985; Pelletier 1994). Such contexts have stricter requirements on the substitutability of their clausal complements than the contexts provided by extensional verbs like indicate. In particular, truth-conditionally equivalent expressions (e.g. (1a), (1b)) which can be substituted *salva veritate* in extensional contexts can often *not* be substituted *salva veritate* in cognitive contexts. The difference in substitutability between extensional and cognitive verbs is exemplified by (2) and (3).

(1) a. Someone added sodium to the solution.
b. Someone added natrium to the solution.

(2) a. The reaction indicates [$_{\text{CP}}$that s.o. added sodium to the solution]. (**T**)
b. The reaction indicates [$_{\text{CP}}$that s.o. added natrium to the solution]. (**T**)

(3) a. Len believes [$_{\text{CP}}$that someone added sodium to the solution]. (**T**)
b. Len believes [$_{\text{CP}}$that someone added natrium to the solution]. (**F**) !

Philosophical semantic theories attempt to answer the above challenge i.a. by explaining the intuitive non-substitutivity of truth-conditional equivalents in cognitive contexts w.r.t. the pragmatic basis of these intuitions (cf. Salmon 1986, Ch. 6) or w.r.t. the presence of hidden indexical arguments of cognitive verbs (cf. Crimmins and Perry 1989; Richard 1990). However, since these accounts undermine speakers' intuitions about substitutivity, require a hard-to-specify notion of pragmatic implicature, or question the structural correspondence between syntax and semantics, none of these answers has been adopted into mainstream formal semantics.

2. Objectives. This talk outlines an alternative answer to the above challenge that follows the lead of two-dimensional semantics (cf. Chalmers 2006; 2011). Like Chalmers' semantics, our answer interprets complement clauses as *unified* contents that include both the truth-conditional content of these clauses (hereafter glossed *semantic content*; cf. (2)) and the kind of content that is required to capture the substitution behavior of cognitive constructions like (3) (hereafter glossed *cognitive content*). Like Chalmer's semantics, our answer explains the different substitution behavior of extensional and cognitive verbs by assuming that these verbs select for different components of the unified content (cf. Chalmers 2006, p. 104). In particular, while extensional verbs like indicate only select for the semantic component of their complement's unified content, cognitive verbs like believe also select for the cognitive component of the unified content.

In contrast to two-dimensional semantics, our answer treats the unified content as an object of *the same* type as the semantic or the cognitive component of this content,[2] and does *not* assume a strict separation between the semantic and the cognitive component of unified content. As a result of the former, our answer does not require the interpretation of different occurrences

[1] See (Werning 2005) for a critical assessment of these arguments for compositionality.

[2] Two-dimensional semantics analyzes unified content (there called *two-dimensional intensions*) as functions from centered worlds to sets of possible worlds (type $\langle s, \langle s, t\rangle\rangle$) and analyzes the semantic and the cognitive component of this content as sets of possible worlds resp. of centered worlds (type $\langle s, t\rangle$).

2

of the same expression as objects of different types and, hence, enables a simpler compositional interpretation of natural language (cf. Liefke and Werning, submitted; Theiler et al. 2016). As a result of the latter, our answer enables the modeling of 'mixed' contexts (i.e. the contexts provided by, e.g., factive verbs like *remember*), which require the combination of particular *aspects* of semantic and cognitive content, without combining *all of* semantic and cognitive content (cf. Marconi 2005, Sect. 4). To reflect the integrability of aspects of the cognitive component of content in semantic content (and *vice versa*), we call the proposed semantics *Integrated Semantics*.

To obtain the integrated notion of unified content described above, Integrated Semantics interprets sentences and clausal complements as functions (called *parametrized unified contents*) from agents' informational situations to partial sets of informational situations. The members of these sets satisfy both a semantic constraint (i.e. the members' support of (the proposition expressed by) the complement) and a cognitive constraint (i.e. the members' encoding of the attitude bearer's information about the complement's subject matter at the bearer's informational situation). Parametrized unified contents serve as input of the compositional machinery. In Integrated Semantics, the different substitution behavior of extensional and cognitive verbs is explained by the fact that extensional and cognitive verbs supply different arguments for the parametrized unified content of their complement and, thus, yield different unified contents of the same complement. In particular, while cognitive verbs like believe supply the *informational* situation of the verb's subject (at the time of cognition), extensional verbs like indicate supply the '*empty*' situation, i.e. the situation which does not support any sentence. Our talk will specify the particular interpretation of extensional and cognitive verbs that brings about these verbs' different substitution behavior, and will give a unified compositional semantics for a small fragment of English containing these verbs.

References.

Chalmers, David J. *The foundations of two-dimensional semantics*, Two-Dimensional Semantics (M. Garcia-Carpintero and J. Macia, eds.), Oxford UP, 2006, pp. 55–140.

_____. 2011. *Propositions and attitude ascriptions*: *A Fregean account*, Noûs **45/4**, 595–639.

Cresswell, Maxwell J. 1985. *Structured Meanings*: *The semantics of propositional attitudes*, MIT.

Crimming, Mark and John Perry. 1989. *The prince and the phone booth*: *reporting puzzling beliefs*, Journal of Philosophy **86**, 685–711.

Hodges, Wilfrid. 2001. *Formal features of compositionality*, Journal of Logic, Language and Information **10**, 7–28.

Liefke, Kristina and Markus Werning. submitted. *Evidence for single-type semantics – an alternative to e/t-based dual-type semantics*.

Marconi, Diego. 2005. *Two-dimensional semantics and the articulation problem*, Synthese **143/3**, 321–349.

Pagin, Peter and Dag Westerståhl. 2010. *Compositionality I*: *definitions and variants*, Philosophy Compass **5/3**, 250–264.

Pagin, Peter. *Communication and the complexity of semantics*, The Oxford Handbook of Compositionality (M. Werning, W. Hinzen, and E. Machery, eds.), Oxford UP, 2012.

Pelletier, Francis Jeffry. 1994. *The principle of semantic compositionality*, Topoi **13**, 211–24.

Richard, Mark. 1990. *Propositional Attitudes*: *An essay of thoughts and how we ascribe them*, Cambridge UP.

Salmon, Nathan U. 1986. *Frege's Puzzle*, Ridgeview Publishing Co.

Szabó, Zoltán Gendler. *Compositionality*, The Stanford Encyclopedia of Philosophy: Summer 2017 Edition (E.N. Zalta, ed.), https://plato.stanford.edu/archives/sum2017/entries/compositionality/.

Theiler, N., F. Roelofsen, and M. Aloni. 2016. *A uniform semantics for declarative and interrogative complements*.

Werning, Markus. 2005. *Right and wrong reasons for compositionality*, The Compositionality of Meaning and Content. Volume I: Foundational issues (M. Werning, E. Machery, and G. Schurz, eds.), Ontos Verlag.

Abstract Concepts, Compositionality

and The Contextualism-Invariantism Debate

Concepts in cognitive science are theoretical constructs that are posited to explain higher cognitive abilities such as object classification, decision-, inference- and analogy-making (Barsalou, 2012; Machery, 2009). According to the majority view, these explanatory posits are thought to be stable mental representations in long-term memory that are retrieved in an automatic and context-insensitive manner (Machery, 2015; Barsalou 2012; Laurence & Margolis 1999; Fodor 1998; Keil 1994). I refer to this view as Invariantism.

According to the competing and increasingly popular view, concepts should be construed as knowledge that is not retrieved by default, but constructed "on the fly", from now on referred to as Contextualism or Variantism. Defenders of this theory (e.g., Casasanto and Lupyan 2015; Kiefer and Pulvermüller 2013; Hoenig et al. 2008; Kiefer 2005; Barsalou 1992, 1987) argue that the available empirical and conceptual evidence is best accounted for by unstable and context-dependent conceptual representations.

The clear advantages of default representations are that they can easily account for the apparent stability of concepts in thought and communication and its computational simplicity. When seeing a tiger, we do not have to retrieve all of our knowledge associated with tigers to construct a concept upon which we base our decision to run away. Instead we can immediately react appropriately because of the fast retrieval of our default concept of tiger that includes the feature 'highly dangerous'.

To explain appropriate behavior in less typical situations, however, default theorists have to posit another kind of knowledge structure, called background knowledge, which is supposed to show how the individual adjusts their default knowledge to an unfamiliar context. For example, in a zoo it would be inappropriate to run away upon hearing that there is a tiger because we have the background knowledge that tigers in zoos are usually behind bars.

I would like to present my forthcoming paper in which I argue that Invariantism cannot account for two important desiderata for any psychological theory of concepts, *compositionality* and *scope*. First, I show that Invariantism cannot explain abstract concepts (concepts without physical referent or whose referents share few features) and how concepts compose and without collapsing into Contextualism. Second, I argue that stability of thought for which Contextualism seems especially problematic, can easily be accounted by it.

Finally, I discuss the empirical evidence for both claims. I argue that Contextualism and Invariantism

are often mischaracterized and that if properly characterized, both views are difficult to be falsified experimentally. In particular, that which is used to defend Invatiantism (typicality effects especially) can support both views and can easily be accounted for by Contextualists.

It will be left for discussion what consequences this result has for a theory of language comprehension and the usefulness of the notion of concept for the psychology of language.

References:

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Leher & E. Feder Kittay (Eds.), *New essays in semantic and lexical organisation* (pp. 21–49). Mahwah, NJ: Erlbaum.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

Barsalou, L. W. (2012). The human conceptual system. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 239–258). Cambridge: Cambridge University Press.

Hoenig, K., Sim, E. J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience, 20*, 1799–1814.

Kiefer, M. (2005). Repetition-priming modulates category-related effects on event-related potentials: Further evidence for Multiple cortical semantic systems. *Journal of Cognitive Neuroscience, 17*, 199–211.

Machery, E. (2009). *Doing without concepts*. New York, NY: Oxford University Press.

Machery, E. (2006). Two dogmas of neo-empiricism. *Philosophy Compass, 1*, 398–412.

Machery, E. (2009). *Doing without concepts*. New York, NY: Oxford University Press. Machery, E. (2010a). Précis of doing without concepts. *Behavioral and Brain Sciences, 33*, 195–206.

Machery, E. (2010b). A reply to Barbara Malt and Jesse Prinz. *Mind & Language, 25*, 634–646.

Machery, E. (2015). By default: Concepts are accessed in a context-independent manner. In L. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 567–588). Cambridge, MA: MIT Press.

Mazzone, M., & Lalumera, E. (2010). Concepts: Stored or created? *Minds and Machines, 20*, 47–68.

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences, 17*, 458–470.

What about the colours? Shoemaker on appearance properties

The objects we perceive can appear differently to different perceives and in different circumstances of perception. According to a number of philosophers, accounting for this everyday phenomenon motivates an appeal to special properties of objects, construed sometimes as perceiver- or experience-dependent properties (cf. e.g. Johnston 1992, McLaughlin 2003, Levine 2006) and sometimes as objective, but context-dependent properties (cf. e.g. Noë 2004, Schellenberg 2008, Genone 2016). I here focus on one of the most sophisticated versions of this approach: Shoemaker's 'appearance properties' view (2000, 2003, 2006). On this view, when an object appears different to a perceiver in spite of retaining its objective, intrinsic properties, this is due to its appearing to have different appearance properties, which are individuated partly in terms of the perceiver's experience. The metaphysical commitment to properties other the ones we usually suppose the objects we perceive to have is supported by the Ways=Properties Principle (Shoemaker 2006): an object's appearing a certain way is always a matter of its appearing to have a certain property. Since sometimes the way an object appears is not a matter of its appearing to have an objective, intrinsic property, such as a colour, shape, or size, Shoemaker argues, we need to suppose that objects also instantiate certain relational properties, which explain how they appear. Given the phenomenon of changing appearances, and in particular the actual and possible cases of intersubjective variation in how things appear, paradigmatically in colour experience, these relational properties are best construed as individuated partly in terms of the phenomenology of the perceivers' experiences.

On the face of it, Shoemaker's appearance properties view seems to have the resources to account for the phenomenon of changing appearances, and for the related variation in the phenomenology of our perceptual experiences. But is it really a satisfactory theory of appearances? In this paper, I discuss a potential problem: the appearance properties view seems unable to account for the intuitively plausible claim that, in perceptual experience, we perceive the objective properties of objects. The problem extends beyond colour experience – Shoemaker's preferred case – to shape and size perception, where, within his framework, we seem unable to perceive experience- and context-independent properties. On Shoemaker's view, we always perceptually experience the experience-dependent appearance properties of objects and these properties constitute the ways things appear and explain the phenomenology of perceptual experience. As a result, despite Shoemaker's claims to the contrary, objective properties such as colours lose their explanatory role: there seems to be no grounds for claiming that we perceptually experience these properties. The problem is a serious one: not only does the appearance properties view fail to respect our pre-theoretical assumption that properties like colours are paradigmatically perceivable, but, as I argue, the view also struggles with explaining an important case of changing appearances, i.e. colour constancy (cf. Kalderon 2008, Allen 2016).

I then discuss two strategies Shoemaker could adopt in response to the above problem, and argue that they are unsuccessful. The first strategy appeals to Shoemaker's (2002) account of colour attribution. He distinguishes between a doxastic and a phenomenal sense of "looking". Saying that an object looks e.g. blue in the doxastic sense means that, taking our visual experience at face value, we would judge that that object

is blue and thus attribute the colour to it. An object looking blue in this sense is compatible with it looking a variety of ways, and thus compatible with it instantiating a variety of appearance properties. By contrast, saying that an object looks blue in the phenomenal sense does not amount to attributing a colour to it, but rather an appearance property. However, Shoemaker holds, the object's phenomenal way of looking can ground our colour attribution, because there are some privileged appearance properties that an object with a certain colour instantiates when perceived in standard viewing conditions. When an object phenomenally looks blue, then, it instantiates one the "canonical" appearance properties associated with the colour blue. On the basis of this account, Shoemaker could argue that perceiving one of the canonical appearance properties for a colour allows us to perceive the colour. In response, I argue that the account only explains how we can attribute colours in judgement and come to know about the colours of objects on the basis of our perceptual experience. Because colours only determine the doxastic ways of looking, they do not contribute to how objects appear phenomenally and thus, I argue, to perceptual phenomenology. At most, Shoemaker can defend the claim that perceptually experiencing one of the canonical appearance properties for a colour puts us in a position such that we "count" as perceiving that colour. This, however, means conceding that objective properties can at most be perceived in a different sense than appearance properties – the former can at most be perceived indirectly in virtue of perceiving the ways they appear.

The second strategy Shoemaker could adopt is to revise his view, and argue that what explains how things appear are not appearance properties, but the "qualitative characters" of objective properties such as colours (2006). Since these qualitative characters are aspects of the colours themselves and not, as the Ways=Properties Principle prescribed, further properties of objects, Shoemaker can claim that perceiving a qualitative character is perceiving a colour, and so that, in accordance with our pre-theoretical assumptions, in colour experience we directly perceive the colours. At a closer look, however, Shoemaker is not really offering a new view, but merely giving a more plausible characterisation of his old view. It is not clear, I argue, that the new view rejects the Ways=Properties Principle and that the qualitative characters are not properties distinct from the colours, just like appearance properties were. If so, then formulating his view in terms of the new notion does not help Shoemaker with solving the problem I raised: when it comes to our perceptual experience, the colours themselves are still out of the picture.

What can we learn from this discussion? I conclude by briefly considering two questions that go beyond my assessment of Shoemaker's view. First, whether the prima facie counterintuitive consequence that we do not perceptually experience experience-independent properties such as colours and shapes should be regarded as a sufficient reason for rejecting a theory of appearances. Second, whether this consequence is the symptom of a more fundamental problem for the appearance properties approach, namely, that this approach is unable to fully explain how perceptual experience can gives us direct awareness of the mind-independent world.

The Normative Role of Logic- Abstract

Logic gives us a very clear picture of what inferences are valid, if sets of sentences are consistent, and what form of sentences are tautological or contradictory. Accordingly, logic can tell you what sentences necessary follows from your beliefs, if your beliefs are consistent, and which of your beliefs are necessarily true or false. But does this information come with any normative force? To ask this question is to ask if logic places epistemic norms. I think the answer to this is yes, primarily because belief aims at truth, and to recognise a sentence is true but to resist believing it would be a mistake; it would be *irrational.* Accordingly, to recognise a sentence is logically true, or entailed by your beliefs, means you should believe those sentences in *some* sense. The real difficulty arises in nailing down exactly what that sense is. I defend a novel version, on which logic does place positive epistemic norms, but an agent's rationality is only constrained by the norms of the logic they accept.

Consider a very simple characterisation of the epistemic norms placed by logic:

N1: Where A ⊨ B, if you believe A you ought to believe B

This sort of norm is clearly unsatisfactory, not least for reasons pointed out by Harman (1986). For one, it is too demanding. This means you always *ought* to believe the logical consequences of your current beliefs. But it's not even possible to recognise all the consequences, so we can't consider someone irrational for failing to believe them all. Secondly, this would lead to being obligated to clutter our minds with irrelevancies. From our belief A, it follows that A v A, and A v A v A… and with the norm *N1* we would be obligated to believe them all.

For these sorts of reasons MacFarlane (2004) suggests a different characterisation:

N2: Where A ⊨ B, you ought to see to it that if you believe A then you don't disbelieve B

This wouldn't make the logical norms demanding or cluttering; you aren't irrational for failing to believe logical consequences of your beliefs, but only if you disbelieve them. However, it is far too weak. According to *N2,* an agent who accepts A but refuses to make a judgement on A v A, an obvious logical consequence, couldn't be considered irrational. MacFarlane suggests that whilst the norm cannot oblige the agent to accept the disjunction, we can say that logic gives her a (defeasible) reason to accept it.

I submit that this is far from satisfying; the agent is clearly making a mistake when she refuses to accept A v A, she is being *irrational*. And it is not because she fails to believe something for which there is a *pro-tanto* reason to believe.

Consider for example someone who fully believes that 'grass is green', and 'the sky is blue', and acknowledges that those together entail 'grass is green and the sky is blue', but refuses to take any attitude to the conjunction. This person is clearly making a mistake, they are being irrational. But this irrationality must go further than simply failing to do something for which there is a reason to do. As this reason is *pro-tanto,* an agent could very rationally fail to believe or disbelieve 'the grass is green and the sky is blue' if there were some reason to do so, which outweighed the reason to believe provided by the logic. Someone who recognises the validity of conjunction-introduction to two of their beliefs, and refuses to take an attitude to the conjunction is making a more fundamental error. They recognise that the *truth* of some proposition is guaranteed by some propositions they believe to be *true*, but refuse to take an attitude towards it. It is like claiming 'P is true but I don't believe it or disbelieve it'.

We need a characterisation of the norms that is strong enough to capture the mistake here, but avoids being demanding or cluttering.

Hartry Field (2009) gets something better:

N3: Where it is obvious A ⊨ B, you ought to see to it that if you believe A then you believe B, where A and B are in question

This avoids being demanding or cluttering. An agent isn't irrational for failing to believe all the consequences of their beliefs, as they are only obliged where those entailments are obvious and in question; as long as they're *disposed* to believe them when they do come into question, they are acting rationally. This also captures the irrationality that *N2* could not. However, I'm suspicious of the use of obviousness- obvious to who? An agent could be deemed irrational for failing to map their beliefs to disjunction-introduction, even if they're not aware it's a valid entailment, if that is considered an obvious entailment. In this respect, the norms could end up too strong. They could also end up too weak: I know that Fermat's last theorem is true, but it certainly less than obvious based on it's very complicated proof. Given the complicated but successful proof, it *should* be irrational to accept the axioms of number theory and deny Fermat's last theorem, but *N3* might not capture that.

I suggest the following characterisation:

N4: Where you accept that A ⊨ B, you ought to see to it that you're disposed to believe B if you believe A

This avoids demandingness and clutter in the same way as *N3.* However, I think the notion of being normed by a consequence relation you accept, rather than that is obvious, is better. It wouldn't be irrational to fail to map beliefs over an entailment you're not aware, and *would* be irrational to fail to map belief over one that you accept but isn't obvious.

MacFarlane objects to principles like *N4* because of a priority issue; it seems backwards to think logic only norms our belief when we have logical knowledge. This means we are free to believe what we like if we aren't aware of entailment- but don't we seek logical knowledge to discover how we *should* conduct our beliefs, even in our state of ignorance?

Whilst someone could make this complaint about *N4*, I suggest that it's right we don't consider someone irrational for failing to reason in accordance with inferences they don't accept or aren't aware of. If someone is an open advocate of a paraconsistent logic, with strong reasons for that advocacy, fails to use modus ponens, they shouldn't be deemed irrational. But there is still a way to capture the sort of priority MacFarlane has in mind. I suggest that there could be competing epistemic norms placed by logic. *If* there is such a thing as the correct logic, then that logic can place norms also. However, if suggest that only the norms of the logic one *accepts* places norms that constrain rationality.

Imagine two consequence relations, $\models_1$ and $\models_2$, with the following properties:

A $\models_1$ B, A $\not\models_1$ C

A $\not\models_2$ B, A $\models_2$ C

Suppose agent S accepts $\models_1$, but $\models_2$ is the correct consequence relation. S believes A, should she also be disposed to B or C? I think, in line with *N4,* we shouldn't think S irrational for accepting also B, but not C, as those are the entailments she recognises. But there is a *sense* in which she should accept C, but not B, as those as the entailments of the correct logic. There is a simple reason why; her beliefs would better track the truth she they did. The only norms relevant to her rationality are those placed by her accepted consequence relation, but a correct logic can still place competing norms of the following form:

N5: Where A ⊨ B, you ought to see to it that where you believe A, you are disposed to believe B

There is a sense that S ought to be disposed to believe B, as captured by *N5*, because her beliefs would better track the truth. This sense can capture what it is MacFarlane is pushing for in his priority concern. But *N4* captures the epistemic norms places by logic that constrain an agent's rationality.

A commitment to haecceitism in Robert Stalnaker's solution to indexical beliefs

The aim of my paper is to discuss a commitment to haecceitism implied in Robert Stalnaker's account of indexical thoughts. With his solution to the problem of indexical thoughts Stalnaker aims to preserve the two main tenets of the traditional theory of beliefs: (i) beliefs are two-argument relations between the thinker and a proposition and (ii) propositions have absolute truth values. To this end, he adopts a constraint that every case of a lack of a *de se* belief (or other kind of an indexical belief) can be represented as an ignorance with respect to which world is the actual one. The constraint requires two controversial assumptions: (i) that identification of oneself by reference to the thought one is entertaining guarantees rigid designation across all epistemically accessible worlds (ii) haecceitism. The role of the first assumption is to provide a way of identifying oneself, which is immune to the problem of misidentification and thus does not bring in the problem of essential indexicality. As Claes Weber (2016) argues, it works only as long as we exclude the possibility of having singular thoughts about someone else's thoughts (and by the same token having a third-person access to singular thoughts about one's own thoughts). I will consider whether we have grounds to exclude such a possibility reflecting on what it might mean to have a singular thought about a thought.

A New Metapher Illustrating the Vector Based Information Processing in Systemtheoretical Cognitive Neuroarchitectures in Connectionistic Cognitive Science

The Mind/Brain may be considered as one and the same nonlinear, complex dynamical system, in which information processing can be described with vector and tensor transformations and with attractors in multidimensional state spaces. Thus, an internal neurocognitive representation concept consists of a dynamical process which filters out statistical prototypes from the sensorial information in terms of coherent and adaptive n-dimensional vector fields. These prototypes serve as a basis for dynamic, probabilistic predictions or probabilistic hypotheses on prospective new data (see the recently introduced approach of "predictive coding" in neurophilosophy). Furthermore, the phenomenon of sensory and language cognition would thus be based on a multitude of self-regulatory complex dynamics of synchronous self-organization mechanisms, in other words, an emergent "flux equilibrium process" ("steady state") of the total collective and coherent neural activity resulting from the oscillatory actions of neuronal assemblies. In perception it is shown how sensory object informations, like the object color or the object form, can be dynamically related together or can be integrated to a neurally based representation of this perceptual object by means of a synchronization mechanism ("feature binding"). In language processing it is shown how semantic concepts and syntactic roles can be dynamically related together or can be integrated to neurally based systematic and compositional connectionist representations by means of a synchronization mechanism ("variable binding") solving the Fodor-Pylyshyn-Challenge. Since the systemtheoretical connectionism has succeeded in modeling the sensory objects in perception as well as systematic and compositional representations in language processing with this vector- and oscillation-based representation format, a new, convincing theory of neurocognition has been developed, which bridges the neuronal and the cognitive analysis level.

With reference to the models of Werning's "Oscillatory Networks", Freeman's "K0-KV (Katchalsky) Set Attractor Network Models", Kohonen's "Self-Organizing Feature Map (SOFM)", Grossberg's and Carpenter's "Adaptive Resonance Theory (ART)", Abeles' "Synfire Chains and Corticonics", Jaeger's "Echo State Network (ESN)" for example, the nature of this "vectorial form" of neurocognitional information can be illustrated by means of a metapher, referred to as "the mountain lake and mountain creek metapher": the mode of neural information processing in the human brain and therefore in the functioning of the human mind can be best modeled by self-excited, self-amplifying and self-sustained waveforms superimposing each other in fluid multiple-coupled feedback cycles. Thus, the neural information storage and retrieval in the long-term memory, for example, can be understand by means of computational adaptive resonance mechanisms in the dominant waveforms, or "modes" and by warming up and annealing of oscillation modes by streams of informational processes in the context of computational "energy functions", like in Smolensky's "Harmony Theory": The information processing of sensorial data and syntactic-semantic information can compared with a mountain lake, which spontaneously oscillate in such a manner that, in a

wide range, various wave patterns will occure on its surface. 1. Perception and Sensation: Percepting something can be described in such a manner that a sequence of rain showers impinge on the lake's surface, where a rain drop stands for a sensorial element, for instance a photon hitting a ganglion cell in the retina. The way the raindrops impinge on the wave patterns on the lake's surface favor the generation and the continuity of certain wave structures according to the similarity and frequency of these showers. Thus, one or a low number of wave forms will prevail, or "win," which code or represent such sensorial information. A rain shower stands for a state of affair to be stored, for example an object or an event. The more frequent and more similar a specific sort of rain showers hits the wave motions, the higher the degree of probability that a specific complex wave structure emerges as a superposition of the single rain showers. 2. Long-Term Memory: In the course of time a gradually informational annealing of the mountain lake will take place, with the effect that – at least temporarily – the wave forms will be frozen. Thus, a permanent and relatively stable long-term storage of sensorial data and syntactic-semantic information will be guaranteed, which – on demand – can be defreezed, for example, if new information must be embedded in already existing activation patterns. 3. Short-Term Memory, Working Memory and Reasoning: The information-processing style of a mental process, for example in the form of a cognitive inference process, can be illustrated with a net of backwards-coupled mountain creeks, in other words, with positive and negative feedback loops. This (hyper- and ultra-)cyclic mechanisms – in the sense of hydro- and electrodynamics in physics – can create a laminar or turbulent flow of water with the respective water vortices at the rocks lying in the creek. 4. Thinking and Problem Solving: A mental process can be described by building flux equilibrium processes at certain spatial bounded regions in form of this relativ stable, stationary water vortices, despite of the continuous stream of the creek water. This regions which are equivalent with internal representational concepts are consistent with recurrent process mechanismus in self-organizing artificial neural networks building co-oscillate cycles with resulting resonant processes. Furthermore, one can imagine that the inflowing warm creek water streams in the mountain lake with the result that the already existing wave forms might be adapted if required. These streams including a new wave motion puts the already stored wave patterns in an interferencial oscillation with a higher degree of informational energy so that they are partly superimposed or overwritten.

Challenging Evolutionary Psychology with the eternalist view of the world

1) Arguably, Evolutionary Psychology (EP) gets the right picture of human mind. I will start the talk by briefly motivating this claim. Basically, I'm going to describe EP as a research program (Lakatos, 1970), whose "hard-core" – the idea that our mind is comprised by evolved information-processing mechanisms (widely known as the 'massive modularity hypothesis') – became clear in the 90s. Based on some interesting developments in the "protective belt", I'm going to state that this program is indeed empirically progressive.

2) EP may be challenged with the eternalist view of the world. To make the case, I'm going to bring in the Moving Spotlight Theory. This theory combines eternalism – the view that the world is a big concrete four-dimensional block – with the idea that one moment in time, the 'now', instantiates an A-property: presentness, and is therefore metaphysically privileged. All moments are equally real, and presentness is a property that moves from one moment to the next, like a spotlight.

3) Considering that Moving Spotlight theory has an interesting claim about the architecture of the world, one might wonder: how does a massively modular mind handle such a world? What could a massive modularist say in reply to this question? I'm going explore one possible answer: that human mind counts one a module for event processing. To address this properly, I will sketch a computational explanation – an account of *what* the module does and *why* (Marr, 1982) – of that mechanism. To accomplish this enterprise, I will follow a set of guidelines proposed by Anderson (1990). In short, the module works with three goals: *track* presentness; *link* temporal parts that already instantiated this property and, based on the sort of motion that that sequence manifests, *predict* forthcoming temporal parts of that event (ie, temporal parts that will instantiate presentness shortly). Based on empirical literature, one may propose that this module maximizes fitness (i) by assigning intentionality to events that present motion with biological timing; and (ii) by processing events from the framework of a third person.

4) Finally, I'm going to highlight a couple of limitations of this answer. First, it is not clear what could be the evolutionary story of this module. Second, it seems compatible only with Moving Spotlight Theory, which is highly controversial on its own.

Keywords: Massive modularity; Eternalism; Motion.

Rational Action and De Se Skepticism

A long lasting and quite influential tradition of philosophers, which has its roots in the work of Perry (1977; 1979), Lewis (1979) and Castañeda (1966), believes that First-Personal Self-thinking (FPT) – and indexical thought in general – is special or essential in some sense and cannot be reduced to other types of thoughts a thinker can grasp about him- or herself. What this essentiality consists of is not always clear, but the most common way to define the essentiality of first-person thought is by positing a necessary or at least in some sense special relation between FPT and agency. David Owens (2011: 267) for example claims: "It is widely agreed that agents need information in an egocentric form: they must think of places as 'here' and 'there', times as 'now' and 'then' if they are to be able to act on what they know." Simon Prosser comes to the same conclusion and summarizes: "The most common view about the essentiality of indexicals is that without the kind of thought normally expressed using an indexical, it is not possible to act." (Prosser 2015: 212)

Recently this line of thought has been questioned by so called "De Se Skeptics" (cf. Torre 2016), such as Millikan (1990), Devitt (2013), Cappelen/Dever (2013) and Magidor (2015). They maintain that the proposed connection between rational action and first-person thought is not as tight or even necessary as believed by many adherents of the Perry-Lewis-Castañeda-tradition.

Cappelen and Dever (2013) for instance claim: "First, it is not necessary for an indexical element to enter into the rationalization. Second, on our view the agent doesn't even need to be represented in a non-indexical way in an adequate action rationalization. Nor does any part of the agent's body need to be represented (indexically or not)." (Cappelen/Dever 2013: 37) Furthermore, Cappelen/Dever (2013) argue that their Action-Inventory-Model (AIM) is able to explain rational actions without an appeal to indexical thought or even any agent-related contents. Even persons with the same set of non-indexical beliefs and desires are able to act differently in the same circumstances, since they have different actions available to fulfill their desires. They do not need to have any beliefs about their available actions in order to act rationally.

In the proposed talk I will firstly offer a critique of Cappelen's and Dever's AIM and present cases their model is not able to account for. I will conclude with an inference to the best explanation that we should rather adopt a model for rational action that appeals to beliefs (or other forms of representation) about the available actions than to the action inventory itself. Then, I will argue why those beliefs (or other forms of representation) about the available actions should be regarded as first-personal. I will identify three peculiarities of FPT that support this claim.

The first feature is the reflexivity of FPT. When persons think via first-person thoughts about themselves they usually belief that they do so. They think about themselves as themselves in virtue of the type of thought they have grasped. This might be different in cases where persons use a definite description or a demonstrative concept and accidently happen to be the object to which those concepts apply. The second distinctiveness that relates to the first one is that first-person thoughts do not require any sophisticated information in order to be used to reflexively think about oneself. Even an amnesiac

who has lost almost all information about himself is able to think about himself via first-person thoughts. The only thing he really needs to know about himself is that he is the producer of the thought in question. This information seems to be more secure and less laborious to gain than other forms of information he might have about himself. The last peculiarity that I would like to point out is the fact that when we think with certain indexicals we think about the things our thoughts are about via certain relations in which they stand to us. Hence, if a speaker or thinker uses "I" he is not solely referring to himself but he is referring to himself in virtue of and by presupposing a particular relation that he bears to the object he is referring to and that is the relation of identity. This relation is not presupposed by non-indexical thoughts with the same truth-conditions. For example, when I am thinking "I am in pain." I presuppose that I am identical to the thing my thought is about. This seems to be not presupposed by a thought like "The author of this proposal is in pain."

In the final section, I will sketch how these features of FPT – i.e. reflexivity, effortlessness and relational presuppositions – relate to rational action in general. I will conclude that it is not only important that we have beliefs about our available actions but that we additionally have the information that these available actions are also *ours*. Accordingly, it is not enough for rational action that we have an action inventory, we also need to know which of the many action inventories is *ours*. I will conclude that the identified features of FPT – and thus FPT itself – play a crucial role in maintaining this information.

References

Cappelen, Herman/Dever, Josh. 2013. *The inessential Indexical - On the Philosophical Insignificance of Perspective and the First Person*. Oxford University Press.

Castañeda, Hector Neri. 1966. *He: A Study in the logic of self-consciousness*. In: *Ratio*. Vol. 8. 130 – 157.

Devitt, Michael. 2013. *The Myth of the Problematic De Se*. In: *Attitudes De Se – Linguistics, Epistemology, Metaphysics*. CSLI Publications.

Lewis, David. 1979. *Attitudes De Dicto and De Se*. In: *The Philosophical Review*. Vol. 88. No. 4. 513 – 543.

Magidor, Ofrar. 2015. *The Myth of the De Se*. In: *Philosophical Perspectives*. Vol. 29. 249 – 283.

Millikan, Ruth Garrett. 1990. *The Myth of the Essential Indexical*. In: *Noûs*. Vol. 24. No. 5. 723 – 734.

Owens, David. 2011. *Deliberation and the First Person*. In: *Self Knowledge*. Oxford University Press.

Perry, John. 1979. *The Problem of the Essential Indexical*. In: *Noûs*. Vol. 13. No. 1. 3 – 21.

Perry, John. 1977. *Frege on Demonstratives*. In: *The Philosophical Review*. Vol. 86. No. 4. 474 – 497.

Prosser, Simon. 2015. *Why are Indexicals Essential?*. In: *Proceedings of the Aristotelian Society*. Vol. CXV. Part 3. 211 – 233.

Torre, Stephan. 2016. *De Se Thought and Communication – An Introduction*. In: *About Oneself: De Se Attitudes and Communication*. Oxford University Press.

How to account for Subjective Character?

There is widespread agreement between many contemporary philosophers of mind that phenomenally conscious states involve in addition to their qualitative character some kind of subjective character. (among others cf. Block 1995; Frank 2011; Gallagher/Zahavi 2008; Kriegel 2009; Kriegel/Horgan 2007; Levine 2001; Metzinger 2000; Schlicht 2014; Strawson 2011; Williford 2014, 2006; Zahavi/Kriegel 2016; Zahavi 1999) If a person experiences something he or she is not only experiencing the object of that experience but in addition that him- or herself is having that very experience. Accordingly, in the case of seeing something red it is not only something like (cf. Nagel 1974) to see red but it is at the same time something like that this red experience is for me or mine. Ned Block (1995) has named this phenomenon "me-ishness" and many authors have followed him in believing in it. What is not so clear though is how one is supposed to understand the nature of this experience accompanying subjective character or me-ishness.

There are different accounts that have been provided to answer this question. In a first instance, I will offer a taxonomy of these accounts of subjective character and explain what sets them apart from each other. I will introduce four main approaches to subjective character, which are self-representational accounts (cf. Kriegel 2009), self-acquaintance accounts (cf. Williford 2014), pre-reflective accounts (cf. Henrich 1967; Frank 2011; Zahavi 1999) and mode of presentation accounts (cf. Recanati 2013; Musholt 2015; Kriegel/Zahavi 2016). In the second part, I will review some of the central assumptions and questions that form the motivation for many of the accounts of subjective character. Central to this part of my presentation are issues concerning the purportedly special epistemic and semantic status of first-personal self-ascriptions of phenomenally conscious states. Many adherents of the subjective character view rely on the phenomenon of immunity to error through misidentification relative to the first-person pronoun (Shoemaker 1968) (cf. Musholt 2015; Recanati 2013; Zahavi 1999, 2004) and take it as evidence for a form of subjective character. Others maintain that first-personal linguistic self-reference is only possible because there are pre-linguistic, experiential forms of self-awareness. (Bermúdez 1998; Lang 2010; Musholt 2015; Nozick 1981) This part of the presentation will focus on those aspects but is not limited to them. In the third and final section, I will offer some concluding remarks of what kind of account will be best suited to give the most satisfactory answers to the questions introduced in part two and discuss individual problems that each of the main accounts of subjective character possibly generates.

Literature

Bermúdez, José Luis Bermudez. 1998. *The Paradox of Self-Consciousness*. Cambridge MA. MIT Press.

Block, Ned. 1995. *On a confusion about a function of consciousness.* In: *Behavioral and Brain Sciences*. Vol. 18 (2). 227 – 247.

Frank, Manfred. 2011. *Ansichten der Subjektivität*. Berlin. Suhrkamp.

Gallagher, Shaun/Zahavi, Dan. 2008. *The Phenomenological Mind*. New York. Routledge.

Henrich, Dieter. 1967. *Fichtes ursprüngliche Einsicht*. Frankfurt. Klostermann.

Kriegel, Uriah. 2009. *Subjective Consciousness – A Self-Representational Theory*. Oxford. Oxford University Press.

Kriegel, Uriah/ Horgan, Terry. 2007. *What is Consciousness that we may know it so well?* In: *Philosophical Issues*. Vol. 17, N. 1. 123 – 144.

Lang, Stefan. 2010. *Spontaneität des Selbst*. In: *Neue Studien zur Philosophie*. Vol 23. Jürgen Stolzenberg, Konrad Cramer, Reiner Wiehl (ed.). Göttingen. V&R unipress.

Levine, Joseph. 2001. *Purple Haze: The Puzzle of Consciousness*. New York. Oxford University Press.

Metzinger, Thomas. 2000. *The Subjectivity of Subjective Experience: A Representationalist Analysis of the First-Person Perspective*. In: *Networks 3/4*. This is an extended and considerably revised version of a book chapter that first appeared in Thomas Metzinger (ed.). 2000. *Neural Correlates of Consciousness – Empirical and Conceptual Questions*. Cambridge, MA. MIT Press.

Musholt, Kristina. 2015. *Thinking about oneself - From nonconceptual content to the concept of a self*. MIT Press.

Nagel, Thomas. 1974. *What is it like to be a bat?* In: *The Philosophical Review* Vol. 83, N. 4. 435 - 450.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge MA. Harvard University Press.

Recanati, Francois. 2013. *Mental Files*. Oxford. Oxford University Press.

Schlicht, Tobias. 2014. *Explaining Subjective Character: Representation, Reflexivity, or Integration?* In: Open MIND. Thomas Metzinger, Jennifer Windt (ed.). URl = http://open-mind.net/papers/explaining-subjective-character-representation-reflexivity-or-integration

Shoemaker, Sydney. 1968. *Self-reference and self-awareness*. In: *Journal of Philosophy* 65 (19). 555 – 567.

Strawson, Galen. 2011. *The minimal subject*. In: *Oxford Handbook of the Self*. Shaun Gallagher (ed.). Oxford. Oxford University Press.

Textor, Mark. 2006. *Brentano (and Some Neo-Brentanians) on Inner Consciousness*. In: DIALECTICA 60 (4). 411 – 432.

Williford, Kenneth. 2014. *Representationalisms, Subjective Character, and Self-Acquaintance*. In: Open MIND. Thomas Metzinger, Jennifer Windt (Hg.). URl = http://open-mind.net/papers/representationalisms-subjective-character-and selfacquaintance

Williford, Kenneth. 2006. *Zahavi versus Brentano: A rejoinder*. In: *Psyche* 12 (2).

Zahavi, Dan/Kriegel, Uriah. 2016. *For-me-ness: What it is and what it is not*. In: *Philosophy of Mind and Phenomenology: Conceptual and Empirical Approaches*. D.O. Dahlstrom et alii. (ed.) London. Routledge.

Zahavi, Dan (ed.). 2004. *Back to Brentano?* In: *Hidden Resources – Classical Perspectives on Subjectivity*. In: *Journal of Consciousness Studies* Vol. 11, N. 10 – 11. 66 – 88.

Zahavi, Dan. 1999. *Self-awareness and Alterity – A Phenomenological Investigation*. Evanston. Northwestern University Press.

Sub-sentential speech acts: a situated contextualist account

1. Introduction

The most commonly given examples of sub-sentential speech acts are expressions such as "Nice dress", "From Spain", "Where?" used in such circumstances in which speakers uttering them are regarded as "making moves in a language game", e.g. asserting, asking, promising etc. If Mum knows that Johnny is looking for his ball and says „Under the table" she probably means – and will be understood as meaning – „Your ball is under the table" (or „The ball is under the table", "It is under the table" etc.). If "Under the table" were uttered in response to the question "Where is my ball, Mummy?", Mum's utterance would be a straightforward case of ellipsis. However, since in our example it is uttered in a discourse initial position, with no (articulated) linguistic antecedent, regarding it as an ellipsis is much more problematic.

Philosophers who discuss sub-sentential utterances might roughly be divided into those who think that such utterances are genuinely sub-sentential but nevertheless are speech acts, because they have propositional and truth-evaluable contents (most notably Stainton, but also e.g. Elugardo, Carston, Hall) and those who think that they are speech acts but are not really sub-sentential, since they can be treated as a special kinds of ellipsis (e.g. Merchant, Ludlow, Martí). I side with the former "pragmatically-oriented" (Stainton 2006) group and I'm going to suggest a contextualist account of sub-sentential speech acts, which is based on Recanati's moderate relativism and which is, in my view, a promising way of developing Stainton's insights[1].

2. Situation-relativity: Perry and Corazza

In his 1994 paper Perry uses Wittgenstein's builder's language game to argue that nouns used in isolation (such as "slab" or "pillar") in the right circumstances might e.g. express the command (e.g. that the builder's assistant is to pass a slab) even though such words are not part of sentences. He claims that there are two ways of expressing complete thoughts (Perry 1994: 33): either one can utter a complete sentence whose elements correspond to all necessary constituents of a proposition or one might only "complete a thought": i.e. utter a word or a phrase which supplies only the missing constituents necessary to express a thought whose other constituents are part of the situation of utterance. Using Perry's 1980 terminology we may say that those other elements present in the situation are unarticulated constituents and things that the sub-sentential utterance *concerns* rather than *is about*[2].

More recently Corazza (2011) has used Perry's framework to argue that sub-sentential speech acts are best seen as situated unenriched illocutions. He appeals to Perry's distinction between reflexive and incremental truth-conditions and to his idea of situations providing necessary elements and argues that we do not need to appeal to ellipsis or pragmatic enrichment to claim that sub-sentential speech acts might be successful communicative acts. In his view it is enough for the communication to succeed if the hearer grasps a general thought expressed by a sub-sentential utterance (i.e. its reflexive truth-conditions). If John says to Jane "Hidden on top of the shelf", his communicative act will be successful if Jane entertains the thought "There is something hidden on top of the shelf". She doesn't have to grasp a singular thought expressed by John (such as "Cigars are hidden on top of the shelf") (Corazza 2011).

There are two main problems with Corazza's view interpreted as an account of sub-sentential speech acts. Firstly, Corazza is much more interested in successful communicative acts than with speech acts. Communication may be successful even if no words are uttered, whereas in order to talk about speech acts we need an utterance with a more or less determinate content and illocutionary force. By saying that such utterances need not be enriched, Corazza likens them to nudges and kicks under the table, which indeed might be communicatively successful but cannot be regarded as speech acts. In order for a sub-sentential utterance to be e.g. an assertion it must have asserted as well as otherwise communicated content. And Corazza would be the first to acknowledge that "There is something hidden on top of the shelf" is not the asserted content of John's utterance.

The second problem is that Corazza considers in any detail only two sub-sentential utterances: "Hidden on top of the shelf" and "Reserved". In both cases it is relatively easy to supply reflexive truth conditions. However, there are many utterances for which it would be a challenge. For instance, John might have said only "On top of the shelf". If we do not enrich his utterance it will be unclear whether its reflexive truth conditions are "There is something on top of the shelf" or for instance "Put something on top of the shelf".

[1] Stainton (2006) argues that the sub-sentential utterances' content must be pragmatically enriched, but in my opinion he is not specific enough in explaining how the enrichment is to be done.

[2] According to Perry there are also unarticulated constituents that are part of the content. These are the things that the proposition *is about*. See below.

3. Situated contextualism

I think that there is a better way of using Perry's ideas in accounting for sub-sentential speech acts. The way to do it is to adapt Recanati's framework of moderate relativism. Recanati's relativism presupposes two principles: duality and distribution, which say that to assign a truth-value to a proposition we need both a content and a circumstance of evaluation (duality), and that the determinants of truth-value distribute over content and circumstance (distribution). The principle of economy says that the richer the content, the poorer the circumstances can be and vice versa. (Recanati 2007: 33-34) Thus, for instance, the content of "It's raining here and now" uttered on the 27th January 2017 in Paris will be that it is raining on 27.01.2017 in Paris and the circumstance will consist only of the possible world, whereas the content of "It's raining" uttered in the same situation will be that it is raining, and not only the world but also time and place will belong to the circumstances. The explicit content (*lekton*) plus the situation of evaluation add up to the complete content. Recanati argues at length (2007: part 9) – *pace* Perry (1986) – that *lekta* do not contain any unarticulated constituents. According to him, all unarticulated constituents are things that the proposition concerns and hence belong to the circumstances of evaluation. Perry argues that invariant, externally determined unarticulated parameters are constituents of the situation, whereas those that are not invariant and are cognitively discriminated are constituents of the content. Recanati claims that invariability can be so extended as to include also cases in which unarticulated constituents are cognitively determined, so there is no need to posit unarticulated constituents in the *lekton*.

In my view sub-sentential speech acts might be seen as a limiting case of the working of the principle of economy. However, I'd like to argue that in the case of such acts we have to postulate a two-staged principle of distribution: it's not only the case that the determinants of truth-value distribute over content and circumstance but also the content itself is distributed over the locutionary what is said and the situation of the utterance. In other words, I claim – *pace* Recanati and following Perry – that at least in the case of sub-sentential speech acts *lekta* have unarticulated as well as articulated constituents. Postulating only such unarticulated constituents that the proposition concerns will not do. The reason for this is the following. We understand utterances "It's raining" and "It's three o'clock" without even thinking about the place and time of the utterance and about time zones respectively. The situation of the utterance provides all the relevant parameters and we do not need to consider them. Therefore, the proposition does not have to be *about* them and they may be 'left' in the situation. The case of sub-sentential utterances is different, however. It doesn't suffice that the missing elements are in the situation; they have to be cognitively articulated (in thought). If John utters "From Spain" holding a letter, but his utterance is not accompanied in Mary's thoughts by any representation of the letter, then Mary cannot be said to have grasped the relevant proposition. If John himself doesn't have any representation of the object he's talking about, his utterance cannot be regarded as a speech act. That is why unarticulated constituents of the content are needed. Hence, although I'm using Recanati's relativist framework, the account I'm proposing is contextualist.

I suggest that we use disquotational indirect reports test[3] to distinguish those unarticulated constituents that the proposition is about (i.e. those that belong to the *lekton*) from those that it concerns (i.e. elements of the situation). If an unarticulated constituent has to be added in order to make the indirect report correct, it should be regarded as a constituent of the *lekton*. Thus, since we may report someone's utterance "It's raining" just as "He said that it was raining", unarticulated constituents (place, possibly time) are not elements of the *lekton*; they are items the proposition concerns. On the other hand, since in order to report John's utterance "From Spain" the reporter must add the missing object (e.g. "John said that *the letter* was from Spain"), "the letter" should be counted as an unarticulated constituent of the *lekton*.

References

Cappelen, H., Lepore E. 2005. *Insensitive Semantics*. Blackwell Publishing.

Corazza, E. 2011. Unenriched Substentential Illocutions. *Philosophy and Phenomenological Research* 83: 560-580.

Perry, J. 1986. Thought without Representation. *Proceedings of the Aristotelian Society*, supp. vol. 60: 137-151.

Perry, J. 1994. Davidson's Sentences and Wittgenstein's Builders. *Proceedings and Addresses of the American Philosophical Association*, vol. 68(2): 23-37.

Recanati, F. 2007. *Perspectival Thought, A Plea for (Moderate) Relativism*, Oxford University Press.

Stainton, R. 2006. *Words and Thoughts: Sub-sentences, Ellipsis, and the Philosophy of Language*. Oxford University Press.

[3] Cappelen and Lepore (2005) propose inter-contextual disquotational indirect reports test for different reasons: it is one of their context-sensitivity tests.

Handling Concurrency through Consciousness

For any biological trait, the elaboration of a good evolutionary explanation presents remarkable difficulties[1] in relation to the adopted methodology or definition of adaptation[2], and to the frequent scarcity of empirical data. Furthermore, as far as consciousness is concerned, some philosophical theories such as Conscious Inessentialism[3] or the By-product Accident View[4] deny that the property of being conscious could confer a real selective advantage to organisms, making its evolutionary explication even more arduous.
Nonetheless, it may be reasonable to look at consciousness as a useful tool that, from certain points of the phylogeny onwards, succeeded in handling the organism-environment interaction in a novel – and better - way.

In vertebrates, the ever-increasing complexity of terrestrial life had implied a corresponding complexification of the sensory, cognitive, and motor systems. Such a complexification involved "a rich and intricate set of issues in logistics, control, and resource utilization"[5]. In other words, living beings – especially after the colonization of terrestrial environment – had to handle a huge number of stimuli, inputs and ongoing different sub-systems, and to do so evolution provided them with many mechanisms of stimulus-response. Nonetheless, we could not expect that these mechanical reflexes explain all animal and human behaviours, as reflexes are automatisms which work in a deterministic way, and, what is most important, they handle stimuli in order of arrival, with any possible behavioural planning in the short or long run. On the contrary, living beings live in complex environments in which external (and internal) stimuli concurrently overlap, and this means that they have to recognize the salience of certain information at the expense of others, and have to handle more important stimuli before less important ones. A criterion like the stimulus-response's one, which only focuses on the sequential order of arrival of the inputs, and always responds in a rigid fixed way, unrelated to the context, will fail in handling this kind of multi-layered concurrency.

Therefore, it seems that at some points of the phylogeny, a new cognitive unit emerged, namely a conscious mental space able to centralize, and consequently process more efficiently and flexibly the abundant information provided by the environment. What was "needed", teleologically speaking, was a unitary centre of control of both the requests of the environment, and the available resources, which hold in consideration, at the very same time, the targets and the consequences of possible behaviours, according to past experiences and – consequently – future expectations. Even in cases of "minimal consciousness"[6], hence, the cognitive structure at work should be remarkably complex, for it should entail the acquisition of a cognitive model of the world, and, nested in it, a cognitive model of the self. Moreover, this nested structure should be updated in real-time, and all these features implied memory and learning abilities[7].

---

[1] Brandon, R. (1990). *Adaptation and Environment*. Princeton, NJ: Princeton University Press. Polgar, T., & Flanagan O.J. (.1999). *Explaining the Evolution of Consciousness: The Other Hard Problem.* Manuscript.
[2] Lewens, T. (2007). *Adaptation*. In D. Hull, M. Ruse (Eds.), *The Cambridge Companion to the Philosophy of Biology* (pp. 1-21). Cambridge, UK: Cambridge University Press.
[3] Flanagan, O. J. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.
[4] Piccinini, G., Robinson, Z., & Corey, J. (2015). Is Consciousness a Spandrel?. *Journal of the American Philosophical Association*, 1(2), pp. 365-383.
[5] Merker, B. (2005). The Liabilities of Mobility: A Selection Pressure for the Transition to Consciousness. *Animal Evolution*. *Consciousness and Cognition*, 14, pp. 89–114.
[6] Bronfman, Z. Z., Ginsburg, S., & Jablonka, E. (2016). The Evolutionary Origin of Consciousness. *Journal of Consciousness Studies*, 23(9-10), pp. 7-34.
[7] Ginsburg, S. & Jablonka, E. (2005). The teleological Transitions in Evolution: A Gántian View. *Journal of Theoretical Biology*, pp. 55-60.

As several studies have proposed[8], the origin of this cognitive unit can be traced back to the early Amniotes, which present features that suggest the emergence of a unitary dimension of cognition which could facilitate the discrimination and the ranking of the stimuli and the selection of the adequate response behaviours. Hence, through evolution it seems that this conscious strategy emerged from the complexification of the reflexive stimulus-response one, where the latter was able to rapidly respond to immediate stresses, but unable, unlike the former, to perform goal-directed behaviours.

The complexity of the organism-environment interaction, and the role of consciousness could be successfully compared to the structure of an advanced digital system, and this analogy may shed light on the problem whether *something like a self* could be practically useful or not.
A digital electronic system is composed of a hardware, an operating system and several pieces of software running on it, whether active or in the background[9]. The operating system – especially the *kernel* –, efficaciously arranges and controls the machine's resources (the hardware) in order to cover all the software requests, and this arrangement is realized according to *priority* and *importance*. This last point is of particular interest, as the concurrent digital environment which the operating system handles, presents many similarities with the natural and social environment in which every living being has to live and move. For this reason, the role of the operating system, often defined as a *resource manager*, could be considered similar to that of consciousness, which relates and ranks stimuli and decision processes providing them with a higher or lower priority, and in so doing it allows the organism to successfully interact with a highly concurrent environment
Without this kind of strategy, the ordering of the stimuli, as well as the ordering of the software requests, would be sequentially and blindly handled in order of arrival: nonetheless, what really matters for the purposes of conservation and efficiency is the ability to recognise the priority (or the *psychological salience*) of certain stimuli so as to dose the available cognitive resources.

So conceived, consciousness gave living beings a great evolutionary advantage: the ability to handle the complexity and the concurrency of their interactions with the environment, which means the ability of selecting the right behaviour in the right environmental conditions.

[8] Århem, P., Lindahl, B.I.B., Manger, P. R., & Butler, A. B. (2008). On the Origin of Consciousness. Some Amniote Scenarios. In H. Liljenström, P. Århem (Eds.), *Consciousness Transitions. Phylogenetic, Ontogenetic and Physiological Aspects* (pp. 77-96). New York, NY: Elsevier. Cabanac, M., Cabanac, A. J., & Parent, A. (2009). The Emergence of Consciousness in Phylogeny. *Behavioural Brain Research*, 198, pp. 267-272. Cabanac, M. (1996). On the origin of Consciousness, a Postulate and its Corollary. *Neuroscience and Biobehavioral Reviews*, 20, pp. 33-40.
[9] Tanenbaum, A.S. & Woodhull, A.S. (2006). *Operating Systems Design and Implementation*, Upper Saddle River, NJ: Prentice Hall. Bacon, J. & Harris, T. (2007). *Operating Systems: Concurrent and Distributed Software Design*, Boston, MA: Addison-Wesley.

Cognitive complexity and mathematical problem solving

In cognitive sciences, there is a widely accepted distinction (due to Marr 1982) between three levels of explanation: the computational level, the algorithmic level, and the implementation level. In this paper, we want to approach that distinction from one particular field of cognitive phenomena: mathematical problem solving. In the context of a mathematical problem, the three levels can be understood, respectively, as the computational complexity class of the problem, the algorithm used to solve the problem, and the neural action involved. It is commonly accepted in cognitive modelling that the first level, the computational explanations, should be the main focus. In this paper, we want to challenge that idea by asking whether it provides a sufficiently rich framework for explaining cognitive complexity in mathematical practice.

Computational complexity is an important subject in theoretical computer science and mathematics, but also in studying mathematical cognition. Based either on the space or the time required for a Turing machine to solve a problem, mathematical problems can be divided into complexity classes. The great fruitfulness of this approach makes it easy to see strong connections between cognitive complexity and computational complexity. When studying cognitive phenomena, the computational approach (e.g. Frixione 2001) aims to single out a function to model the particular phenomenon. Under this approach, the complexity of the cognitive task of solving a mathematical problem is identified with the complexity class of the problem, and nothing is assumed about algorithms and their implementations.

However, we want to suggest that for an account of cognitive complexity that aims to illuminate human problem solving in mathematics, focusing on the computational level is insufficient. While the implementation (neuronal) level is beyond our explanations at this point, the algorithmic level of cognitive explanations in mathematics can provide important insights. In the literature, computational complexity is often seen as the primary concept, with the complexity of problems determining the complexity of algorithms needed for solving them. This, however, is a characteristically abstract subject involving ideal computational agents, not one of actual mathematical practice and human agents.

In particular, we are focused on the heuristic aspects of mathematical problem solving. It has been suggested that heuristic procedures (e.g. Ausiello et al. 1999) can provide partial of approximate solutions to computationally intractable problems. While that approach seems misguided when it comes to mathematics, where such inexact solutions have limited value, heuristic procedures can have important relevance of another type when it comes to cognitive explanation in mathematical problem solving. Computational complexity looks for *computationally* optimal strategies for problem solving, but these can differ from *cognitively* optimal strategies.

Research from mathematics education shows that there can be a great difference between the optimal computational solution and students' capacity to understand it. The purpose of this paper is to elucidate this difference, thus contributing to a wider notion of cognitive complexity that also takes mathe-

matical practice into account. Beginning from such commonplace procedures as drawing pictures, human problem solvers use a wide array of heuristic tools which increase the algorithmic complexity of the solution, but which nevertheless makes it easier to understand. This way, human agents do not choose algorithms based only on minimal computational complexity, even understood locally within a specific field of mathematical practice. Heuristic procedures, connections to different areas of mathematics, as well as other areas of cognitive ability, provide an important part of mathematical understanding and problem-solving.

Finally, the topic is discussed in a wider context concerning philosophy of mind. We will argue that mathematical problem-solving gives us a well structured way to approach the larger question of explanations in cognitive science and their connections to philosophy of mind. There have been arguments by Lucas (1961) and Penrose (1989) against mechanical models of mind based on the incompleteness of arithmetical theories and the supposed human ability to supercede these limitations. Under the present approach, such arguments are seen to be misguided as algorithmic explanation is established in its proper context.

References

Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A., & Protasi, M. (2012). *Complexity and approximation: Combinatorial optimization problems and their approximability properties*. Springer Science & Business Media.

Frixione, M. (2001). Tractable competence. *Minds and Machines*, 11(3), 379–397.

Lucas, J. R. (1961). Minds, machines and Gödel. Philosophy, 36(137), 112-127.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman and Company.

Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics,* Oxford: University Press.

SEMANTIC ANTIREALISM AND THE SELF-ASCRIPTION OF ATTITUDES

I'll take semantic antirealism to be the doctrine that truth must be constrained by whatever epistemic notion fits a particular area of discourse and inquiry, so that if a statement is true, it must be possible to determine that it is true by means suited to the appropriate area. The doctrine is inherited from Dummett (see, e.g., *Dummett [1973] 1978*), with the proviso that Dummett's aim isn't to advocate it across the board. My purpose, though, isn't exegetical, but to find out what one's commitments must be should one wish to advocate the doctrine with respect to statements that register self-ascriptions of propositional attitudes, i.e., claims made by self-ascribers to the effect that they entertain beliefs, intentions, desires, and other attitudes. Should the semantics of ascriptions of the form "I ϕ that p," where "I ϕ" plays the part of the main clause and "that p" that of the subordinate declarative clause providing the content of whatever is ϕd, be of an antirealist persuasion?

There is room for a particular kind of Davidsonian interpretative strategy in the self-ascription case, where the charitable interpreters of their self-ascribed attitudes strive to eliminate internal inconsistencies and avoid, or at least curb, internal disagreements. I'll propose answers to three questions. (*i*) How should self-ascribers construe the warrants of their private self-ascriptions? In other words, what are the necessary and sufficient conditions for something to be a warrant of a self–ascription? Introspective checkings being likely candidates, what must they be like in structure and content from the ascribers' own subjective point of view? (*ii*) Are there any compelling reasons to think that these warrants must be effectively obtained in a finite time by the putative bearers, and should the limitations of their cognitive capacities be taken into consideration in order for the corresponding self-ascriptions to be legitimately deemed true? (*iii*) What is the logic of self-ascriptions? What formally valid inferences may self-ascribers legitimately draw, that would be justified by relations of logical consequence holding betwen self-ascriptions? In particular, which semantic principles should self-ascribers accept or, on the contrary, reject, that would justify the choice of a particular logic for statements of the form "I ϕ that p"?

These questions aren't psychological, at least not straightforwardly so. The point isn't about the sincerity of the putative bearers but about the *reasons* they must be able to provide for claims of the relevant form. One might be honest in claiming that one entertains a particular propositional attitude and yet be mistaken about what one believes, intends or desires. There are also cases where the retrievals that will yield justifications for self-ascriptions might involve a great deal of reflective introspection, as well as belief and desire revision.

My answers to the three questions involve two key points: a phenomenological point pertaining to the experiencing of warrants, and an epistemological point pertaining to their retrieval and

defeasability. I shall argue that, conjointly, they strongly suggest (rather than strictly speaking prove by reductio) that semantic antirealism for the self-ascriptions of attitudes fails on three counts.

Answer to question (*i*). The appropriate construal of the self-ascriber's subjective point of view requires that she experiences herself as ϕing that *p* so that the awareness of being the one who enjoys the ϕing is built in the warrant that a successful positive instrospection might uncover. Such experiences do represent the subject in whose mental life they occur (see, e.g., *Coliva 2002*). I shall argue that the ascriber's subjective awareness of her own ϕing that *p*, although phenomenal, is *also* propositional, and that checkings are *not* indifferent to the referents of the that-clauses of first-order attitudes, in particular beliefs (see, e.g., *Pryor 2005* for a discussion).

Answer to question (*ii*). The surveyability constraint requires that self–ascribers check whether they ϕ that *p* so that the checking *doesn't* tolerate marginal increases in lengh and complexity beyond a given level *L*. The cognitive limitations constraint requires that sub-optimal capacities be taken into consideration when predicating truth. Both constraints yield psychological and epistemological puzzles, both when self-ascriptions are stable under the rescrutiny of further accredtions to our knowledge (see *Wright [1982] 1993*) and in the case of so-called "entitlements."

Answer to question (*iii*) The properties at stake are the doxastic positive introspective property for bimodal propositional logic ($B\phi \rightarrow KB\phi$) and the positive introspective property for epistemic propositional logic ($K\phi \rightarrow KK\phi$) holding in all Kripke structures. Are the axiom schemata of Kraus and Lehmann's bimodal system KB_{CD} (*Kraus and Lehmann 1988*) and of Voorbraak's *S*5+*KD*45 system (*Voorbraak 1992*) legitimate? The systems have a clear advantage: they provide separate accessibility relations for individual knowledge (K_i) and individual belief (B_i) such that the bridge axiom schema $B_i\phi \rightarrow K_iB_i\phi$ holds. This is exactly what a semantic antirealist would want. The crux of the matter is whether these axiom schemata are psychologically and epistemically fit, or indeed relevant, when highly mediate warrants must play an epistemic role and when positive introspection, phenomenologically construed, calls for second-order mental states.

MINIMAL REFERENCES

Coliva (Annalisa), *2002*, "On What There Really is to Our Notion of the Ownership of a Thought. A Reply to John Campbell," *Philosophy, Psychiatry and Psychology* 9/1, pp. 41-6.

Dummett (Michael), *[1973] 1978*, "The Philosophical Basis of Intuitionistic Logic," in *Truth and other Enigmas*, Duckworth, London, 1st ed., pp. 215-247.

Kraus (Sarit) and Lehmann (Daniel), *1988*, "Knowledge, Belief and Time," *Theoretical Computer Science*, vol. 58, pp. 155-174.

Pryor (Jim), *2005*, "There is immediate justification," *Contemporary Debates in Epistemology*, E. Sosa and M. Steup, eds., Basil Blackwell, Oxford, pp. 181-202.

Voorbraak (Frans), *1992*, *Generalized Kripke models for epistemic logic*, Morgan Kaufmann Publishers, San Mateo, pp. 214-228.

Wright (Cripsin), *[1982] 1993*, "Strict Finitism," *Realism, Meaning and Truth*, Basil Blackwell, Oxford, 2nd ed., pp. 107-175.

A Neurobiological Case for Animal Rationality

In the following, I make the case that non-human animals, including such phylogenetically distant creatures as rats, are rational. Recent work in systems neuroscience, I argue, provides powerful evidence that other animals are capable of both deliberation and reflection. Consonant with demands by proponents of human uniqueness, the underlying mechanisms are flexible, finding application in a variety of contexts, and capable of dealing with abstract, relational concepts. Though many regions are involved, I focus on an evolutionarily conserved circuit consisting of prefrontal cortex, hippocampus and their homologues. Both, I contend, play essential and preserved roles in human and non-human reasoning. Though human thought may have greater depth and scope than its animal counterpart, and while it is augmented by language and by sophisticated sensorimotor processes, the basic processes of reasoned decision-making are fundamentally the same.

Now the term rationality has many different meanings. There are senses in which humans aren't rational and notions where the simplest of organisms are. Even limiting ourselves to notions from comparative psychology and related disciplines, the sense at issue here, there is room for debate. For the purposes of argument, I adopt what I take to be the more conservative standards available. To be rational (or capable of "human-like" cognition), an animal will need to engage in an explicit process of planning and assessment, "decentring" from any habitual, stimulus-driven response tendencies (Hurley and Nudds, 2006) and drawing inferences in a rule-governed manner. In addition, and in keeping with demands laid out by prominent skeptics in the animal cognition literature (e.g., Penn et al., 2008), the representations employed in these processes must be both schematic and flexible. In other words, the same explicit, relational structures should be deployed not simply in one or two kinds of cases but across domains and degrees of abstraction. That humans benefit from cognitive mechanisms of the sort described above is commonly taken for granted. Structural relations like membership, for instance, pop up everywhere from family trees to time. Yet the extent to which they are present in other species has long remained a point of controversy. Even closely related species like chimpanzees have been argued to lack such abstract and flexible thought.

If I am correct, they may be present in monkeys and perhaps even rats. To make the case, however, it is necessary to draw on a body of research that has received comparatively little attention in the debate: namely, systems neuroscience. I focus specifically on research involving PFC and hippocampus, which I argue can be taken to implement broad planning and detailed assessment, respectively. In humans and animals alike, I contend, PFC and its homologues cue and maintain abstract strategies and task-relevant information in working memory. This is done in a wide variety of contexts and problems, working not only with perceptually-grounded features like as color but with abstract categories organized by things like numerical value (Nieder, 2016) and outcome (Tsutsui et al., 2016). In each case, I show, firing patterns provide evidence of a schematic representational structure corresponding to the type of problem faced. From PFC, task information is relayed to hippocampus (Ishino et al., 2017), which draws out inferences from previous experience and, to some extent, imagination. The region is perhaps best known for containing "cognitive maps," cellular assemblies that encode spatiotemporal relations in a topologically organized format. In recent years, however, it has also been implicated in temporal, social, and other forms of relational memory (Eichenbaum, 2017). Most strikingly, work conducted over the last decade has found that the region underwrites a process known as "vicarious trial and error" in which activity within cognitive maps literally traces alternative paths ahead of the animal one after the other (Redish, 2016). The locations that the paths go to need not be immediately perceptible, and as the recent work of Mou and Ji (2016) demonstrates, it need not have ever been run by the animal itself.

Given the functional properties of the regions discussed and their consistency with functions ascribed to human homologues, I argue, many animals have a strong case. The fact that these results were all obtained in studies of (relatively) distantly related animals does, however, raise a few questions. In closing, then, I consider what these results might tell us about the evolutionary functions of "higher"

reasoning and why, if present, such reasoning has been so difficult to conclusively show in behavioral tests of phylogenetically closer species like chimpanzees.

References

Dennett, D. (1996). Kinds of minds: Toward an understanding of consciousness. Basic Books.

Eichenbaum, H. (2017). Memory: Organization and Control. *Annual Review of Psychology*, *68*(1), 19–45.

Hurley, S., & Nudds, M. (2006). "The questions of animal rationality: Theory and evidence" in Hurley and Nudds (Eds.). *Rational animals?* (1-83). Oxford: Oxford University Press.

Ishino, S., Takahashi, S., Ogawa, M., & Sakurai, Y. (2017). Hippocampal-prefrontal theta phase synchrony in planning of multi-step actions based on memory retrieval. *The European Journal of Neuroscience*, *45*(10), 1313–1324.

Mou, X., & Ji, D. (2016). Social observation enhances cross-environment activation of hippocampal place cell patterns. eLife, 5, e18022.

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366–382.

Penn, D., Holyoak, K., & Povinelli, D. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. Behavioral and Brain Sciences, 31(02), 109–130.

Tsutsui, K., Oyama, K., Nakamura, S., & Iijima, T. (2016). Comparative Overview of Visuospatial Working Memory in Monkeys and Rats. *Frontiers in Systems Neuroscience*, *10*.

Recalcitrance Poses No Threat to Cognitivism

The primary argument against cognitivist accounts of emotion relies on an appeal to recalcitrant emotions to show that cognitive content such as judgment or belief is not a proper part of emotion (e.g. Greenspan 1981, D'Arms and Jacobson 2003). However, many mental processes that we are happy to deem cognitive exhibit the very inconsistency with which the argument from recalcitrance takes issue. Hence, this paper aims to show that appeals to recalcitrant emotion cannot tell against cognitivist theories of emotion and cannot establish that emotions are irrational. Toward this end I will argue both that the original argument from recalcitrance fails, and that Grzankowski's (2016) recent modification to that argument fails.

The primary argument against cognitivism relies on the purported puzzle about recalcitrant emotion. This argument is originally due to Greenspan (1981), but it has enjoyed quite a long and robust record, with theorists including D'Arms, and Jacobson (2003), Prinz (2006), Deonna and Teroni (2012), Benjabi (2013) citing recalcitrant emotion as the major problem facing cognitivism. In her original formulation, Greenspan claims that in order to accommodate recalcitrant emotion the cognitivist must attribute inconsistent judgments to the subject experiencing recalcitrant fear in the face of a dog whom the subject knows to be friendly. The cognitivist must, according to Greenspan, attribute to the subject at once the judgment that Fido is dangerous and the judgment that Fido is not dangerous. By the non-cognitivists' lights this amounts to attributing undue irrationality to the subject having the recalcitrant emotion: The cognitivist must say that the agent is systematically irrational whenever undergoing recalcitrant emotion. This, Greenspan urges, is problematic and is motivated only by the cognitivist's desire to preserve cognitivism.

I take issue with the assumption above that a subject who has inconsistent cognitive states (e.g. the belief that *p* and the belief that *–p*) is irrational. Indeed I take it that other areas in philosophy of mind have widely accepted this kind of inconsistency of judgment or belief as compatible with rationality. Within philosophy of mind there is wide acceptance that consistency and closure cannot be requirements for rationality (e.g. Lewis 1982, Stalnaker 1984, Davidson 2004, Egan 2008, Greco 2015, Yalcin 2016). Attributing contradictory judgments or beliefs to otherwise rational subjects seems necessary for making sense of human behavior. The literature on belief fragmentation and the underlying phenomenon that fragmentation seeks to explain, though so far overlooked in the debate over recalcitrant emotions, provides an excellent footing for the cognitivist to object to the argument from recalcitrance. Consider Lewis' famous case of the person who believes that Nassau Street runs North/South, that the railroad tracks run East/West, but also believes that Nassau Street and the railroad tracks are parallel to one another. This person has never considered all three beliefs together and so has not noticed that they are mutually inconsistent. Humans make these sorts of mistakes all the time. And these inconsistencies are not the plight only of those we would deem irrational. On the contrary, we all seem to suffer these inconsistencies fairly regularly—though perhaps more likely than we'd like to admit.

Thus it is fairly uncontroversial that even paradigmatically cognitive states such as belief exhibit the inconsistency that Greenspan finds objectionable in cognitivist accounts of recalcitrant emotion. If it is right that we have inconsistent beliefs regularly, and if we also preserve the assumption that people are generally rational, then having some inconsistent cognitive states must be consistent with rationality. That is to say, if inconsistency of cognitive states is widespread in normal subjects, then the attribution of some inconsistent judgments or beliefs to a person cannot amount to the attribution of irrationality. This runs counter to the anti-cognitivist assumption mentioned above.

However, Grzankowski (2016) argues that there is a more intractable problem from recalcitrance than Greenspan's formulation of the argument from recalcitrance suggests. In his view, to preserve the assumption of rationality the cognitivist will have to say that every time the subject has a recalcitrant emotion the subject is entertaining the contents of the relevant cognitive states under different modes of presentation or using different concepts. And yet, Grzankowski claim, the subject having the recalcitrant emotion will almost always take herself to be using the same modes of presentation or concepts for both of these contents when she has recalcitrant emotion. By his lights, introspection normally allows us to tell

whether two current thoughts we are having are entertained under the same mode of presentation or using the same concepts. Furthermore, Grzankowski urges, in cases of recalcitrant emotions, the subject will upon introspection usually report that she is entertaining the relevant contents under the same mode of presentation or using the same concepts. However he argues that the cognitivist, to preserve the assumption of rationality and also to preserve cognitivism, must claim the subject is always entertaining the contents of recalcitrant emotion under different concepts or modes of presentation *despite what she might introspectively report*. Grzankowski seems to think this would be tantamount to claiming that the subject of a recalcitrant emotion is *never* able to introspectively determine under what modes or concepts she is entertaining the contents of her thoughts during recalcitrant emotion. This would indeed amount to an attribution of undue, systematic opacity

My argument against Grzankowski's modified argument from recalcitrance runs as follows. Grzankowski supposes that a subject who has two inconsistent beliefs is not irrational if she entertains the contents of those beliefs using different concepts or under different modes of presentation. This might be so, but it could very well turn out that in everyday cases of inconsistent beliefs we actually *do* entertain contents under the same modes of presentation or the same concepts. Perhaps we have conflicting beliefs about the very same content *entertained under the very same mode of presentation*, but fail to notice the inconsistency because it does not influence our behavior, and does not create problems for our actions and goals, since we never happen to entertain the two beliefs under the same circumstances.

Secondly, Grzankowski claims that when a subject of a recalcitrant emotion introspects, she usually discovers that she is entertaining the contents in question under the *same* mode of presentation. However, it is not at all clear why this would be. As far as I can tell he doesn't give any reason in defense of this claim—that is, aside from stressing that the subject in Greenspan's original example would report that her inconsistent beliefs are both about *the very same dog* considered in *the very same way*. It is not obvious that this sort of thing is introspectable for most individuals, not least because it is probably not clear to most subjects whether they're entertaining a given content under one mode of presentation or another. What, after all, does it mean to think about something *in the very same way*? This is a very abstract philosophical question and not one, I think, that most subjects can readily make use of in introspection. Furthermore, even if most subjects *could* introspectively report on this, it is not obvious that they would report that during recalcitrant emotions they are entertaining the same contents under the same modes of presentation in their inconsistent cognitive states.

However even if my above arguments go through there is still the apparent puzzle of recalcitrant emotions, which seems to count against the cognitive nature of emotion even from the layperson's perspective. To dismantle this puzzle I suggest that the reason that emotions may seem more recalcitrant, or more often recalcitrant, or even more problematically recalcitrant, than other kinds of mental states is that emotions are more motivationally central to us and so we are more likely to *notice* their recalcitrance. Additionally, almost always the emotions that we discuss when we discuss recalcitrant emotions are so-called 'negative' emotions. We don't usually want to get rid of 'positive' emotions such as joy, hope, etc., though of course there are some cases where we do. Nevertheless, it seems that our general negative attitude toward the 'negative' emotions makes us more likely to notice them as recalcitrant than to notice a recalcitrant, inconsistent belief.

Submission for PLM4

A Puzzle for Semantics

Thesis

The thesis of this talk is twofold:

(A) Semantics faces an important trilemma in light of the liar paradox.
(B) To resolve the trilemma, we may need a scientific anti-realist attitude towards Semantics.

Structure and Main Arguments

1. Introduction

Semantics is the science of linguistic meaning. It involves using linguistic data to construct formal *semantic theories*, which purport to describe the meaning of the linguistic expressions of natural language.

2. Three widely held views

Here are three views about the aim of semantics that are individually and jointly popular.

(i) *Semantics is in the business of modelling semantic competence.*
E.g., semantic theories might be taken as models of the *processing* that takes place in the semantic module, or as models of what the competent speaker *knows*, or as models of what is *conventionally encoded* by words/sentences, etc.

(ii) *Semantics is in the business of assigning truth conditions to utterances (or sentences-in-contexts, etc.)*.
The view I am describing here is *not* that truth conditions *exhaust* meaning. I take (ii) to be compatible with the view that semantic theories *also* assign e.g. Austinian illocutionary force, or something else non-truth-conditional, to utterances.

(iii) *Semantics is in the business of discovering the truth*.
That is, as a science, semantics is aiming at the truth. E.g., if Semantics ultimately yields a semantic theory that proves the theorem "X is true iff *p*", then it is in fact *true* that: X is true iff *p*.

Two clarifications:

- While (i)–(iii) are popular, none is universally accepted.
- I make no claim that anyone thinks that (i)–(iii) provide a *complete* statement of the aims of Semantics.

An example: on Emma Borg's Minimal Semantics, each of (i), (ii) and (iii) are true.

3. Trilemma

I present an argument that, in light of the liar paradox, we cannot coherently endorse all of (i)–(iii).

A couple of preliminary points.

- The liar paradox arises when we consider the truth value of the liar sentence, λ:

  > λ is not true.

  Intuitively, the truth of λ would imply its untruth; and the untruth of λ would imply its truth. Thus, by the law of the excluded middle, we apparently derive that λ is *both* true *and* not true.

- Liar sentences arise in natural languages. E.g.:

  > The only sentence in the abstract titled "A Puzzle for Semantics" submitted for PLM4 that mentions the word "foxtrot" is not true.

  For simplicity, I henceforth assume that λ, as introduced above, is a sentence of English.

The main argument for (A) proceeds by assuming (i) and (ii), and then arguing for not-(iii):

- Assume (i) and (ii).
    - By (i), a semantic theory should match the competent speaker's semantic intuitions and judgements. Which is to say, we can use linguistic and semantic evidence in the usual way to test semantic hypotheses for λ.
    - By (ii), a semantic theory should assign truth conditions to sentences. Which is to say, a semantic theory for English should ultimately assign a truth condition to λ.
- In light of evidence concerning λ that I have published elsewhere, we have good reason to think that Semantics will ultimately assign a truth condition to λ along the following lines:

  (T) "λ is not true" is true iff λ is not true.

  In the talk, I will briefly sketch linguistic evidence that: λ can be embedded under logical operators to form intuitive truths, which is explainable only if λ has content *that λ is not true*; λ passes standard 'cancellability' tests for possessing content *that λ is not true*; and λ fails standard tests for context sensitivity.
- Given the fact that λ *is* "λ is not true", (T) classically entails the following:

  (X) λ is true and λ is not true.

  However, (X) is a contradiction and thus (*pace* dialetheism) is not true. Thus (by *modus tollens*) (T) is not true.
- Thus, in the light of the liar paradox, we have good reason to think that Semantics will ultimately yield an untrue semantic theory.
- Thus, Semantics is not in the business of discovering the truth. That is, not-(iii).

If this argument is right, then we cannot coherently endorse all of (i), (ii) and (iii).

4. Towards scientific anti-realism

Some unattractive responses to the trilemma:

- Deny (i). Objection: the use of evidence in Semantics is by-and-large coherent only on the assumption that (i) is true. Denying (i) is too revisionary.
- Deny (ii). Objection: some of the most successful results in Semantics have been in the truth-conditional paradigm. We should not reject it outright by denying (ii) unless absolutely necessary.
- Deny my treatment of the linguistic data. Objection: the broad aims of Semantics ought not to be dependent upon the specific interpretation of a few pieces of data.

Thus, I recommend denying (iii): Semantics is *not* in the business of discovering the truth.

But how do we make sense of this conclusion? To do so, it seems we should adopt a scientific anti-realist position towards Semantics.

- Scientific anti-realism gives us the resources to explain how science can (rationally) lead to theories that are not true.

References

Austin, J. L. 1962. *How to Do Things with Words*. Oxford: Oxford University Press.
Borg, E. 2004. *Minimal Semantics*. Oxford: Oxford University Press.

Degrees of Analyticity

Overview Recent work by Gajewski (2008) and Chierchia (2013) argues that some tautological or contradictory propositions are marked as ungrammatical. For example, Gajewski argues that the exceptive in (1) entails both that it's not that case that some women smiled, contradicting the assertion. Chierchia argues the distribution of NPIs like *any* can be explained along similar lines.

(1) *Some women except Mary smiled.

In this contribution, we claim that the logical status of propositions explains a deviance recently derived in the literature on scalar implicatures. Pistoia-Reda (2017b) shows that scalar implicatures are not generated when quantifiers range over non-existent and inconceivable entities (e.g. round squares). On the other hand, he also shows that they can be generated when quantifiers range over non-existent but conceivable entities (e.g. present Greek kings). His account of the asymmetry is based on the notion of modalized presupposition. The idea is that scalar implicatures can only arise when the existence presupposition attached to the universal alternative proposition is satisfied in at least one metaphysically possible world. However, we show that this account provides only a partial explanation of the data. In particular, the account explains why the scalar implicature is not derived when quantifiers range over non-existent and inconceivable entities, but it does not explain why the base proposition is deviant. We note that the base proposition is necessarily false and that the universal alternative proposition is necessarily true. This would account for the absence of the scalar implicature and for the deviance of the base proposition if we assumed that analyticity results in ungrammaticality.

Background Pistoia-Reda (2017b) argues that a sentence like (2a) does not give rise to any scalar implicature because the quantifier ranges over inconceivable entities. On the other hand, he also argues that a sentence like (3a), where the quantifier ranges over non-existent but conceivable entities, does in fact give rise to a mismatching scalar implicature, and that this is the source of the observed oddness of the sentence. In particular, the scalar implicature to the effect that the alternative proposition (3b) is false contradicts the piece of contextual knowledge according to which all Greeks come from the same country, irrespective of their occupation or social status (see Magri (2009), Pistoia-Reda (2017a) for discussion).

(2) a. Some round squares are rectangles
b. Alt = { All round squares are rectangles }

(3) a. Some Greek kings come from a warm country
b. Alt = { All Greek kings come from a warm country }

The asymmetry is confirmed by the observation that, when taken in isolation, the universal alternative of (2a), repeated as (4), sounds clearly infelicitous whereas the universal alternative of (3a), repeated as (5), seems to improve. Of course, this could be due to the fact that the universal quantifier is the highest element in the Hornian scale of positive quantifiers, and so in a standard linguistic environment there is no stronger potential alternative to be negated, i.e. no scalar implicature can be derived (Horn (1989); but see also Spector (2014), Katzir (2007)).

(4) All round squares are rectangles

(5) All Greek kings come from a warm country

The existing account Pistoia-Reda suggests that the absence of the scalar implicature in the case of (2a), as well as the oddness of(4), could be due to the fact that universal propositions are associated with a *modalized* rather than with an *existence* presupposition, as assumed in standard accounts. The modalized presupposition, which is defined in terms of compatibility with the common ground, is assumed to be satisfied whenever the entities quantified over by the universal quantifier are actual in at least one metaphysically possible world, i.e. whenever the standard existence presupposition is satisfied in at least one metaphysically possible world. Since the modalized presupposition carried by (4) is not satisfied, the alternative is undefined, and the scalar implicature cannot possibly arise. When the universal proposition is taken in isolation, the observed oddness is also expected. On the other hand, since the modalized presupposition carried by (5) is satisfied, the alternative

is defined; thus, the mismatching scalar implicature can arise, and is in fact generated, yielding the observed oddness.

The present account We would like to submit that the oddness of the universal proposition (4) would be expected if we assumed, following work by Gajewski (2008) and Chierchia (2013), that analytic truths result in ungrammaticality. Note that this proposition is analytically true. Since round squares do not exist in any metaphysically possible world, the antecedent of the implication contained in the proposition is necessarily false; as a consequence, the entire implication will be necessarily true. Interestingly, if we assume that the logical properties of propositions can result in deviance, our suggestion would also explain the oddness of (2a). Since round squares do not exist in any metaphysically possible world, the intersection between the set of round squares and the set of rectangles is necessarily empty; the existential proposition will consequently be necessarily empty. The deviance of (2a) would be immediately accounted for. In addition, it is important to underline that the present proposal also accounts for the oddness of other existential cases, like (6) below. This proposition is construed with the quantifier ranging over non-existent and inconceivable entities but the scalar implicature possibly attached to the proposition in question would not amount to a contradiction. In other words, if we expect oddness effects only from mismatching scalar implicatures the following deviance would be surprising.

(6) Some round squares are green

Pistoia-Reda is right in maintaining that a sentence like (6) does not generate any scalar implicature (note that the modalized presupposition would be unsatisfied in this case). However, we note that his account has nothing to say when it comes to the deviance of the two base existential propositions construed with non-existent and inconceivable entities, that is (6) and (2a). What we are saying is that the modalized presupposition account provides merely a partial empirical account of the available data, for it is limited to the cases involving universal propositions and their presuppositions.

References

Chierchia, Gennaro. 2013. *Logic in grammar*. Oxford: Oxford University Press.

Gajewski, Jon. 2008. NPI *any* and connected exceptive phrases. *Natural Language Semantics*, **16**(1), 69–110.

Horn, Laurence. 1989. *A natural history of negation*. Chicago: The University of Chicago Press.

Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy*, **30**(6), 669–690.

Magri, Giorgio. 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural language semantics*, **17**(3), 245–297.

Pistoia-Reda, Salvatore. 2017a. Contextual blindness in implicature computation. *Natural Language Semantics*, **25**(2), 109–124.

Pistoia-Reda, Salvatore. 2017b. On conceivability and existence in linguistic interpretation. *Lecture Notes in Artificial Intelligence*, **10257**, 203–212.

Spector, Benjamin. 2014. Scalar Implicatures, Blindness and Common Knowledge: Comments on Magri (2011). *Pages 146–169 of:* Pistoia-Reda, Salvatore (ed), *Pragmatics, semantics and the case of scalar implicatures*. Palgrave studies in pragmatics, language and cognition. Basingstoke, UK: Palgrave Macmillan.

Two critical remarks to classical identity theory

Contemporary empirical science about mind/consciousness usually describes the relation between mental states and states of brain applying the term "correlation" (Koch et al. 2016; Crick and Koch, 1990; Chalmers 2000). The concept is general enough to respond flexibly to the future knowledge that might specify a particular form of the relation. In search for the best matching concept, however, both philosophers and empirical scientists have proposed a number of other concepts that could be suitable candidates for mind-brain relation. So one can encounter formulations like that neural processes "determine" consciousness (Block 2007), or that conscious phenomenal contents are "entailed by activity of neuronal systems that create consciousness" (Edelman 2003). In neuroscientific texts, especially those dealing with research into consciousness, there can be found terms as "association", "correspondence" (Frith et al. 1999), "signature" (Dehaene 2014), or the concept of "causality" (Neisser 2011).

In current naturalistically oriented philosophical literature pursuing this issue, there are number of attempts to account for relation between mental states and their founding neurobiological mechanisms. Although none of them has got prevailing position, some attempts acknowledge that mental states must somehow be the brain states. Scholars thus speak about realization, constitution, determination, composition or supervenience (Keaton 2015; Bennett 2011; Kozuch, Kriegel 2015; Bickle 2013). One of the most significant concepts directly supporting the view that states of mind are neural states is the concept of identity.

Identity theory belongs to historically well-known attempts capturing the mind-body relation (Smart 1959; Place 1956; Lewis 1966; Feigl 1958). Although the concept of identity has been rejected for number of reasons as an inadequate solution (e.g., Kripke 1980; Putnam 1967; Chalmers 1996), its advantages are currently debated again (Gozzano, Hill 2012), especially with regard to the type identity theory.

The purpose of the paper is to support the tendency to revitalize identity by identifying two novel issues within the traditional philosophical discussion of identity theory. These issues are aimed as clarification for future research into identity theories. First issue concerns Kripke's consideration in which he refuses to apply identity to the relation between pain and C-fibers firing. Kripke argues that empirical identities like “light is electromagnetic radiation” or “heat is molecular motion” are in fact not analogical to cases like “pain is C-fibers firing” (PisC) and therefore they can hardly support PisC to be an identity at all, the less empirical identity. The aim is to show that Kripke is wrong in this regard. He does not distinguish sufficiently the meaning of individual relata that enter the identity relation. According to him the term pain is categorically different from the term heat: pain is feeling of pain, but heat is not feeling of heat. This misunderstanding then prevents him to apply identity of heat and molecular motion in the same way as in case of pain and C-fibers firing. Therefore his criticism of the analogy between heat and molecular motion, and pain and C-fibers firing, seems to be wrong. If the relata are properly defined, then the cases have to be analogous.

The second issue concerns Lewis' application of Fregean semantic distinction the Morning Star – the Evening Star. While Kripke argues against identity theory, Lewis believes that it is the right solution to the mind-body/brain problem. The way he argues in favour of this thesis is based on Fregean distinction. Applying this distinction on mind-brain identity thesis seems to have a fatal shortcoming that, to my best knowledge, has not yet been addressed. This paper argues that the statement of identity between the Morning Star and the Evening Star is not a well-done analogy to the assertion about the identity between feeling of pain and C-fibers firing. The most confusing is a question what kind of object is supposed to be in the position of Fregean reference (Bedeutung) if “pain” and “C-fibers firing” label different ways of being given (Sinn).

The goal of this paper is not to reject the theory of identity as such: it is only to point out that some of its traditional semantic and categorical delimitations are not justifiable and, moreover, are not compatible with how contemporary science considers the relation between mind and brain. Current formulations of identity theorists thus must be free from the above mentioned misleading analogies. Despite these points, the theory of identity remains viable as a theory of mind-brain relation and paying more attention to it makes good sense.

Context and communicative success

There are a range of competing accounts of communicative success in the literature. For example, some authors adopt a 'shared content' approach to communication. On this kind of account, for communication to succeed, the hearer in a speech exchange must grasp the very content that the speaker expressed with her utterance. This kind of account is appealed to in objections to theories of content which cannot posit widespread shared meaning (Newman 2005, Cappelen and Lepore 2007); the view also often underpins accounts of how we gain knowledge from testimony (Burge 1993, Fricker 2006, Goldberg 2007). In opposition to the shared content view are various similarity-based approaches to communication. For example, some authors have suggested that communicative success requires, not sameness of content, but merely some suitable degree of similarity between the content grasped by the hearer and the content expressed by the speaker (Pollock 2015; Bezuidenhout 1997; Carston 2001). A related approach is to claim that communicative success is graded: communication succeeds to the degree that the hearer understands the speaker's utterance (Bezuidenhout 1997). Whilst these various views offer different pictures of communicative success, they all endorse a kind of invariantism: they claim that the conditions on communicative success remain the same across different contexts. For example, shared content views claim that a hearer must always grasp the same content that the speaker expressed; graded views claim that success is always proportional to the hearer's degree of understanding. In this talk, I will argue that these invariantist approaches are problematic; I will propose instead that the standards for communicative success vary with the context of the speech exchange.

The problem with the invariantist approach is that its picture of communicative success doesn't reflect the reasons for which we attempt to communicate with each other. When we attempt to communicate, it is always with the intention to produce some effect in the hearer – for example, to persuade our interlocutor to pass the salt, or to convince her that it is raining. These intentions are perlocutionary intentions in Austin's (1962) sense. These different intentions can make different demands on the hearer. To take a toy example: if I utter 'There is a cassowary in the field' with the intention that you form the belief that there is a flightless bird in the field, it will matter that you understand that a cassowary is a flightless bird. However, if I produce this same utterance with the intention of warning you that there is something dangerous in the field (and convincing you to heed this warning), you need only understand that cassowaries are dangerous – your further beliefs concerning the nature of cassowaries are not relevant to success, considered relative to my intention.

In this talk, I present a view of communicative success that takes this variation in speaker intention into account. The view I will defend is a kind of similarity view, in that it claims that communicative success requires that the speaker and hearer's understanding of an utterance must be similar to some specified degree (rather than claiming that content must be shared). However, it differs from the invariantist views sketched above in that it claims that what counts as sufficiently similar understanding will vary with context. This view is thus a context-dependent view rather than an invariantist one; in particular, the view is that a speaker's perlocutionary intentions determine contexts with respect to which the success of the exchange should be judged. There are different ways to flesh out the details of a context-dependent view. For example, one might claim that success requires some contextually determined degree of *overall* similarity in utterance understanding between speaker and hearer. However, I will argue for a slightly more complex version of the view. Utterance understanding can vary along different dimensions. To return to the example from above: an individual can understand that a cassowary is something dangerous without understanding that it is a flightless bird (and vice versa). Only certain of these dimensions will be relevant to the speaker's intentions in a given speech exchange. Given this, I

will argue that the most attractive version of the context-dependent view is one which claims that communicative success requires, not overall similarity, but similarity along certain *relevant* dimensions of utterance understanding. On this version of the view, the speaker's intentions determine not just how similar two interlocutors' utterance understanding must be, but also the relevant dimensions along which this similarity must hold.

The role of context in the *processes* by which hearers recover interpretations of speakers' utterances has been well explored by relevance theorists (Sperber and Wilson 1986). However, authors have rarely investigated how context might figure in an account of the success relation itself (although see Bezuidenhout 1997). What is distinctive of my proposed view is that it allows us to say both (a) that a communicative attempt can succeed even when the hearer's understanding of the speaker is quite poor overall and (b) that a communicative attempt can fail even when the hearer's understanding is very good. Furthermore, the account can allow that the same exchange may be judged a success with respect to one context and yet a failure with respect to some further context (given that speakers can produce utterances with multiple intended effects). Competing accounts do not have this flexibility. I will argue that consideration of a range of examples gives us reason to think that such flexibility is an attractive feature of an account of communication.

Bibliography

Austin, J. L., 1962, *How to do things with words, 2nd edition*, M. Sbisà and J. O. Urmson (eds.), Oxford University Press, 1975.

Bezuidenhout, A., 1997, 'The communication of *de re* thoughts', *Noûs*, 31: 197–225.

Burge, T., 1993, 'Content preservation', *The Philosophical Review*, 102: 457–488.

Cappelen, H., & Lepore, E., 2007, 'Relevance theory and shared content', in N. Burton-Roberts (ed.), *Advances in pragmatics*, Palgrave Macmillan.

Carston, R., 2001, 'Explicature and semantics', in S. Davis & B. Gillon (eds.), *Semantics: A reader*, Oxford University Press.

Fricker, E., 2006, 'Testimony and epistemic autonomy', in J. Lackey and E. Sosa (eds.), *The epistemology of testimony*, Oxford University Press.

Goldberg, S., 2007, *Anti-individualism*: *Mind and language, knowledge and justification*, Cambridge University Press.

Newman, A., 2005, 'Two grades of internalism (pass and fail)', *Philosophical Studies*, 122: 153–169.

Pollock, J., 2015, 'Social externalism and the problem of communication', *Philosophical Studies*, 172, 3229-3251.

Sperber, D. & Wilson, D., 1986, *Relevance: Communication and cognition*, Blackwell.

Is the debate about doxasticism merely verbal?

The aim of the paper is to explore the metaphysical presuppositions of the doxasticism debate, i.e. the discussion whether delusions should count as beliefs. We argue that there are good reasons to worry that this debate is merely verbal, in the sense that both parties might be thought to be right, given charitable interpretation of their theses. The most natural way to avoid such a conclusion is to make a case for the claim that "belief" and "delusion" are natural kind terms. We aim to show that while "delusion" might reasonably be said to refer to a natural kind, it is much more problematic to say that "belief" does, especially given Schwitzgebel's account of folk psychology.

Adherents of the view that delusions are beliefs, typically use arguments from assertion or from first person reports to support their position (Bortolotti & Miyazono, 2015). The response to that is that delusions cannot be beliefs because beliefs are rational and delusions do not behave in a rational way (Currie & Ravenscroft 2002). Some theorists offer arguments based on the apparent irrationality of everyday beliefs to conclude that delusions could be beliefs despite their irrationality (Bortolotti, 2010).

Although support for the doxasticism is very popular among philosophers a variety of alternatives have been presented in recent literature. Some try to fit delusions somewhere between beliefs and non-belief states while others attempt to find other status for delusions. Thus, scholars of the former kind hold the view that delusions are bimaginations i.e. things located between belief and imagination (Egan, 2009; Currie and Jones, 2006) or that delusions are 'in-between beliefs' (Schwitzgebel, 2012). Arguing for the latter view, some scholars try to prove that we should build separate category for delusions. According to them, delusions could be perceptual states (Hohwy and Rajan, 2012) or acceptations generated by strong cognitive feelings (Dub, 2015).

The persistence of the doxasticism debate and lack of clear conclusions may lead to a question whether there is indeed such a thing as a right answer to the question whether delusions are really beliefs. A metaphilosophical sceptic might argue that the lack of clear conclusions in this case proves that this is yet another case of a merely verbal debate, in which philosophy abounds.

We are going to assume the conception of verbal debate developed by Eli Hirsch (2005; 2009). In order to show that a debate is verbal one needs to show that it is possible to read the claims of the participants of the debate in such a way as to show that on one reading the contested sentence turns out to be true and on the other – false. In our case the sentence in question is "Delusions are beliefs". If it was possible to provide such interpretations of the both key terms of this statement as to get the result that this claim is true on one reading and false on the other, then the debate would be verbal.

A metaphilosophical sceptic might claim that the disagreement between a doxasitcist and an anti-doxasticist stems from the fact that doxasticists and anti-doxasticist differently understand the terms "belief" and "delusion"; for example they differently conceptualize the definitional aspect of what is needed to count as genuine belief and genuine delusion. Consider Bortolotti's account of beliefs in her modest doxasticism (Bortolotti, 2010). She argue that beliefs might be characterized as opposed to desires and difference between those two might be cashed out in terms of so called dimensions: procedural (relation between beliefs versus relations between desires), epistemic (sensitivity to the evidence), agential (manifestation in behaviour) (Bortolotti, 2010). The core tenet of this account is strict pairing between these dimensions and rationality. Nevertheless central features of beliefs might be characterised in a different manner. Stephens and Graham (2004) stress out that the central features of beliefs are: representational content, subjective conviction and reason-giving. This only partially overlap with Bortolotti's account. Usually, the chosen characterisation of belief is mirrored in a characterisation of delusion, although for the latter there is core characteristics (DSM-5, 2013). Bortolotti's analysis rely heavily on rational dimension in analysis of delusions. This dimension however is not considered as the focal by all the scholars. For example, Currie and Ravenscroft (2002) stress out that the content of belief/delusion is crucial in evaluation and on that basis argue that delusions are cognitive hallucinations (Bayne and Pacherie, 2005)

Thus, it might well turn out that when a doxasticist and an anti-doxasticist disagree over the extension of the predicates “is a belief” and “is a delusion” they are using the terms in question in such a way that they effectively speak past each other.

In order to avoid such a negative metaphilosophical conclusion about the status of the doxasticism debate one would have to assume that there is such a thing as an objective extension of the central terms of the debate, namely “belief” and “delusion”. If this were the case, then the question of the relation of the extension of the terms in question would have an objective answer, independently of any conceptual tinkering by philosophers.

This can be achieved by adopting the natural kinds theory. On the standard account, a natural kind term like “tiger” denotes a set of certain object in virtue of the fact that all members of this set form a natural kind (Putnam, 1975; Koslicki, 2008). The facts that our conception of tigers might be different from the objective facts about the actual tigers, and that people’s beliefs about what count as a tiger differ have nothing to do with the actual extension of the term. This is because “meanings just ain't in the head”: concepts and explicit definitions do not determine extensions (Wikforss, 2008).

If “belief” and “delusion” turned out to be natural kind terms denoting natural kinds, then indeed there would be an objective answer to the question “is every mental state denoted by the term 'delusion' also denoted by the term 'belief'?”. In such a case the fact that one theorist would conceptualize beliefs differently than another would have no bearing on the truth value of the answer to that question. Conversely, should “belief” and “delusion” not function as names of genuine natural kinds then it would seem that there is no guarantee that the doxasticism debate has an objective resolution.Thus, it might be claimed that the claim that doxasticism debate has an objectively correct answer presupposes the thesis that “belief” and “delusion” denote natural kinds.

It might be easily argued that “delusion” can be treated as natural kind term, given certain account of natural kinds. Beebee and Sabbarton-Leary (2010) claim that in order to account for the possibility of claiming that psychiatric terms denote natural kinds one must adopt the account of natural kinds developed by Boyd (Boyd, 1991). Delusion fit in this framework, as they seem to support inductive generalisations and might be thought of as being generated by homeostatic property clusters. Samuels (2009) also explicitly endorsed this account. Similar idea was put forward by Zachar (2014), who argued that mental disorders should be characterised as practical kind.

The question whether “belief” can be treated as term denoting natural kind seems far more difficult. This is due to the facts that, unlike “delusion”, “belief” has not been introduced as a scientific term but rather it functions as a part of our folk practice of talking about minds.

It seems obvious that the answer to the question whether “belief” can be treated as a natural kind term depends on the approach to belief one is willing to adopt. In what follows we shall assume that the neo-behavioristic approach, developed and defended recently by Schwitzgebel. According to him (Schwitzgebel 2013), having a belief, similarly to any other mental state, is a matter of fitting into a certain dispositional profile, i.e. exhibiting certain patterns of behaviour and introspectible qualities; those patterns should be similar enough to the paradigmatic case of the attitude in question in order to count as e.g. belief. On this conception the criteria of attitude possession are, as it were, shallow. This mean that there are no deep hidden facts about the subjects which count as arguments for or against ascribing certain attitudes. Such a conception seems to be incompatible with the idea of “belief” being a natural kind term, as the very idea of a term being a natural kind one presupposes that there are non-manifest application criteria for this term.

Thus, given Schwitzgebel’s account of folk psychology, there is no such thing as a correct answer to the question whether certain mental state is a “genuine” belief; this is consistent with his assumption that there are such things as in-between beliefs (Schwitzgebel 2013). However, the consequence of this approach seems to be that many debates in philosophical psychology, which are related to the question of the extension of “belief” are merely verbal ones.

To sum up, our conclusion is that, although “delusion” might reasonably be claimed to denote a natural kind, “belief”, as conceptualized by neo-behaviorists, is not. This makes the hypothesis that the doxasiticism debate is to a large extent verbal quite convincing.

The Recognitive Function of Explicit Expressions of Belief

This paper examines the pragmatic function of explicit first-person expressions of belief of the form 'I believe that p'. I will build on the disagreement between Arthur Collins and Robert Brandom concerning the analysis of such explicit expressions of belief. Collins provocatively argues against all constitutive accounts of belief, on the basis that any such account would license incoherent statements in which speakers attributed beliefs to themselves but explicitly took no stand as to the truth of the relevant proposition.[1] Brandom argues that this issue arises because of the pragmatic function of expressions of belief and gestures towards his later work as means of analysing this phenomenon of a speaker taking a stand on the truth of a matter and thereby becoming responsible to their interlocutors for having that belief. Thus, Brandom argues that Collins' insights only touch upon what we are doing when we express beliefs and not upon the concept of belief itself.[2] However, Collins disagrees, arguing we must distinguish between two different notions of taking a stand, that of being responsible towards others for the truth of a proposition, and that of being committed to it and thereby being liable to be right or wrong about it.[3]

I will develop this disagreement by critiquing and utilising the framework for the pragmatic analysis of the function of speech acts developed by two of Brandom's followers, Kukla and Lance, in their book *Yo and Lo*. I will argue that their analysis of the pragmatic function of speech acts which serve to recognise the speaker's own observations, which they term 'observatives', provides a model for understanding explicit first-person expressions of belief of the form 'I believe that p'. These speech acts have an agent-relative input and output which serves to recognise to a particular audience the speaker's own epistemic position. Most significantly, these speech acts are inherently voiced as from the speaker, because they serve to recognise the very thing which makes it appropriate to use them.

Kukla and Lance analyse the functions of speech acts in terms of the agent-relativity or agent-neutrality of their inputs and outputs, which are the normative statuses constitutive of entitlement to a given speech act and the normative changes (in the status of the speaker, or of others in the discursive community) that the act strives to produce, respectively.[4] Thus a baptismal speech act which requires a particular authority in order to be entitled to it has an agent-relative input, although it has the agent-relative output of providing a public name.[5] Agent-neutrality is explicitly tied to the public nature of truth. Kukla and Lance categorise truth-claims as having agent-neutral outputs because truths are necessarily true for everyone.[6] However, they also analyse them as having agent-neutral inputs. This leads them to deny that a speech act with an agent-relative input can be a truth-claim.[7] This prevents them from developing a richer view of testimonial practice in the space of reasons.

This is because they treat general knowledge such as that Paris is the capital of France as a paradigm for the content of a statement. This obscures the fact that many statements do proceed from an agent-relative entitlement which consists in the speaker's epistemic position. Because they are committed to truth-claims having an agent-neutral input, Kukla and Lance are unable to properly accommodate truth-claims which clearly do proceed from an agent-relative entitlement, such as observation reports.

---

[1] Arthur Collins, 'Reply to Commentators' (1994) *Philosophy and Phenomenological Research* 54(4) 929-945 p. 931.
[2] Robert, Brandom, 'Expressing and Attributing Beliefs' *Philosophy and Phenomenological Research* 54(4) 905-912 p. 910.
[3] Arthur Collins, 'Reply to Commentators' (1994) *Philosophy and Phenomenological Research* 54(4) 929-945 pp. 941-942.
[4] Rebecca Kukla and Mark Lance, *'Yo' and 'Lo', The Pragmatic Topography of the Space of Reasons* (HUP, 2009) p. 15.
[5] Ibid p. 28.
[6] Ibid pp. 16-18.
[7] Ibid pp. 16-17, 54.

Kukla and Lance develop a specific category of observatives, for speech acts which express speakers' observations. Such speech acts have agent-relative input, because only an observer can count as expressing their observation, although anyone could state the fact of their observation through a declarative speech act.[8] However, they have agent-neutral output because the speaker's expression of their observation purports to license anyone to take them to have observed the relevant thing. Kukla and Lance further claim that observatives are inherently voiced as from a speaker. An observative does not just proceed from an agent-relative entitlement, like a baptism. It also serves to express the very fact which constitutes that entitlement, the recognition of observation. Kukla and Lance use the example of 'Rabbit!' said when observing a rabbit.[9]

They analyse a this as a speech act with agent-neutral output because it licenses others to believe that a rabbit is being observed. However, they deny that it is a truth claim despite it having a bearing on the facts. They offer several reasons relating to expressions such as 'Lo, a rabbit' which have no obvious propositional structure.[10] However, they also claim that observatives cannot be truth-claims because of their agent-relative inputs. This supposedly even extends to cases where the observative is expressed with the same surface grammar as a declarative. Thus, they attempt to distinguish between 'Willard is on the mat!' as an observative and 'Willard is on the mat.' as a declarative, and to maintain that only the latter and not the former is a truth-claim.[11] Without their commitment to truth-claims having agent-neutral inputs, it is possible to enrich their framework in order to capture the particular epistemic positions which entitle various testimonial acts, including such observation reports.

Moreover, this understanding of observatives provides a model for the understanding of explicit first-person expressions of belief of the form 'I believe that p'. These are clearly truth-claims, but Collins insight is that the speaker has a particular entitlement to express their own beliefs and that entitlement is expressed in the very speech act itself. Thus, like an observative under Kukla and Lance's framework, an explicit first person expression of belief has an agent-relative entitlement. Furthermore, it is inherently personally voiced as from the speaker. The speaker is entitled to express their belief because of their recognition of it and their expression of belief expresses this fact, precisely because believing essentially involves recognising one's commitment to the relevant proposition.

This understanding of expressions of belief as truth-apt yet agent-relatively and personally entitled allows us to explain the disagreement between Collins and Brandom. Collins is correct that Brandom's notion of publicly taking a stand by expressing a belief does not capture the speaker's own epistemic commitment. Rather, what Collins has highlighted is the personally voiced and recognitive aspect of a speaker's explicit expression of belief of the form 'I believe that p'. This reveals the importance of a person's recognition, implicit or explicit, in our attribution of beliefs to them. Even when we think people are wrong about their beliefs, this is based in some explanation about how their behaviour otherwise commits them to recognising themselves as acting on contrary beliefs subconsciously. Hence Collins is correct that one cannot coherently ascribe a belief to oneself whilst maintaining that one does not take a stand on the matter. The function of the speech act to give expression to the speaker's agent-relative recognition of their own commitment is thus undermined by the contradictory expression of belief. Thus, Collins' argument does go to the concept of belief and illustrates that beliefs have an essential recognitive aspect which is exemplified in their first person explicit expressions.

[8] Ibid pp. 47-49.
[9] Ibid p. 48.
[10] Ibid pp. 54-56.
[11] Ibid pp. 55-57.

How To Twin-Earth a Phenomenal Concept (and Other Meta-Semantic Shenanigans)
Abstract for PLM 4, Ruhr-Universität Bochum

There is an intuitive distinction between those terms which are *semantically neutral*, in the sense that they are not susceptible to a twin-earth case in the style of Putnam [1975], and those terms which are susceptible to such a case, which we call *semantically valent.* Indexicals, names, and natural kinds terms are not semantically neutral. 'Sibling', 'bachelor', and 'sandwich' are plausibly semantically neutral. Bealer[1996]: 134 hypothesizes "that most, if not all, of the central terms of philosophy are" semantically neutral. This claim plays a central role in his arguments about the nature of philosophy. This paper investigates the nature and extent of semantic neutrality. In particular, I argue that, contrary to widespread philosophical orthodoxy, phenomenal concepts are not semantically neutral. I investigate the ramifications of this claim for anti-physicalist arguments in the philosophy of mind.

A term is semantically neutral if it cannot be "twin-earthed" in the style of Putnam[1975]. For the study of semantic neutrality, twin-earth cases like those in Burge[1979], involving thinkers who incompletely understand their terms or concepts and/or defer to others in their social community, are not relevant. Any term can be given a Burge-style twin-earth case, similar to 'arthritis'. We must exclude those cases, while not excluding Putnam's, to achieve a meaningful notion of semantic neutrality. After some preliminaries, we settle on the following conception of twin-earthability:

A term t is twin-earthable if there is a pair of cases with the following similarities and differences:

SIMILARITIES: Both cases involve agents that…

(a) are intrinsic duplicates

(b) use orthographically and syntactically identical terms t and t'

(c) are competent users of t and t' respectively

(d) do not use t/t' deferentially

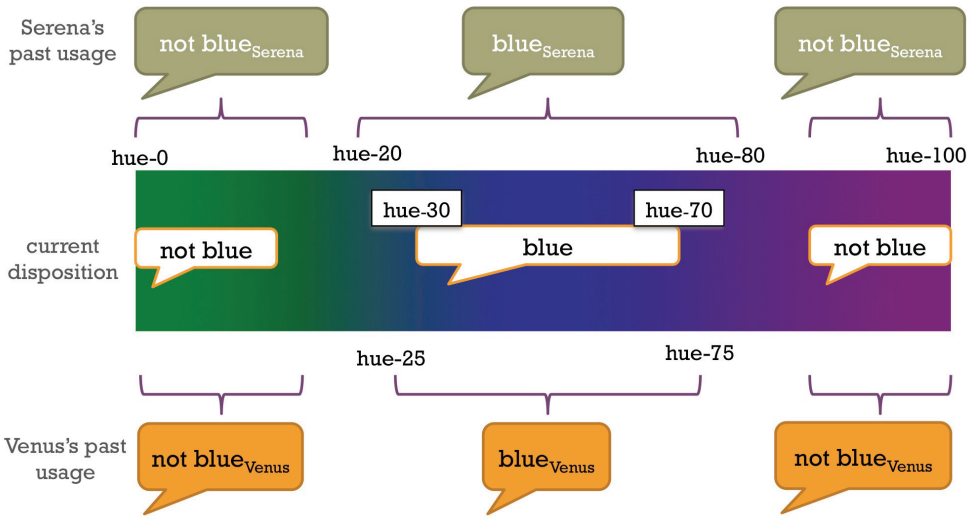
DIFFERENCES:

(e) t and t' have different intensions

We now come to the argument that phenomenal terms and concepts are not semantically neutral. A phenomenal term or concept characterizes an experience according to how that experience feels to the subject. It does not characterize the type of surface or frequency of light that tends to cause that experience. I use 'red', 'blue', and 'painful' exclusively in the phenomenal sense. I demonstrate that (phenomenal) 'blue' is not semantically neutral by providing a twin-earth case involving two intrinsic duplicates, Serena and Venus. To avoid complications involving incomplete understanding, deference, and/or social externalism, we take drastic measures. Serena and Venus were, separate from each other on their own respective planets, raised by wolves, completely separated from the rest of human society. Despite these considerable developmental obstacles, Venus and Serena grow into conceptually and linguistically sophisticated thinkers. They dabble in philosophy and lean toward dualism. They know what phenomenal concepts are and employ them.

Serena and Venus are intrinsic duplicates. They have identical dispositions to apply the phenomenal term 'blue'. We label the various hues of color experience from a clear case of green (and not blue), which we'll call hue-0, to a clear case of violet, which we'll call hue-100. Hue-50 is a clear and paradigm case of blue. Serena has a well-entrenched and long history of applying 'blue' to hues 20 through 80. She has a similar practice of predicating 'not blue' of hues 0-10 and 90-100 and abstaining from judgement for hues 10-20 and 80-90. But, at the current moment ("time t"), Serena is in a stingy mood. She is disposed to apply 'blue' only to hues 30-70. Venus, being a duplicate of Serena, has the same dispositions. But Venus has different past dispositions, and a different past practice of use, for her term 'blue'. Venus has a well-entrenched and often-exercised practice of applying 'blue' to hues 25-75. But, at time t, Venus, like Serena, is in a stingy mood, and is disposed to apply 'blue' only to hues 30-70. A diagram of Venus and Serena's past usage and current dispositions can be found below.

I claim that Venus and Serena's phenomenal term 'blue' and its corresponding phenomenal concept BLUE have different extensions (and intensions), despite the fact that Venus and Serena are intrinsic and atom-by-atom duplicates. If so, then 'blue' is not semantically neutral.



In the paper, I respond to a litany of objections against the thought experiment. The most important is that Venus and Serena, because their dispositions do not match the extension of their term, somehow fail to understand their own term or concept, and/or defer to a linguistic community partly consisting of their own former time-slices. My main reply to this objection is that it relies on an overly generous conception of "deference", according to which any disposition to accept correction of any type constitutes deference. But this can't be correct. If it were, we all defer all the time, and Putnam-style twin-earth cases are impossible. (For example, Putnam's Oscar, like Serena or Venus, would similarly accept correction were he to learn certain facts about his own past usage of 'water'. This does not entail that Putnam's thought experiment fails to demonstrate the semantic non-neutrality of 'water').

The claim that phenomenal concepts are semantically neutral plays a vital role in conceivability arguments against physicalism. As far as I am aware, that claim has never been challenged. The alleged neutrality, and thus transparency, of phenomenal concepts, provides the anti-physicalist their answer to the pressing question of why one should trust conceivability as a guide to possibility (Chalmers[1996, 2002, 2010]). I investigate various ways in which the proponent of conceivability-style arguments might respond, including resorting to other concepts or to determinate phenomenal concepts. In the end, I believe that conceivability arguments can be resurrected, but will remain significantly weakened.

Finally, I turn to more properly meta-semantic considerations. I offer an explanation for 'blue's failure of neutrality and hypothesize about the extent of semantic neutrality (or, more aptly, lack thereof). A *demonstrative labeling* occurs when a representational agent singles out some particular in the environment (often by perception) and labels it as an instance of a representational type ('blue', 'water', 'Feynman'). Demonstrative labelings function as a type of meta-semantic glue that fasten the representational type onto that which is labeled. Since the history of demonstrative labelings is not transparent to the agent, nor need it be reflected in the agent's current intrinsic state, demonstrative labelings can create semantic valency. This is what has happened with Venus and Serena. Their history of past labelings generates a difference in their current meaning, and a corresponding lack of neutrality. A hypothesis suggests itself: terms and concepts for which demonstrative labelings play an important role in reference-determination will not be semantically neutral. If so, then the realm of the semantically non-neutral extends well beyond the well-trodden cases like indexicals, names, and natural kind terms. Many non-natural kind terms, including 'mud' and 'sandwich', will not be semantically neutral either. I investigate the degree of non-neutrality that results from this phenomenon, and compare to the conventional cases involving natural kind terms. I also look at some big-picture issues in meta-semantics, and conclude by offering a picture on which the generation of meaning is a much less cognitively sophisticated affair, and meaning itself far less transparent, than has traditionally been thought.

Predication, Meaning and Transparent Hyperintensional Logic

Outline. The paper focuses on foundations of *Transparent Hyperintensional Logic* (THL) that resides in a substantial modification of Tichý's Transparent Intensional Logic (TIL; cf. Tichý 1988, 2004; Duží et al. 2010); THL has recently been suggested by Kuchyňka (yet unpublished). THL, as well as TIL, is based on Tichý's *type theory*. Similarly as *Montague's semantics* (Montague 1974 and a plenitude of other writings) and some more recent type-theoretic approaches (cf. Chatzikyriakidis and Luo 2017) THL attempts to provide a formalization of *natural language* (NL) while offering rules emulating NL 'natural logic'. (Cf. also Francez 2015 for type-theoretic approach to meaning from general perspective.)

Some general features of THL. One of the chief contributions of THL in comparison to TIL or Montague-like *possible world semantics* (PWS) is that 'intensionality' is resolved on the level of hyperintensions, not PWS-intensions. In fact, it is *predication* that is explained differently. Subsequently, the aboutness of NL expressions, i.e. what is supposed to be their *meanings*, is revised. From certain viewpoint, the proposal is not extremely revolutionary: many predictions concerning semantic behaviour of NL expressions made esp. by TIL remain. Yet THL brings even a significant notational economy, which also means that it yields a welcome simplification of the whole model. Finally, THL's different explanation of predication amounts to an accent on entities (called conditions, cf. below) which have easier computer implementation.

Meaning-relevant intuitions.. To illustrate the basic ideas of THL, consider possible explanation and then formalization of the following two sentences:

a. "Fido is a dog."

b. "Alice believes that Fido is a dog." (or: "Alice speaks about that Fido is a dog.")

There are several linguistic 'intuitions' $I_{(i)}$ pertaining to a. and b. (selection):

I_1 By affirming a./b., the speaker *reports* an observable *state of affairs* that consists in this-and-that.

I_2 The goal of affirming a./b. is to *speak about truth*.

I_3 By affirming a./b., the speaker *eliminates the possibilities which are incompatible* with the semantic content of a./b. (e.g., one excludes that Fido is a non-dog).

I_4 As regards their meaning, expressions a. and b. are *structured*.

I_5 None of a. and b. entails the other one.

How PWS implements $I_1 - I_5$. The reader surely knows that PWS successfully copes with $I_1 - I_5$ while it offers an elegant model of meaning: a. expresses a certain PWS-proposition P, i.e. the set of worlds consisting of those Ws in which Fido is a dog and eliminates those W's in which this not the case; a. speaks about P, while P is even the model of the state of affairs *that Fido is a dog*. Similarly for b., while b.'s meaning is also explained as somehow related to (though logically independent from) the meaning of a. As pointed out by many writers, the implementation of I_4, which concerns *fine-grainedness of meaning* and its (hierarchical) organization into a complex whole, is unsatisfactory in PWS. Moreover, many noticed another drawback of PWS: the meaning of a set of logically equivalent, yet non-identical expressions is the same (it is one PWS-intension); to provide a materially adequate analysis of meanings, however, one needs rather structured *hyperintensions*.

How Tichýan hyperintensional logic implements $I_1 - I_5$. Tichý's earliest proposal (from 1971 onwards) consist in capturing Montague-like predication quite differently on the formal level. But both these authors share (of course, suppressing now irrelevant differences) that a. ascribes to Fido that it is a dog, i.e. that Fido falls in the extension S (viz. a set of individuals) of the *PWS-property be a dog*. The property is modelled as an PWS-intension, namely a function from worlds Ws (and moments of time, which I will suppress) to sets such as S. In Tichý's formalism ("X_w" is an abbreviation of "$[X\ w]$"),

a.$_{TIL}$ $\lambda w[\text{Dog}_w\ F]$

Though Tichý claimed that the meaning of a. or b. is not a PWS-proposition, but its abstract structured *algorithmic computation* called by him *construction*, he does not fully embraced this hyperintensional theory of meaning until 1988. (There are various rivals of this proposal, but the nearest ones are *neo-Fregean semantics* by Muskens 2005 and Moschovakis 2006 who postulate certain algorithms computing the expressions' denotata as meanings of expressions as well.)

The analysis of b. is then:

b.$_{TIL}$ $\lambda w[\text{Bel}_w\ A\ {}^0[\lambda w[\text{Dog}_w\ F]]]$

where the (higher-order) construction ${}^0C$ yields the construction C, not its value (i.e. P). In other words, A speaks about C, not about its value P.

5. *How THL implements $I_1 - I_5$.* In notation, the artificial game with λ-abstraction over Ws and its subsequent elimination, so typical for TIL, is abandoned:

a.$_{THL}$ $[\text{Dog}\ F]_w$

b.$_{THL}$ $[\text{Bel}\ A\ {}^0[\text{Dog}\ F]_w]_w$

In THL, a proposition P (yielded e.g. by $[\text{Dog}\ F]$) is a medadic case of *condition*. *Properties* are modelled as monadic conditions, i.e. functions from objects to propositions. *Predication* is explained as *satisfying a condition*. Satisfaction of a condition is easily verifiable/observable: consider e.g the case when Fido satisfies the condition *be a dog*. According to PWS and TIL, on the other hand, one is obliged to identify the *entire* extension S of a PWS-property in W and then ask whether Fido is in S – which is hardly ever implemented, since it is often practically impossible.

As regards I_1, P is a state of affairs textslthat Fido is a dog; note that P can also be understood as the state of affairs *that Fido is a dog and Fermat's Last theorem holds*. But by affirming a. one does not intend to refer to this state of affairs P, but to the fact (if it is a fact) that P is currently true (cf. I_3). By expressing a.$_{THL}$, one eliminates $\neg[\text{Dog F}]_w$; cf. I_3. (As regards I_4, THL shares hyperintensionality/fine-grainedness of meaning with TIL; of course, I_5 is implemented in THL as well as in TIL and PWS.)

References

Chatzikyriakidis, S., Luo, Z. (eds.) (2017): *Modern Perspectives in Type-Theoretical Semantics* (Springer).

Duží, M., Jespersen, B., Materna, P. (2010): *Procedural Semantics for Hyperintensional Logic: Foundations and Applications of Transparent Intensional Logic*. Springer.

Francez, N. (2015): *Proof-theoretic Semantics*. College Publications.

Montague, R. (1974): *Formal Philosophy*. New Haven, London: Yale University Press.

Moschovakis, Y.N. (2006): A Logical Calculus of Meaning and Synonymy. *Linguistics and Philosophy* 29(1): 27–89.

Muskens, R. (2005): Sense and the Computation of Reference. *Linguistics and Philosophy* 28(4): 473–504.

Raclavský, J., Kuchyňka, P., Pezlar, I. (2015): *Transparent Intensional Logic as Characteristica Universalis and Calculus Ratiocinator* [in Czech]. Brno: Masarykova univerzita.

Tichý, P. (1988): *The Foundations of Frege's Logic*. Walter de Gruyter.

Tichý, P. (2004): *Pavel Tichý's Collected Papers in Logic and Philosophy*. V. Svoboda, B. Jespersen, C. Cheyne (eds.), Dunedin: University of Otago Publisher, Prague: Filosofia.

How to Represent the Characters of Indexicals

Indexicals are context dependent expressions; that is, their semantic value varies from context to context. This variation happens according to a rule (commonly called "character") associated with the indexical; for instance, "I" typically picks out the agent of the context, etc. Within the Kaplan tradition, characters are represented in a formal semantics by having the semantic value of the indexical delivered by a function defined on n-tuples, which in turn represent contexts, so that each place in the n-tuple is associated with a particular indexical expression. In this paper, I argue that the representation of the characters of indexicals requires semantics to place certain restrictions on n-tuples, lest we introduce arbitrary assignments in the wrong place. The more general claim is that there are certain features of indexicals that must be represented in the semantics of, and thus in the logic of, indexicals, and that these features lead to some inescapable, substantial metaphysical commitments. The semantic task of describing indexicals thus speaks against the traditional insistence on separating logic and semantics from metaphysics.

In fact, most philosophers have imposed such substantial restrictions on assignments, while postponing the worry about metaphysical commitments; I will call their theories "complicated theories". Kaplan (1989), for example, required that the speaker must exist at the space / time / possible world point.

But not everyone agrees: recently, Predelli (2013) has argued that we should allow no such restrictions; that we should allow any and all n-tuples in our semantics of indexicals. I will call such theories "simple theories".

The goal of this paper is to show that simple theories are incorrect. I argue that the semantics of indexicals requires fairly strict restrictions on n-tuples, though the reasons that Kaplan (1989) offered are not quite sufficient for this conclusion. Along the way, we also track one kind of consequence of complicated theories: they generate special, always-true sentences, variously called "logical truths" (Kaplan (1989)), or "truths in virtue of meaning" (Russell (2008)), or "truths in virtue of character alone" (Predelli (2013)).

I argue that simple theories leave out part of what they are supposed to do. The simplest way to see the problem is to look at the way a simple theory treats the first and second person singular pronouns. A simple theory represents the character of each pronoun as a function from an index to individuals. It thus distinguishes indexicals from, say, proper names, whose semantic value does not vary with the indices. But it does not, in

any interesting way, distinguish between the first person and the second person. It treats them differently, to be sure: "I" gets its value from the first element of the n-tuple, not the second. So it captures the fact that the two pronouns are different, but, it does not capture anything about the difference between them.

So how can formal semantics capture the difference between 'refers to the speaker' and 'refers to the addressee'? The simple answer: through a complicated theory. Here is the picture. Start from a context of utterance c_u (we are not assuming that all contexts are contexts of utterance, but this is the simplest case). Someone is speaking to someone else. It is not a trivial task to spell out what exactly makes one person the speaker, and another the addressee. Speaking is not sufficient for the former: when I read aloud what Frege said, I do not count as the speaker in the sense relevant here (e.g. tokens of "I" do not refer to me). Simply being in the vicinity does not make one the addressee: there may be several people around, and only one addressee. But I assume that, whatever the fundamental facts may be, the semantically relevant fact is that, in our context, A is the speaker and B the addressee. Any theory, simple or complicated, allows an index i_u which corresponds to our context, at least inasmuch as A is the first member of i_u, B the second, etc.

A complicated theory does more. It spells out the relationships among the members of indices, as they reflect the relationships between the parts of contexts. It does not assign values arbitrarily: the first member of an index is there because that index represents a context where that index member is the speaker. But this richness comes at a cost. The challenge for complicated theories is to give a theory of contexts, i.e. to specify the features of possible worlds that provide context sensitive expressions their semantic values, and to account for the relations between these features. A theory of contexts will then serve to limit indices to proper ones, i.e. those representing contexts.

References

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

Predelli, S. (2013). *Meaning without Truth*. Oxford University Press.

Russell, G. (2008). *Truth in Virtue of Meaning*. Oxford University Press.

How to talk about subjective experience?

1. INTRODUCTION

When describing our subjective conscious experiences, very often we simply describe the external objects and features that we perceive. But even on those occasions (supposing there are any) where we are concerned to speak of the subjective phenomenal character of experience *itself*, it can still be very natural to use some of the same predicates we use to characterise familiar external physical objects. E.g. A subject who is enjoying a detailed visual hallucination – and knows perfectly well that she is hallucinating – might describe some hallucinatory element in, or aspect of, her subjective visual field as 'hexagonal', 'blue', 'growing in size', 'rotating clockwise', 'to the left of' some other element etc. How literally should we understand such descriptions? Can there be phenomenal elements to an experience which are quite literally square or blue or rotating clockwise? Or are the terms 'square', 'blue', 'rotating' etc. being used here to do something other than literally predicate square-ness, blue-ness, rotation etc. There are 2 sorts of issues here, metaphysical and semantic. As a rough first pass, we can put the 2 questions as follows:

- METAPHYSICAL QUESTION: Can subjective phenomenal elements of/in experience, possess some/any of the familiar properties that external physical items can possess?
- SEMANTIC QUESTION: Do our predicates have the same meaning when we use them to describe subjective phenomenal elements of/in experience, as when we use them to describe external physical items?

As well as being (I hope) of some intrinsic interest, these questions are also crucially important to settling a further issue that has occupied a number of prominent philosophers, viz: are inter-subjective comparisons of phenomenal properties well-defined? Or is the notion of phenomenal similarity only well-defined for intra-subjective comparisons? This latter topic was taken up by Frege (1918), Wittgenstein (1953), Schlick (1948), Craig (1982, 1986, 1997), Shoemaker (1981, 1984, 1996, 2006) and Stalnaker (2003, 2006). For after all – if phenomenal elements in the experiences of two different subjects *could* both literally and correctly be described as, say, 'square', then we would plainly have a perfectly well-defined dimension of inter-subjective similarity among phenomenal features of experience.

After making some initial clarifications to our two questions [section 2], I consider how various standard positions on the metaphysics of experience will be committed to answering them "yes" or "no". I also consider to what extent there may be 'theory-neutral' or phenomenological reasons to favor answering 'yes' or 'no' and suggest that there is at least *some* defeasible phenomenological support for answering 'yes' [section 3]. I finish by considering how the questions connect with some very general issues about the nature of empirical properties/predicates and modality. I tentatively suggest that what Robert Brandom (2015) has recently labeled 'The Kant-Sellars Thesis', might provide a very general reason to answer 'no' to our questions [section 4].

2. CLARIFYING THE QUESTIONS

In clarification of the metaphysical question, we need to say more precisely *which* properties we are concerned with. This means ruling out various kinds of properties which sensations and physical objects could trivially share – e.g. being self-identical, being non-identical to Caesar etc. Also: negative properties (e.g. being non-hexagonal) and disjunctive properties (e.g. being either square or painful). To narrow things down, I will focus on spatial properties.

Of course there is a common metaphorical use of the term 'space' to speak of 3 or more dimensions of variation – e.g. colour space (hue, saturation, brightness), logical space, modal space etc. We are concerned with whether phenomenal elements/aspects of experience are *literally* spatial – not just that they have various dimensions of variation that could be represented by 3 or more axes in a diagram.

As well as metaphorical uses of spatial terms, another kind of non-literal usage is 'short-hand'. This is plausibly a common linguistic manoeuvre when describing at least some sensations – e.g. if I talk

of a 'burnt taste sensation' I am not literally ascribing the property of being burnt to the sensation. Finally, concerning the semantic question, same-ness of meaning is very plausibly something that comes in degrees – so we must be sensitive to the possibility that there might be partial overlap of meaning.

3. REASONS TO ANSWER "YES" OR "NO"

If you already have a theory about the metaphysics of experience then you may well already be committed to answering 'Yes' or 'No' to the Metaphysical Question:
Naïve-Realism/Phenomenal Externalism = YES (phenomenal properties are simply identical to external features)
Sense-Data: standardly answer = YES, but it is logical option to answer = NO
Reductive Physicalism/Identity Theory = NO (Neural structures/processes may instantiate micro-physical shapes but the wrong kinds of shapes!)
Non-Reductive/Emergentist Physicalism: it is at least a logical option to answer = YES
If you hold that experience does *not* have a genuinely 'act-object' structure – rather experience is a monadic state of the subject, as Adverbial theorists do, but also as most orthodox representational theorists also hold then you will answer = NO. (States of a subject are just the wrong logical category to instantiate spatial properties.)
Other Relevant Assumptions:
'Phenomenal Principle' (Robinson, 1994) = YES
Strong versions of the 'Transparency of Experience'= YES

A number of classic sense-data theorists argued that we should answer 'YES' on phenomenological grounds (e.g. Broad 1925, Price 1932). I suggest – following Peacocke (1983) and Mehta (2013) – that there *is* at least some defeasible phenomenological support for answering 'yes' to the metaphysical question. I also briefly rehearse a line of thought (found in Jackson 1977 and Byrne & Hilbert 2008) in favour of answering 'YES' to the semantic question, and consider possible empirical support for this line of thought from Perky's (1910) famous 'Banana Experiment'.

4. PROPERTIES & MODALITY

Following Brandom (2015) we can distinguish between *modally isolated* and *modally involved* properties/predicates:

- MODALLY INSULATED: a property, *F*, is modally isolated iff whether some particular, *o*, instantiates *F* in a possible world, *w*, depends only what is *actually* the case with o in *w*.
- MODALLY INVOLVED: a property, *F*, is modally involved iff whether some particular, *o*, instantiates *F*, in a possible world, *w*, depends on what *would be the case* with *o* in worlds other than *w*.

With this conceptual distinction in hand, we will naturally want to ask which properties, if any, are modally-insulated and which, if any, are modally involved. Brandom suggests we should endorse the following:

- "KANT-SELLARS THESIS": *Every* empirical property/predicate is modally involved. [There are *no* modally isolated empirical properties/predicates.]

How is this relevant to our metaphysical and semantic questions specifically concerning *experience*? My tentative suggestion is that *if* the 'Kant-Sellars Thesis' is correct, then private phenomenal elements/features in subjective experience could not instantiate familiar empirical properties, such as spatial properties – for they would not have the right kind of counterfactual profile. Phenomenal elements in experience would lack the right sort of connections to laws of nature and causation that, according to the 'Kant-Sellars Thesis' are *necessary conditions* for being, say, square or circular etc. So this Kant-Sellars thesis, if correct, would provide a quite *general* reason to answer 'NO' to our questions. Conversely, if one wants to hold onto the idea that inner' phenomenal features of experience are *literally* spatial – as the classic sense-datum theorists maintained – then one should probably reject the Kant-Sellars Thesis.

Attitude verbs as reference-shifting operators

Frege famously claimed that the referent of a name is shifted if it is used in the scope of an attitude verb. However, Frege had very peculiar views concerning the nature and result of this shifting process. For him the shifting was a shifting in *kind* and *type*. That is, the referent on a name is on this basis shifted from an ordinary individual to a mode of presentation of this individual.

Russell, on the other hand, claimed that the distinction between a *de dicto* and a *de re* reading of an attitude ascription that contains a name is on the level of logical form a distinction between a narrow-scope and a wide-scope use of this name with respect to the attitude verb. However, Russell's claim is based on a very controversial and problematic logical analysis of proper names. He treats names as truncated definite descriptions and definite descriptions as quantifier expressions.

In my paper, I will show that Frege and Russell made moves in the right direction concerning the semantic analysis of attitude ascriptions, but they made some unnecessarily problematic additional assumptions to validate their basic claims. I will introduce an analysis that agrees with Frege and Russell concerning the outlined basic claims, but that makes use of a very different formal semantic framework. I will make use of the tools of *model-theoretic world semantics* as it is developed in the works of Kripke (1959, 1963, 1965) and Priest (2005, 2008, 2016) to justify the mentioned two theses in a completely different way than Frege and Russell.

According to my analysis, attitude verbs are formally represented by a specific sort of modal operator. Namely, as modal operators that are relativized to subjects and that access normal and non-normal worlds. A normal world is a complete and consistent possible world. A non-normal world is either incomplete or inconsistent. Correct attitudes towards actually existing things can be represented by normal worlds. Attitudes about fictional and non-existent objects or about actual objects that are represented in an incorrect way by a subject can only be represented by means of non-normal worlds. Hence, for example, if a person thinks that Udo Jürgens and Udo Jürgen Bockelmann are two different persons, such a belief can only be represented by a non-normal world that contains two different non-actual individuals that are somehow related to the actual persons and that bear these two different names.

According to my semantic framework, names are semantically represented by individual constants. These individual constants can be without or without a denotation relative to the actual world. If they have an actual denotation, they have the same denotation relative to all *normal* worlds. If they do not have an actual denotation, they have no denotation relative to all *normal* worlds. But names behave completely anarchic with respect to non-normal worlds. Any name can have very different denotations or no denotation at all relative to different non-normal worlds. In this sense, the semantic referent of a name can be *shifted* if it is used in the scope of an attitude verb and if the modal operator that represents the semantics of this verb has access to a non-normal world, where the name has a different denotation than it has in normal worlds. But the semantic referent of a name can only then be shifted, if it is used in the scope of such a shifting operator.

On this basis, we cannot only capture the core idea behind Frege's mentioned claim, but also the core idea of Russell's way to draw the de dicto/de re distinction. A sentence like the following has two different readings

and we can capture these two readings by means of two different logical forms:

(1) Peter assumes that Mark Twain is an author.

If we use (1) with a *de dicto* reading with respect to "Mark Twain", then the logical form of (1) is the following and "Mark Twain" is used *in the scope* of a modal operator according to this logical form:

(1L1) cASS[Aa]

If we use (1), on the other hand, with a *de re* reading with respect to "Mark Twain", then the logical form of (1) is different and can be capture as follows:

(1L2) $\exists x(x=a \wedge cASS[Ax])$

Relative to the logical form (1L2) the name "Mark Twain" is used *outside the scope* of a modal operator and has his ordinary denotation. The logical forms (1L1) and (1L2) are not logically equivalent. Furthermore, neither (1L1) entails (1L2) nor the other way round.
On this basis, of our formal semantic framework which is following Priest (2005, 2016) concerning the formal representation of attitude verbs, we can show that (1L1) and (1L2) have different truth-conditions. (1L2) is only then true with respect to the actual world iff the assumption-worlds that represent the assumptions of c are such that the actual bearer of 'a' has the property expressed by 'A' according to all of them. (1L1), on the other hand, can also be true relative to the actual world if the assumption-worlds that represent the assumptions of c are only non-normal worlds. It is possible that relative to all accessible non-normal worlds 'a' refers to a different objects than the actual referent of 'a'. Whatever this object may be, (1L1) is true if this object has the property expressed by 'A' relative to those non-normal world that formally represent c's assumptions. This shows that our account has the resources to distinguish the desired different readings of attitude ascriptions like (1).

Kripke and the Formal Indeterminacy of the Physical

In the mid 1980's, Saul Kripke (1984; cf. Buechner 2011) gave a series of lectures on functionalism in which he argued that our thinking about, and reasoning in terms of, mathematical functions cannot be accounted for in terms of the way physical computing systems realize mathematical functions because such physical systems are, roughly speaking, incapable of excluding in their realization incompossible or incompatible abstract functions from being at the same time and in the same respect equally well satisfied. The reason for this, according to Kripke, is that nothing in the objective, physical causal history of the system determines which function it computes, and that any mathematical functional determinacy a physical computing system can have is the one given to it by an agreed upon interpretation, that is, by an extrinsic and observer-dependent fact. Kripke, however, resists the temptation to conclude that therefore that there's no fact of the matter as to whether we do mathematics (since he believes it is an obvious fact that we do). The conclusion to be drawn, rather, is that we are not (entirely) physical computing machines, or alternatively that our mathematical reasoning is not (entirely) a matter of physical computational processes. He likewise resisted the temptation to conclude that therefore there's something immaterial at hand, since (he claimed) an immaterial process would suffer the same fate. The conclusion to be drawn, again, was that whatever we are (whatever our mathematical reasoning consists in), we simply are not (it simply does not consist in) what the computationalist model of the mind tells us we are (tells us it consists in), at least not entirely. Kripke aims to show that functionalist computational models of the mind do not even work for those states for which it has generally been agreed to succeed, namely those "of the most intellectual kind": mathematical reasoning and understanding.

Kripke's problem for functionalism can be stated in the following way: which function a physical computing system computes is relative its idealization (i.e. to how it is idealized) and which idealization is true or correct of that physical system is, in turn, relative to an agreed upon or chosen interpretation (in the case of a computer or pocket calculator, relative to the choice and intentions of its designer(s)). In other words, neither the function a physical computing system purportedly computes nor the correct or true idealization and interpretation are part of the objective and intrinsic physical facts, features, and laws of the physical system. But if our mathematical reasoning is or consists in the physical computation or physical realization of abstract, mathematical functions, in the way physical computing machines realize such functions, then which mathematical function we compute or realize when reasoning mathematically is likewise extrinsic and relative to the choice and intentions of a designer. Our mathematical reasoning would be objectively and intrinsically indeterminate, a mere extrinsic and observer- or designer-dependent fact. This is, surely, a preposterous situation. The recommendation Kripke gives is that we abandon computational functionalist accounts of our abstract (i.e. mathematical) reasoning. And since Kripke thinks his argument can be extended, *mutatis mutandis*, to *any* functionalist account (e.g. to those who employ Ramsey sentences), the implication is clear: to abandon functionalism *in toto*.

The argument can be traced back to Kripke's notorious plus/quus considerations, but refined and excised of the skeptical paradox and newly employed in the context of assessing the plausibility of explaining mathematical reasoning in terms of the way physical computing machines and processes realize mathematical functions. Despite Kripke's intentions, however, it can be argued that his lectures ('til this day unpublished) constitute, at bottom, an argument for the *formal* indeterminacy of the physical. A fascinating but perplexing feature of Kripke's argument against functionalism is that it need not be limited to mathematical functions or structures like finite state automata nor to physical computing machines like computers or calculators, but rather it can be formulated with respect to *any* pure formal structure, whether geometrical, logical, or even grammatical (in Chomsky's sense), and *any* physical structure, like brains or force fields. The unsettling proposition with which we end up (though not entirely new to the philosophy of science) is that no physical thing or process can realize or satisfy in a determinate and objective way a unique and specific abstract function (i.e. a determinate pure formal structure), to the exclusion of any other incompossible one. There will always be a level of formal indeterminacy in anything physical.

The present talk will thus present and examine Kripke's unpublished argument in detail and will highlight that the argument's considerations are potentially generalizable to *any* sort of formal reasoning or understanding, including reasoning in accord with the rules constitutive of games like

Chess and Sudoku and the morpho-syntactic structures of languages. (Thus, for example, if the linguistic understanding and abilities of humans involves the realization of pure formal structures, functions, rules, or principles (whether in the manner of Universal Grammar or in some other way), then if Kripke's argument is sound, such understanding and abilities can be neither (1) (wholly) physical computational abilities nor (2) (wholly) the outcome or function of mere physical processes. But then, what could they possibly be?) Such startling consequences might warrant a prima facie presumption that something must be wrong with the line of reasoning, but if nothing is thus wrong, one must make way for an alternative conceptual framework if we hope to explain what it is to realize or compute abstract functions. The present talk aims to introduce this interesting but controversial argument in the hopes that some light may be shed on its defects and limitations and/or on its virtue and promising traits.

References

Buechner, J. 2011. 'Not Even Computing Machines Can Follow Rules,' in Alan Berger, ed., *Saul Kripke.* Cambridge: Cambridge University Press.

Kripke, S. 1982. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

——. 1984. "Lessons on Functionalism and Automata," *The Kirchberg-am-Wechsel Lecture*, 23 August (#KRM-$_{56}$ in The Saul Kripke Center, Kripke Archives. The Graduate Center, NY)

Chalmers on Indexicals

In 'Propositions and Attitude Ascriptions: A Fregean Account', Chalmers holds that:[1]

(T_1) Every referential expression is associated with a primary intension and a secondary intension. Primary intensions are functions from scenarios (epistemically possible worlds) to extensions. Secondary intensions are functions from possible worlds to extensions.

(T_2) When the extension of a complex expression depends compositionally on the extensions of its parts, its primary intensions depend compositionally on the primary intensions of its parts, by applying the compositionality of extensions across scenarios.

(T_3) The extension of an expression coincides with the value of its primary intension at the scenario of utterance.

(T_4) A sentence S is *a priori* (epistemically necessary) iff the primary intension of S is true at all scenarios.

This is also known as 'Chalmers' Two-Dimensional Semantics'. For now, we will say that scenarios are possible worlds,[2] and that the scenario of utterance simply is the world of utterance. Following this, we can say that the primary intension of an expression e assigns to a possible world W the extension of e if W turns out to be actual.[3] For example, if a world where Neptun is the brightest star in the evening turns out to be actual, the extension of 'Hesperus' turns out to be Neptun, and not Venus. Therefore, the primary intension of 'Hesperus' assigns to such a world Neptun, and not Venus.

In 'On Sense and Intension' Chalmers points out that in connection with indexicals his Two-Dimensional Semantics leads to a problem.[4] For example, Chalmers notes that an utterance of 'It is now Saturday' can be true today and false tomorrow. Together with (T_3), it follows that the primary intension of my utterance of 'It is now Saturday' today has to be different from the primary intension of my utterance of 'It is now Saturday' tomorrow. From this, in turn, it follows together with (T_2) that the primary intension of one of the constituents of 'It is now Saturday' chances between the two occasions of utterance. As Chalmers points out, "the obvious source of the difference is the indexical expression 'now' ".[5] However, prima facie, two arbitrary tokens of 'now' seem to have the same primary intension, since, at an arbitrary scenario W, the primary intension of a token r of 'now' always picks out the time of utterance of r. I will call this 'the problem of indexicals'.

Prima facie, the most promising way to solve the problem of indexicals within Chalmers' Two-Dimensional Semantics seems to be to advocate a token-reflexive theory according to which a token r of 'now' has the same primary intension as the description 'the time

[1] David Chalmers, 'Propositions and Attitude Ascriptions: A Fregean Account', 597, *Noûs* **45** (2011), 595–639.

[2] David Chalmers, 'On Sense and Intension', 145, *Philosophical Perspectives* **16** (2002), 135–82.

[3] Op. cit. note 2, 145.

[4] Op. cit. note 2, 155–6.

[5] Op. cit. note 2, 155.

of utterance of this token', where 'this token' refers to r. However, Chalmers[6] opts for a different solution. Chalmers rejects the claim that scenarios are objective worlds, and, instead, holds that scenarios are *centered* worlds: ordered triples of worlds, individuals and times with the individual and the time being the *center* of the world. This is motivated by the fact that an objective description of the world might leave open who I am and what time it is now. According to Chalmers,[7] these epistemic possibilities are best represented by a centered world. Moreover, such a solution has the advantage that two arbitrary tokens of 'now' have the same primary intension; the primary intension of 'now' picking out the time marked at the center of any given scenario. However, at the cost of the primary intension of an indexical sentence like 'It is now Saturday' not having an absolut truth-value anymore, but being true or false only relative to a subject and a time.

Another consequence of Chalmers' theory of indexicals is that an utterance of 'Now = now' is a priori true if true, since, at an arbitrary scenario W, both the primary intension of the first token of 'now' and the primary intension of the second token of 'now' pick out the time marked at the center of W. The same is true of an utterance of 'Here = here', since, according to Chalmers' theory, both the primary intension of the first token of 'here' and the primary intension of the second token of 'here' pick out the place marked at the center of any given scenario. However, in this talk, I will argue that there are both *a posteriori* true, i.e. epistemically contingent, utterances of 'Now = now' and a posteriori true utterances of 'Here = here'. I will call this 'the problem of a posteriori truths'. For example, I will argue that we can think of a very slow utterance of 'Here = here' such that we cannot rule out a priori that it is false. Take a speaker who produces a token of 'here' at a place p, goes for a long walk, comes back to p, and produces a token of 'is here' at p. Such a speaker certainly cannot rule out a priori that his utterance of 'Here = here' is false. This means that there is a possible world W such that his utterance of 'Here = here' turns out to be false if W turns out to be actual. Since the primary intension of an expression e assigns to a possible world W the extension of e if W turns out to be actual, according to Chalmers' Two-Dimensional Semantics, the speaker's utterance of 'Here = here' is epistemically contingent, i.e. a posteriori true.

The obvious solution to the problem of a posteriori truths within Chalmers' theory of indexicals will be to say that in case of an utterance of 'Now = now' and 'Here = here' the center of the scenario of utterance contains two times of utterance and two places of utterance (respectively). Since there are scenarios where the two places of utterance differ, the explanation goes, there are scenarios where the primary intension of the above utterance of 'Here = here' comes out false. However, this seems to be true for every true utterance of 'Here = here'. Following this, I will argue that such a solution to the problem of a posteriori truths does not provide an explanation of the fact that there are also a priori true utterances of 'Here = here' and 'Now = now', and that, therefore, Chalmers' theory of indexicals does not provide a solution to the problem of a posteriori truths.

Concluding, possible alternatives to Chalmers' theory of indexicals within Chalmers' Two-Dimensional Semantics like the token-reflexive theory are discussed. I will argue that just like Chalmers' theory of indexicals the token-reflexive theory does not provide a solution to the problem of a posteriori truths which also explains that there are a priori true utterances of 'Here = here' and 'Now = now'. The conclusion will be that an advocate of Chalmers' Two-Dimensional Semantics has to come up with a new solution to the problem of indexicals which (a) explains that at an arbitrary scenario W the primary intension of a token r of 'here' picks out the place of utterance of r, and (b) provides a solution to the problem of a posteriori truths which also explains that there are a priori true utterances of 'Here = here' and 'Now = now'. In light of the problems discussed in this talk, this will be no easy task.

[6] Op. cit. note 1, 598, and op. cit. note 2, 155.
[7] Op. cit. note 2, 156.

Naive Russellians and the Goldbach Puzzle

Following the work of Marcus, Donnellan, Perry, Kripke, and Kaplan, so-called *Neo*-Russellians like Salmon, Soames, and Braun hold that:

(*S*) The propositions we say and believe are *Russellian* propositions; i.e. structured propositions whose basic components are the objects and properties our thoughts and speech acts are about.

(DR_N) Names are directly referential terms; i.e. the content of 'n is F' in a context c is the singular proposition $\langle o, \Phi \rangle$, where o is the referent of the name n in c and Φ is the property expressed by the predicate F in c.

This is also known as the Neo-Russellian theory. Moreover, Salmon, Soames and Braun advocate the following theory of belief reports:

(*G*) A sentence of the form 'n believes that S' is true in a context c iff the referent of the name n in c believes the proposition expressed by S in c.

The theory consisting of (*S*), (DR_N) and (*G*) is sometimes referred to as 'the *Naive* Russellian theory'.

Against the Naive Russellian theory, it is often objected that we have strong intuitions that there are circumstances in which (1) is true and (2) is false.

(1) Ralph believes that Karol Wojtyła is Polish.

(2) Ralph believes that John Paul II. is Polish.

The Naive Russellian theory could not explain this. In response, Naive Russellians like Salmon, Soames, and Braun claim that our intuitions regarding the truth-values of sentences like (1) and (2) can be misleading. For example, Salmon, Soames and Braun would argue that in the above example both (1) and (2) are true. Therefore, for these philosophers, the above example neither shows that (DR_N) is false nor that (*G*) is false.

It suggests itself that within the Naive Russellian theory also (DR_G) is true.

(DR_G) The content of 'n believes that S' in a context c is the singular proposition $\langle \langle o, p \rangle$, believing$\rangle$, where o is the referent of the name n in c, and p is the proposition expressed by S in c.

However, in this talk, I will argue that together with (DR_N) and (*G*) this leads to unacceptable consequences.

The argument runs as follows. Since 'Goldbach's Conjecture' designates the (Russellian) proposition *that every even number greater than two is the sum of two primes*, according to (DR_N) and (DR_G), (3) expresses the same proposition as (4); i.e. the singular proposition $\langle \langle$Ralph, Goldbach's Conjecture$\rangle$, believing$\rangle$.

(3) Ralph believes Goldbach's Conjecture.

(4) Ralph believes that every even number greater than two is the sum of two primes.

Together with (G), it would follow that (5) is true if and only if (6) is true.

(5) Peter believes that Ralph believes Goldbach's Conjecture.

(6) Peter believes that Ralph believes that every even number greater than two is the sum of two primes.

However, as the following considerations show, there are circumstances in which (5) is true and (6) is false.

Assume that Peter is a normal English speaker who does not know that Goldbach's Conjecture says that every even number greater than two is the sum of two primes. Nevertheless, he could be disposed to sincerely and reflectively utter (3). Assume, for example, that Peter hears one of his friends sincerely and reflectively utter (3). If he considers his friend to be reliable, he could be disposed to sincerely and reflectively utter (3) without knowing what Goldbach's Conjecture says. Together with the following, simple, disquotational principle concerning belief, it would follow that (7) is true.

(DP) If a normal English speaker is disposed to sincerely and reflectively utter 'S', then he believes that S.

But since Peter does not know that Goldbach's Conjecture says that every even number greater than two is the sum of two primes, we can simply assume that (6) is false. I will call this 'the Goldbach Puzzle'.

An advocate of the Naive Russellian could object that just like our intuitions regarding the truth-value of (2) our intuitions regarding the truth-value of (6) can be misleading. I will reply that although a Naive Russellian may hold that a normal English speaker who accepts (7) believes the same thing as a normal English speaker who accepts (8), the same cannot be true for a normal English speaker who accepts (3) and a normal English speaker who accepts (4).

(7) Karol Wojtyła is Polish.

(8) John Paul II. is Polish.

A normal English speaker who accepts (4) not only believes that Ralph believes a certain proposition, but he also believes that the proposition Ralph believes has a certain content and certain truth-conditions. The same is not necessarily true for a normal English speaker who accepts (3), since a normal English speaker who accepts (3) does not necessarily know what Goldbach's Conjecture says. It will follow that the solution to the Goldbach Puzzle cannot be that (5) and (6) have Peter believing the same thing.

An advocate of the Naive Russellian theory could respond that the Goldbach Puzzle only shows that we have to reject (DR_G). I will reply that the Goldbach Puzzle already arises in connection with (S), (DR_N) and (G). For example, I will argue that according to (S) the property of believing that every even number greater than two is the sum of two primes simply is the property λx [Believe (x, Goldbach's Conjecture)]. It will follow that according to (S), (DR_N) and (G) both (5) and (6) are true if and only if Peter believes the singular proposition ⟨Ralph, the property of believing that every even number greater than two is the sum of two primes⟩. However, as we have seen above, there are circumstances in which (5) is true and (6) is false. It will follow that we have to reject (S), (DR_N) or (G).

Not all Neo-Russellians accept the Naive Russellian theory. For example, Crimmins and Perry hold that 'believe' expresses a three-place relation holding between agents, Russellian propositions and contextually determined modes of presentation. Concluding, I will show that the objection I have raised against the Naive Russellian theory is not an objection to Crimmins' and Perry's theory of belief ascriptions.

Literal meaning, idiolects, and language learning – the impossible equation

An apparent problem literalist positions in the study of communication face is how language users may come to possess idiolects which attribute to expressions concepts disparate from those of the designated literal meanings. If language-acquisition requires interaction, the fidelity with which such literal meanings are acquired seems insufficient for the purposes of advocates of literalist positions such as the minimal semantics espoused by Prof. Emma Borg (2004; 2011). To the extent that literalist theories require the idiolects of language users to universally approach an approximation of some literal dialect, the advocates of these theories would thus be forced to assume a nativist or Platonist account of people's access to word meaning. Because nativist and especially Platonist accounts posit a preordained foundation for language, they are explanatorily vacuous in the absence of evidence to back the existence of this kind of strict semantic framework.

The argument being presented starts with the premise that literalist positions require expressions being processed to be attributed their literal meaning in the externally defined *literal dialect* of some language. Thus, every individual's *idiolect*, based on which they attribute meanings to expressions, could only include attributions of meaning corresponding to those of the appropriate literal dialect(s). However, no such degree of uniformity between literal dialects and idiolects can be guaranteed if language acquisition relies on interaction between individuals. Consequently, literalist positions are only defensible if they explain access to language to be innate or objectively determined. The line of thought is as follows:

[P1] A dialect *z* of a language *l* is a set of triads M*xyz*: '*x* means *y* (in *z*)', such that

(1) each *x* is a word, or a compositionally arranged set of morphemes involving at least one word, and
(2) each *y* is a well-formed concept.

[P2] For each language *l*, there exists a designated literal dialect *z'*.

[P3] In every literalist position, if an expression *x* is present in a literal dialect *z** to which a person has access, the person attributes expression *x* meaning *y* according to M*xyz'*.

[P4] For every person, there exists an idiolect *z**, such that

(1) *z** is a dialect,
(2) if an expression *x* is present in the person's idiolect *z**, the person attributes expression *x* meaning *y* according to M*xyz**, and
(3) *z** is acquired through interaction with other speakers of a language.

[P5] If an idiolect z* is acquired through interaction with *n* speakers of a language ($n \geq 1$) with idiolects z^*_1, z^*_2, … and z^*_n, idiolect *z** is not *necessarily* a proper subset of the union of z^*_1, z^*_2, … and z^*_n.

[SC1] Thus, no person's idiolect *z** necessarily corresponds to a proper subset of the literal dialect *z* of any language.

[SC2] Thus, for no person, the meaning attributed to an expression *x* necessarily corresponds to meaning *y* derived from M*xyz'*.

[C] Thus, literalist positions can only be correct if idiolects are not acquired through interaction (i.e. if they are innate or objectively determined).

The proposed view on personal idiolects is inspired by the notion of *egocentric* processing of meaning (e.g. Giora, 2003; Barr & Keysar, 2005; Fein et al, 2015). The basic idea is that people primarily attribute to communicative acts the meaning they derive from their personal viewpoint constituted by their

context and their perception of the meaning of signs such as linguistic expressions. However, this view of idiolects is by itself compatible with literalist positions as nothing in the notions themselves precludes each individual's idiolect from corresponding with the literal dialect of some language.

The crux of the presented argument lies in the fifth premise, according to which language acquisition by interaction with other people does not guarantee that the resulting idiolect only includes a proper subset of the exact attributions of meaning present in the idiolects from which it is being derived. The claim that language acquisition by interaction is unreliable is in part founded on well-known criticisms showing the limitations of ostensive naming (Wittgenstein, 1953; Quine, 1960). The wider results of such lack of fidelity are characterised using a model of cultural evolution derived from Sperber's epidemiology of representations (Sperber, 1996; 2006). In effect, even if the prevalence of a literal dialect would provide the grounds for it being reproduced across generations of speakers, any given speaker acquiring an idiolect corresponding to the literal dialect would be contingent.

On this basis, it is argued that only comprehensive, innate conceptual systems may provide the founding for literal meaning required by literalist explanations. If shared literal meaning is to be included as a central factor in the general model of meaning attribution in communication, this literal meaning cannot have been acquired through contingently successful means of acquisition, such as language learning via interaction. While this result is considered evidence against literalist theories of meaning in communication, it does not disprove them as such. If the result holds, however, these theories would seem to be forced to assume likely ultimately indefensible positions on the grounding of linguistic meaning.

References:

Barr, D. J. and Keysar, B., 2005. Making Sense of How We Make Sense: The Paradox of Egocentrism in Language Use. In: Colston, H. L. and Albert, N. K. eds. 2005. *Figurative Language Comprehension: Social and Cultural Influences*. Mahwah, NJ: Lawrence Erlbaum Associates. Ch. 2.

Borg, E., 2004. *Minimal Semantics*. Oxford: Oxford University Press.

Borg, E., 2012. *Pursuing Meaning*. Oxford: Oxford University Press.

Fein, O., Yeari, M., and Giora, R., 2015. On the priority of salience-based interpretations: The case of sarcastic irony. *Intercultural Pragmatics*, Vol. 12(1), pp. 1–32.

Giora, R., 2003. *On Our Mind: Salience, Context, and Figurative Language*. New York: Oxford University Press.

Quine, W. V. O., 1960. *Word and Object*. New edition (2013). Cambridge, MA: MIT Press.

Sperber, D., 1996. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.

Sperber, D., 2006. Why a Deep Understanding of Cultural Evolution is Incompatible with Shallow Psychology. In: Enfield, N. J. and Levinson, S. C. eds. 2006. *Roots of Human Sociality: Culture, Cognition and Interaction*. London: Berg. Ch.16.

Wittgenstein, L., 1953. *Philosophical Investigations*. Revised 4th edition by P. M. S. Hacker and Joachim Schulte (2009). Translated from German by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Oxford: Wiley-Blackwell.

Hume's maxim with impossible worlds

1 Introduction

Priest (2005) develops a logic for intentional operators which makes use of impossible worlds in order to deal with hyperintensionality, failures of logical omniscience and the possibility to mentally represent impossible contents. Recently, he also relied on this framework to model a notion of concevability such that logical inconsistencies are conceivable (Priest 2016). This runs counter to a traditional principle, known as Hume's maxim, according to which

(HM) If it is conceivable that ϕ, then it is possible that ϕ. (SBN p. 32)

The question I would like explore in this paper is whether the admission of impossible worlds in one's modeling of intentionality is necessarily incompatible with a modal epistemology based on Hume's maxim. In opposition to Priest's view that allowing for the representation of impossibilities naturally leads to the rejection of Hume's maxim, I would like to put forward an explication of Hume's maxim that is compatible with such a setting and indeed welcomes impossible worlds in order to model the hyperintensionality of conceivings and the cognitive limitations of conceivers.

2 Priest's on Hume's maxim

Priest offers several counterexamples to (HM) - mathematical falsehoods, alternative logics, inconsistent states of affairs and fictions. But the conceivability of all the items in the list is controversial. They may appear more plausible if one accepts, as Priest does, the principle that

(1) everything describable in understandable terms is conceivable.

Priest addresses the objection that his notion of conceivability may not be the one in (HM) and challenges the objector to provide a better characterization. An appropriate answer to Priest's challenge should satisfy the following desiderata:

1. It should not turn Hume's maxim into an empty tautology.
2. It should make Hume's maxim epistemologically useful.

It is well known that the commonsense notions of imagination and conception are not very well defined and demarcated from other notions such as understanding or entertaining. But as Priest himself notices (2016, p. 2658), the principle (1) above breaks the link between conceivability and the appearance of possibility. Now it has been argued (Yablo 1993) that the appearance of possibility is a good criterion for distinguishing the senses of "conceivability" that are relevant to (HM) and the senses of "conceivability" which are not. We agree with Yablo, but note that his own positive proposal, i.e.

(Y) p is conceivable for me if I can imagine a world I take to verify p.

fails to meet Priest's challenge if we allow impossible worlds. (Either the imagined verifying world is required to be a possible world or it is not. If it is, then Hume's maxim becomes tautological. If it is not, Hume's maxim is false).

3 Conceivability revisited

A first step towards our proposal is to introduce the logic of *ceteris paribus* conceiving described by Berto (2014,2017) which offers a formalization of conceiving respectful of (a) hyperintensionality, (b) bounded rationality and (c) even the conceiving of inconsistensies. In this logic, $[\phi]\psi$ expresses the fact that when some agent explicitly conceives that ϕ, she also implicitly conceives, ceteris paribus, that ψ. The $[\cdot]$ operator behaves like a variably strict conditional, i.e. it is true in w that $[\phi]\psi$ just in case ψ is true in all the members of a set $f(\phi, w)$ of worlds (possible or impossible) accessible from w, given the antecedent ϕ, representing the worlds which are closest to the explicit conceiving of the agent.

My strategy is to constrain conceiving as formalized by Berto to keep the first too features without the third. First, I argue that we should accept a weak form of the "strangeness of impossibility" assumption (presented as plausible by Berto) according to which impossible ϕ-worlds are never closer than ϕ-possible worlds in the competition for being among the set $f(\phi, w)$. It then follows that:

(2) $\quad [\phi]\psi, \Diamond\phi \vDash \Diamond\psi$

Then, I advance the following explication of conceivability:

C $\quad \langle C\rangle\phi := \Diamond\psi \wedge [\psi]\phi$ for some ψ distinct of ϕ.

The insight behind this proposal is that we take some ϕ to be conceivable, in the relevant sense, if ϕ is *certeris paribus* conceived on the basis of some antecedently known possibility ψ, i.e $[\psi]\phi$. In such cases, ψ appears possible to us. So Yablo's constraint is satisfied.

Given the strangeness of impossibility assumption, it follows that

(3) $\quad \langle C\rangle\phi \vDash \Diamond\phi$.

This explication of conceivability respects Yablo's constraint and meets Priest's challenge: it does not turn (HM) into an empty tautology, nor is it epistemologically useless, although it requires some antecedent modal knowledge to begin with. But I argue that there is no vicious circularity here.

It also allows for hyperintensional distinctions between conceptions since, we can prove that

(4) $\quad \phi \vDash \psi \nRightarrow \langle C\rangle\phi \vDash \langle C\rangle\psi$.

and

(5) $\quad \vDash \phi \leftrightarrow \psi \nRightarrow \vDash \langle C\rangle\phi \leftrightarrow \langle C\rangle\psi$.

As a conclusion, some comparisons are made with other recent treatments of imagination (Wansing 2015) and an application of the present framework to the analysis of scientific thought experiments is sketched.

Unpleasant pain = Δ

Abstract:

I propose a new account for the unpleasantness of pain. For this, I explain, analyze, and critique a group of theories that maintain that unpleasantness is a hedonic quality. These are the hedonic theories (HTs). The main issue for all these HTs' proposals is the heterogeneity problem (HP). These HTs claim that there is one common qualitative element, one single shared ingredient, something felt, that is present among all pleasant and among all unpleasant experiences, respectively. However, according to HP there is no single felt qualitative aspect that is shared among all pains, among all unpleasant experiences, or among all pleasant experiences. I will propose and defend a new and different hedonic theory that is able to deal with HP. This is the determinable-determinate theory (DDT).

I develop three different versions of HTs. I begin by the *distinctive feelin*g view, which claims that pleasantness and unpleasantness are distinctive experiences that may accompany other hedonically neutral sensory experiences (Bramble, 2013). In the case of pain, for example, a hedonically neutral pain is accompanied by a distinctive unpleasant feeling. The combination of these two elements constitutes an unpleasant pain. The second theory that I explain is the *hedonic tone* approach. This version of HTs proposes to understand pleasantness and unpleasantness as a *qualitative dimension* along which sensory experiences may vary (Kagan, 1992). In the case of pain, for instance, unpleasantness is explained as the dimension along which a sensory pain can vary. Broadly, according to this proposal, unpleasantness is to pain as what volume is to sound. Third, I consider the *adverbial account*, for which pleasantness and unpleasantness are qualitative modifiers of other mental states such as sensory experiences (Aydede, 2014). In the case of pain, unpleasantness is a modifier of a pain sensory experience. Once the pain sensation is modified by unpleasantness, it counts as an unpleasant pain. I show that, ultimately, none of these HTs is able to provide a satisfactory explanation of the unpleasantness of pain or hedonic experiences in general. If HP is correct, as I will show it is, the three versions of HTs presented are wrong.

I propose then a new theory. It is a hedonic theory, it maintains the strong intuition that unpleasantness is something felt. However, in contrast to all the hedonic theories exposed, the theory that I propose is able to deal with HP. DDT can explain how even if unpleasantness is something phenomenal, something felt, there is no single shared qualitative ingredient that is

present among all unpleasant pains, among all unpleasant experiences, or among all pleasant experiences. There is another account that uses the determinable-determinate distinction to deal with HP (Crisp, 2006). However, this account has an important difficulty: to explain how a neutrally hedonic experience, such as a taste experience, can be either pleasant or unpleasant. In order to solve this problem, I propose a different model from Crisp's. This is what I will refer as the DDT account. According to DDT, an *unpleasant pain* is understood as a *conjunctive determinable*, i.e., a determinable that is composed by the combination of two other determinables. The unpleasant-pain determinable is the conjunction of: i) a pain sensory experience determinable and ii) an unpleasant experience determinable. According to DDT, in the same way that a scarlet equilateral triangle is a determinate of the conjunctive determinable red-triangle, an unpleasant burn experience is a determinate of the conjunctive determinable unpleasant-pain.

I consider and give an answer to two important problems for DDT. First, DDT needs to account for the variation among determinate unpleasant experiences. That is, it needs to explain what makes a determinate unpleasant pain different from another one. I think that DDT is able to capture this feature in terms of the *dimensions* along which different unpleasant pain determinates may vary. Second, it is not clear how an unpleasant pain motivates and justifies action in virtue of being a phenomenal experience, in virtue of being something felt. I will argue that even if DDT cannot offer a fully satisfying answer for this, none of the competing theories can; this is, in fact, problematic for all hedonic theories of unpleasantness. All in all, I argue that DDT is the best available account for understanding the unpleasantness of pain and hedonic experience in general.

References:

Aydede, M.. (2014). How to unify theories of sensory pleasure: an adverbialist proposal. *Review of Philosophy and Psychology*, 5:119–133

Bramble, B., (2013), "The Distinctive Feeling Theory of Pleasure". *Philosophical Studies* 162 (2):201-217

Crisp, R., (2006), *Reasons and the good*, New York: Oxford University Press.

Kagan, S., (1992), *The limits of wellbeing*, In E. F. Paul, F. D. Miller, & J. Paul (Eds.), The good life and the human good, Cambridge, UK: Cambridge University Press.

A problem with armchair arguments against explaining consciousness

The explanatory gap argument is a conceptual argument for a limit in scientific explanations: In order explain consciousness in scientific terms, we would need a functional description of it, but our concepts do not allow for such a description. Thus, no scientific explanation of consciousness can ever be given. Seen from a scientific perspective, such a claim has to be surprising. The cognitive sciences are making impressive progress at understanding the human brain. How can philosophers predict that independent of any future progress, there is something about the human mind we will never understand?
To answer this question, one might be inclined to say the following: As a conceptual argument, the explanatory gap argument is independent of empirical progress in general. Conceptual arguments are a priori, they rely on analytic truth and they are thus necessary. Empirical insights, on the contrary, are a posteriori and thus only give us contingent information. Consequently, conceptual arguments cannot be affected by any progress in the sciences.
I will argue that this is the wrong way to look at the relationship between the explanatory gap argument and empirical progress: Conceptual arguments are affected by empirical insights if they have the goal of saying some meaningful about the world. I will show that there is a way in which empirical progress can affect conceptual arguments: Through conceptual change.

In general, there are two importantly distinct ways of looking at conceptual arguments, which lead to very different views of the relationship between conceptual claims and empirical progress. On the one hand, we can look at the coherence of a conceptual argument: We can look at whether the argument presents a coherent scenario in which the premises allow to infer the conclusion. Our evaluation of the argument will then be based on the descriptive content associated with the concepts by the speaker. Seen this way, conceptual arguments are in fact independent of empirical insights. Whether Spinoza is successful at proving the existence of God, in this view, depends only on how he defined “God” and how the definition relates to his other claims and definitions. This approach explains why conceptual claims are sometimes taken to be independent of empirical progress. If we are merely interested in how a certain word is used, the nature of the actual world is irrelevant.

On the other hand, however, we can look at conceptual arguments and ask whether they describe our world accurately. Here, we are not mainly interested in the descriptive content of a person’s concept, but in whether the description the conceptual claims are based on accurately picks out the intended referent of the claim. Take again Spinoza’s argument for the existence of God. If we look at it within the accuracy approach, we are not merely interested in how he defines “God”, but also whether his definition successfully picks out what we standardly mean by “God”. If we believe his definition is problematic, we might deny the conclusion even if the argument is coherent. Within this approach, scientific insights are relevant to certain conceptual arguments, because they lead us to reach more accurate descriptions of the things we want to pick out with our concepts.

The way empirical insights affect conceptual arguments within the second approach is conceptual change. Conceptual change is a common side-product of scientific progress. New findings motivate scientists to introduce new concepts, to redefine old ones or give them up completely. In general, we use our concepts to categorize the information we gather about the world. Empirical insights are a reason to adjust this categorization in order to describe the

world more accurately. As an example, while it might at some point have seemed a truth of conceptual necessity that “whale-fish” are fish, we know by now they are in fact mammals. Thus, conceptual change can be the reason that we come to see some conceptual premise as false and thus refute the conclusion of a conceptual argument.

How does this all relate to the explanatory gap argument? What seems obvious is that when it comes to the explanatory gap, we are not merely interested in how some people use words like "consciousness". We are not interested in the coherence of the explanatory gap argument whether or not it successfully picks out consciousness. If that were the case, empirical insights would be irrelevant. However, we are interested in the phenomenon itself. Thus, if we have reasons to believe that our current concepts do not describe consciousness accurately, then new insights should motivate us to change those concepts. Such conceptual change could motivate us to change our judgments regarding the explanatory gap argument.

As an example for why we can doubt that our current concepts describe phenomenal consciousness accurately, take a look at the inverted qualia thought experiment, which has been questioned by so-called “qualia space” research. In the inverted qualia thought experiment we are imagining two people who have different phenomenal experiences, while they are identical with respect to their physical and functional properties and thus show no behavioral differences. By our naive concepts of color qualia, this thought experiment appears coherent and thus the existence of such “inverted twins” seems possible. However, this conclusion has been denied by research on qualia spaces. Researchers found out that the space of all distinguishable colors, sorted by their perceived similarities and differences, (i.e. the color qualia-space) is asymmetric. As examples, we see more different shades of green than of any other color and while most colors can be either dark or bright, all shades of yellow are seen as bright. The sorting is directly based on "what it is like" for the subjects to see the different colors. Therefore, it should be different for people who see colors differently. However, given that the resulting space is asymmetric, there is no way to invert the standard experiences while keeping all relations between experiences stable. Thus, there is no way of inverting the color experiences of our inverted twins without behavioral consequences. Our naïve concepts, which allow for the inversion, therefore cannot accurately represent the nature of phenomenal consciousness.

This shows that, despite being a conceptual argument, the explanatory gap argument should not be seen as independent of empirical progress. The inverted qualia case shows that we still lack an understanding of the relationship between experiences and certain behaviors based on those experiences, and further research on that topic should be able to influence our concepts about consciousness. In general, scientific insights have an effect on conceptual arguments when they lead us to a more accurate understanding of the referents of our concepts and thus to conceptual change.

Mentalised prototype frames

1. Introduction

The assumption that concepts have any stable meaning has been attacked from different directions: linguists, psychologists and philosophers argue in theory and empirically against it. Relevant data and arguments for this position are summarised in Casasanto, Lupyan (2015), where it is argued that all concepts are constructed ad hoc, thus meaning-stability is an illusion: "[…] despite people's intuitions, invariance is an illusion. If this is true, then a central goal of research on language and cognition should be to elucidate the fleeting, idiosyncratic neurocognitive representations that people actually use for thinking and communicating—and to explain how apparent stability emerges from pervasive variability." The evidence presented by Casasanto and Lupyan describes phenomena related to the so-called "experiential relativity" of representations between-subjects (linguistic/cultural) and within-subjects (bodily). They thus raise two related questions: a) how (in)variant and b) how idiosyncratic representations are, answering with a) completely variant, and b) completely idiosyncratic.

However, we shall argue that this is not the only possible conclusion that can be drawn from the observation of the various forms of context-dependence of representations. On the contrary, there are good reasons to conceptualize the instability suggested by the data with a different set of answers, namely a) variant, but constrained, and b) idiosyncratic, but sharable. Our argument is based on a toolkit comprising mentalising skills, evolutionary prototypes, and attribute-value structures (frames), which is summarised in part 2. In part 3 and 4, we show how this toolkit allows to argue for constrained variance and the ability of sharing representations.

2. The mentalised evolutionary prototype frame theorist's toolkit

Frames (Barsalou 1992, Petersen 2007) decompose concepts into recursive attribute-value structures. This representation format enables a flexible handling of concepts that allows for addition/deletion of attributes and values as well as for the possibility to switch foci due to their recursivity. In other words, frames allow to give an account of relations among and inside concepts and changes in their structure.

Prototype frames assign weights to the attributes and values in a representation, which reflect their importance for the concept, enabling to include typical, though neither necessary nor sufficient properties of a category. Evolutionary prototypes (Schurz 2012) interpret these weights as subjective probabilities which are based on the evolutionary history of the category to which the concept refers, leading to peaks in the property distribution for prototypical properties due to their direct or indirect contribution to fitness.

Mentalising abilities are what allows humans to attribute mental states, beliefs and intentions to each other contributing to the interpretation of behavior. The complex set of skills underlying this ability is clearly of major importance in communication, where the "effective" speaker takes into consideration what the interlocutor(s) knows and believes. Mentalising skills are formulated as functions adjusting the weights in the frame with contextually relevant information regarding how much is shared in terms of attributes and values, flexibly tuning the frame "online" during communication.

The intersubjective stability of typicality ratings has been confirmed extensively. In Barsalou, Sewell (1984) it has also been shown that adopting different points of view leads to different graded structures, which are intersubjectively stable as well.

3. The variance question

On the one hand, there is a pressure for flexibility of cognitive structures, because of the constantly changing environment, and factors like individual experience, different linguistic communities and such have therefore a role. On the other hand, a pressure for uniformity is also present, as it is necessary to take into account (1) the evolutionary pressure which led to common brain structures and similar ways to categorise the evolutionary-shaped environment, and (2) the presence of mentalising skills in humans that serve as a tool for communicative success. It is thus arguably more plausible to

assume a shared basic structure of representations, shaped by evolution, which is adjusted depending on the context at hand. Prototype frames allow to represent a stable default representation which is sufficiently flexible to account for contextual influences by deletion, addition and re-weighting of attributes and values.

4. The idiosyncrasy question

A distinction between subjective and intersubjective prototypes has the advantage to accomodate two main functions of concepts: predictive and diagnostic reasoning on the one hand, and communication on the other hand. Subjective prototypes contain the whole subjectively relevant information gathered on a category, and it is this information-richness that makes them a powerful reasoning tool. Intersubjective prototypes select only those parts of this structure of which it is assumed that are shared or shareable, making them a great tool for communication. Thus, although subjective representations are idiosyncratic, this does neither mean that they contain no stable parts, nor that people are not aware of this idiosyncrasy and able to adjust them accordingly.

5. Summary

In opposition to a view of representations as completely variant and idiosyncratic, we propose a flexible conceptual structure in terms of prototype frames, grounded in evolution and adjusted "online" by re-weighting attributes and values using mentalising skills and contextual information. First experimental confirmation for MEPs was gathered in a pilot study, asking subjects to not only list characteristic properties of concepts, but also to judge these properties with regard to the probability to which they are shared among friends, their country and the world. The intersubjective stability of these ratings is a first step towards confirming our descriptive approach to model conceptual structure.

References

Barsalou, Lawrence W. (1992): Frames, Concepts, and Conceptual Fields. In Kittay, E., Lehrer, A. (Ed.): Frames, Fields, and Contrasts: Erlbaum, pp. 21–74.

Barsalou, Lawrence W.; Sewell, Daniel R. (1984): Constructing Representations of Categories from Different Points of View. Emory University (Emory Cognition Project Tech. Rep., 2).

Casasanto, Daniel; Lupyan, Gary (2015): All Concepts are Ad Hoc Concepts. In Eric Margolis, Stephen Laurence (Eds.): The Conceptual Mind. New Directions in the Studies of Concepts: MIT Press, pp. 543–566.

Petersen, Wiebke (2007): Representation of Concepts as Frames. In J. Skilters, F. Toccafondi, G. Stemberger (Eds.): Complex Cognition and Qualitative Science, vol. 2 (The Baltic International Yearbook of Cognition, Logic and Communication, 2), pp. 151–170.

Schurz, Gerhard (2012): Prototypes and their Composition from an Evolutionary Point of View. In Wolfram Hinzen, Edouard Machery, Markus Werning (Eds.): The Oxford Handbook of Compositionality: Oxford University Press, pp. 530–554.

Fiction, lies and assertions

Abstract. I use insights from the philosophy of fiction to develop an extension of Stalnaker's framework containing unofficial common grounds to model fiction and lying. I argue that my framework makes better predictions with respect to metafictive discourse and bald-faced lies than Stokke's and Eckardt's theories.

Fiction and lying in a Stalnakerian framework. In Stalnaker's (1970) widely adopted pragmatic framework assertions are modelled as updates of the 'common ground'. When I assert "It is raining", I propose to update the common ground between me and my addressee with the proposition that "it is raining". Because Stalnaker focusses on cooperative information exchanges it can be challenging to model speech acts in a Stalnakerian framework that do not seem to follow standard Gricean maxims, such as lying and fiction-telling which both involve telling a story that you know is false. Although these speech act are traditionally treated on a par, there are crucial differences (cf. Maier, 2017) such as that stereotypical lies, contrary to fictional statements, involve an intention to deceive.

Both Eckardt's (2014) linguistic approach to modelling fiction in Stalnaker's framework and Stokke's (2013) philosophical approach to modelling lying and fiction in Stalnaker's framework, involve separate "unofficial" (Stokke, 2013) common grounds containing the content of fictional stories. Whereas lying is (like assertion) a proposal to update the official common ground, a fictional statement is a proposal to update or create an unofficial common ground related to the relevant fictional work. Take a fictional statement like (1), taken from Tolkien's novel *The Lord of the Rings*:

(1) Frodo had a very trying time that afternoon (Tolkien, *The Lord of the Rings*)

As I read (1) I update the unofficial common ground related to the *Lord of the Rings* stories with the proposition that "Frodo had a very trying time on a particular afternoon". It seems that in Eckardt's framework this results in one official common ground and several coexisting unofficial common grounds that are continuously accessible. Alternatively, Stokke analyses unofficial common grounds as essentially temporal, contrasting them with "more permanent, 'official', common grounds" (Stokke, 2013:21) to account for the intuition that fictional statements are only accepted for the purpose and duration of conversation.

I argue that these views encounter difficulties when modelling different types of fictional discourse. After reading *The Lord of the Rings*, although I no longer accept the content that I entertained while engaging with the narrative, I do not forget it. I for example remember that Frodo was adopted by his cousin and would correct someone who stated otherwise. Stokke explains this phenomenon by claiming that "an unofficial common ground need not be temporary in the sense of lasting a short time. There are arguably [unofficial] common grounds [that] continue to be operative for a very long time" (Stokke, 2014:23). But in what sense are unofficial common grounds that remain accessible after engaging with a fictional narrative (as in Eckardt's theory) temporal? It seems there are two conflicting intuitions we want to account for in modelling fictional discourse: First, the acceptance of fictional truths is temporary (only for the duration of the conversation) and second, we do retain information about the fictional content after engaging with the narrative. Eckardt's theory does not address the first intuition and Stokke runs into difficulties trying to account for both, ending up with unofficial common grounds that are both essentially temporal and continuously operative.

Metafictive discourse. The intuitive distinction made above between different types of engagement with fiction has been observed and labelled in the philosophy of fiction. For instance Currie (1990) distinguishes between fictional statements like (1) taken from Tolkien's *The Lord of the Rings* – which are analysed as direct mandates from the author to imagine certain events; and "metafictive"[1] (Currie, 1990) statements such as found in a discussion on *The Lord of the Rings* (e.g. me stating "Frodo was adopted by his cousin") – which are analysed as abbreviations of assertions about the content of the book (i.e. "In *The Lord of the Rings*, Frodo

[1] It is important to distinguish metafictive statements from so called 'meta-fictional' statements (e.g. "Frodo does not exist") which are also non-fictional assertions but not abbreviations of assertions about the fictional content.

is adopted by is cousin") and report actual beliefs. I propose to model this distinction in a Stalnakerian framework. Following Stokke and Eckardt, I analyse fictional statements as proposals to update or create a separate unofficial common ground. However, in my framework the unofficial common ground exists solely for the purpose of the fictional discourse and evaporates as soon as the conversation ends. As a result of engaging in the fictional discourse the participants update the *official* common ground by adding the total unofficial common ground to it under a Lewisian "In the worlds compatible with fiction *F*"-operator. (cf. Lewis, 1978) This is meant to model the fact that once a fictional discourse ends (e.g. once I put down *The Lord of the Rings*) participants are no longer invited to imagine or entertain any propositions but do maintain metafictive beliefs about the content of the fictional narrative (e.g. "In the worlds compatible with *The Lord of the Rings* Frodo is adopted by his cousin"). A fictional statement thus consists in a proposal to update an unofficial common ground but also in a proposal to update the official common ground with metafictive beliefs concerning the content of the fictional narrative. As we engage in different fictional narratives we update the official common ground using different fiction operators so that a typical official common ground will consist of beliefs about the actual world and metafictive beliefs about several distinct fictional worlds.

Bald-faced lies. Using a recent insight from the philosophy of fiction due to Matravers (2014) we can extend the described framework to nonfictional narratives which results in an alternative analysis of so called 'bald-faced lies'. Matravers challenges the widely adopted view that whereas nonfictional truths are to be *believed*, fictional truths are to be *imagined*. In fact, our primary engagement with fictional narratives (e.g. *The Lord of the Rings*) involves essentially the same processes as our primary engagement with nonfictional narratives (e.g. A vivid biography). In my framework we likewise adopt a uniform analysis of fiction and nonfiction. Assertions and lies are (like fictional statements) proposals to update or create an *unofficial* common ground used for the purpose and duration of the conversation. What differentiates assertion and lying from fiction is how, as a result of updating an unofficial common ground, they propose to update the *official* common ground – whether the content of the unofficial common ground will be adopted as belief (in the case of nonfiction) or as metafictive belief (in the case of fiction). Assertion and lies are thus proposals to directly update the official common ground with the content of the unofficial ground (although in the case of lies the content is known to be false by the speaker) and fictional statements are proposals to update the official common ground with the content of the unofficial common ground under the relevant fiction operator.

My framework generates an alternative analysis of bald-faced lies – statements that carry the label of 'lie' but involve no intention to deceive. For instance in *The Godfather II* Pentangeli testifies in court "I never know no Godfather" to ensure that mafioso Corleone is not convicted and Pentangeli's family remains safe. However, it is common knowledge in the courtroom that Pentangeli *did* know the Godfather. In Stokke's framework if we maintain that a lie is a proposal to update the *official* common ground, and count bald-faced lies as proper lies, we have to switch to a notion of common ground that is doxastically neutral; Pentangeli does not propose that anyone *believes* that he knows no Godfather but instead proposes that it becomes commonly *accepted* that he knows no Godfather. I argue that this is not a convincing characterization of the bald-faced lie. Rather, Pentangeli wants it to be *on the record* that he said the right things in court (especially with Corleone). His bald-faced lie is thus ultimately focused on making it common *belief* that "According to the account that Pentangeli gave in court, he knows no Godfather". As a speech act the bald-faced lie therefore has a stronger resemblance to a fictional statement than to a stereotypical lie (cf. Kreiser, 2016); The bald-faced lie is a proposal to update an unofficial common ground (in which it is true that Pentangeli knows no Godfather) and, as a result of that, update the official common ground with a 'metafictive' *belief*: "In the worlds compatible with Pentangeli's account in court, Pentangeli knows no Godfather".

Selected references: **Eckardt, R. (2014)** *The Semantics of Free Indirect Discourse*, Brill • **Currie, G. (1990)** *The Nature of Fiction*, CUP • **Kreiser, J. (2016)** 'Bald-faced Lies', *Phil Stud* • **Maier, E. (2017)** 'Lying and Fiction', *Oxford Handbook of Lying* • **Matravers, D. (2014)** *Fiction and Narrative*, OUP • **Stalnaker, R. (1970)** 'Pragmatics', *Synthese* • **Stokke, A. (2013)** 'Lying, Deceiving, and Misleading', *Phil Comp*

My Body is the Subject's Body

0. Bodily experiences[1] typically are mental states suitable to ground judgments that are *de se* in the sense that subjects endowed with a conceptual system or language would express them by qualifying the felt body with a first-person indexical. Authors in the literature often express this by saying that bodily experiences typically involve a *sense of bodily ownership* (SBO). To put it more specifically,

> [***SBO***]: For one to have a sense of bodily ownership is for one to be aware of the body one feels in bodily experience as being one's own[2].

At this point, the following question emerges: what is the specific character of this awareness? The aim of my paper is to offer an account on the SBO – namely to offer an answer to this question.

In the first part of the paper I will present the three goals that, I contend, any account on the SBO should aim at fulfilling (section **1**). I will then put forward the main lines of the view (section **2**). Finally, I will pin it down by arguing that it satisfies the goals defined in the first place (section **3**).

1. The foregoing introduction already suggests the first demand we should impose on any account on the SBO: it must allow us to explain the first-personal character of all judgments that have bodily experiences as their basis and are *de se* in the sense now relevant. This yields the following

> [***Epistemic Goal***]: any account on the SBO must explain the fact that we self-attribute the felt body for all judgments based on bodily experiences in which we do so.

On the assumption that we want a theory on the SBO to fulfill the *Epistemic Goal*, which concerns the relation between bodily experiences and proprioceptive judgments, it is also crucial that it accounts for a basic intuition regarding this relation: that of a *seemingly necessary* link between the occurrence of bodily experiences and the involvement of the first-person in the judgments by which we report them. This intuition stresses a contrast between bodily experiences and exteroception – e.g. visual perception[3]. This brings in our

> [***Psychological Goal***]: any account on the SBO must specify it in terms that, picking up a distinctive enough element of bodily experiences *vis à vis* exteroception, accounts for the seemingly necessary link that the former, but not the latter, has with the awareness of the experienced body as one's own.

The intuition just mentioned, however, is challenged by the case of somatoparaphrenic patients (Vallar and Rochi, 2009). Besides somatoparaphrenia, in fact, a vast array of empirical cases needs to be taken into account in this context. Hence our last goal is an

> [***Empirical Goal***]: any account on the SBO must leave room for the specific, sometimes abnormal relations between bodily experiences and the awareness of the felt body as one's own we seem to have evidence for in some pathological and experimental cases.

2. The view to be defended in this paper is the following: for a subject to have a SBO is for her to be aware of **(i)** the constitutive relation holding between the properties she experiences as instantiated in the felt body and the very experiences of which these properties are a content; and **(ii)** the relevant experiences (that is, bodily experiences) as being her own.

Point **(ii)** will be assumed in this paper. It is an uncontroversial fact about phenomenally conscious experiences in general, and about bodily experiences in particular, that, may the subject that undergoes them have the capacity to make judgments in which she reports them, she will typically use the pronoun "I" in the subject position. Assuming **(ii)** amounts to assuming that there is such a thing as a *sense of experience ownership* (SEO):

> [***SEO***]: For one to have a sense of experience ownership is for one to be aware of the phenomenally conscious experience one undergoes as being one's own.

[1] Bodily experiences include proprioception (namely experiences relative to bodily movement and posture), experiences related to balance, touch (i.e. experiences of bodily contact with other objects), feelings of bodily temperature, pain and interoception (e.g. hunger or thirst). I will use "proprioception" and its derivatives in a broad sense to refer to all bodily experiences.

[2] "Awareness" here is used non-factively, and in the most general sense of awareness we can assume is involved in any sincere judgment – in particular in those intended as reports of experiences.

[3] Since it seems perfectly conceivable that a subject S is in a visual state that has S's body as its content, and yet S is not aware that the body she is visually perceiving is her own.

As a result of **(i)** and **(ii)**, my view implies that the SBO is constitutively dependent upon the SEO typically present in bodily experiences.

The main goal of this part of the paper is to spell out **(i)** – that is, on the one hand, to motivate it, and on the other hand, to substantiate it. Steps **a** to **d** below summarize the motivation for a defense of **(i)**:

- **a.** Given a property P of body B, it is in principle possible for subjects to pick up P proprioceptively and exteroceptively. For instance, I can perceive my knee's Damage proprioceptively as Painfulness and exteroceptively as Redness[4].
- **b.** In contrast to Damage, Painfulness and Redness are *experience-dependent* properties: they could not be instantiated unless the experiences they are a content of were occurring. This does not tell against the fact that they are experienced as properties of B[5].
- **c.** As mentioned in section 1 above, experiences of Painfulness seem to necessarily involve awareness of B as one's own. This is not the case for experiences of Redness.
- **d.** The difference between experiences of Painfulness and experiences of Redness in this regard is explained by the fact that in the former, but not in the latter, subjects are typically aware of the *experience-dependent* nature of the property[6].

On my view, then, we are aware of the body we feel in bodily experiences as being our own in virtue of the fact that we pick it up via properties we are aware of as dependent on (our own) experiences. At this point, substantiating **(i)** essentially consists in clarifying what it means to *be aware of the experience-dependency* of properties we perceive as instantiated, and in defining what this is like in the specific case of bodily experiences.

This general clarification will be provided by the following reasoning, which draws on Peacocke (1984) and (2008)'s considerations on the visual field: in the case of sensory experiences, experience dependent-properties can be cashed out as properties of a sensory field. This being so, for a property of this sort to be *experienced as* experience-dependent means for it to be *experienced as* a property of a sensory field. Thus, being aware of the experience-dependency of perceived properties means being aware of a sensory field *as such*.

As an instance of this general reasoning, my argument for the specific case of bodily experiences points out, firstly, that the properties the body is proprioceptively experienced as having are properties of a sensory field – of the *bodily field*, as I shall call it. Secondly, this field is experienced as such in all non-pathological bodily experiences. Thus, typically, the content of bodily experiences is to be specified in terms of a sensory field we are aware of as such. I will defend that it makes sense to talk about a *bodily field* we are aware of *as such* on the basis of phenomenological considerations about the experience of haptic touch. These considerations, I will argue, can be extended to bodily experiences in general.

3. On this account, bodily properties are typically experienced in bodily experiences as constitutively dependant of *my own* experiences – namely of experiences I would report by using a first-person indexical in the subject position. When this is the case, this fact justifies the use of the first-person pronoun to qualify the felt body, thus allowing the view to meet the ***Epistemic Goal***.

Furthermore, the element that, on this view, grounds the SBO, is distinctive of the phenomenology of bodily experiences *vis à vis* experiences of external perception, such as visual experiences of color. The fact that our bodily experiences typically and distinctively involve awareness that the perceived properties are dependent on ourselves *qua* subjects allows the view to meet the ***Psychological Goal***: this is how non-pathological subjects conceive of ordinary bodily experiences.

Finally, the defended relation of dependence of the SBO on the SEO leaves room for cases such as somatoparaphrenia, in which subjects' reports indicate the lack of SBO but not the lack of SEO. Conversely, cases like depersonalization, in which the SEO and the SBO are simultaneously impaired, would seem to reinforce this relation. Thus, the view defended seems to stand in a good position when it comes to the ***Empirical Goal***.

[4] For the sake of simplicity, I shall continue to refer to experiences of Painfulness in 2-4. In the full paper, however, I shall make clear that these points are generalizable to all bodily experiences.
[5] In **a** and **b** I borrow and reformulate Dokic (2003)'s points about the transparency and reflexivity of bodily experiences.
[6] This point certainly needs more motivation than I have space to provide here. I will mainly follow Boghossian and Velleman (1989) on the differences between the experience of pain and colour for this motivation.

The Acceptability and The Probability of the Indicative Conditionals

June 19, 2017

My presentation is devoted to the probability and the acceptability of indicative conditionals. It will be organized around three influential theses, namely Stalnaker thesis:

ST $p(A \to B) = p(B|A)$

Adams thesis:

AT $ac(A \to B) = p(B|A)$[1]

and the quantitative version of Adams thesis, the so-called Qualitative Adams Thesis:

"(QAT) An indicative conditional "If A, B" is assertable for/acceptable to a person if and only if the person's degree of belief in $Pr(B|A)$ is high."(Douven and Verbrugge, 2012, p. 483)

In the first part of my presentation I will discus the results of the empirical experiments devoted to test the three hypothesis and try to show that none of them is well supported. AT was disconfirmed by Skovgaard-Olsen, Singmann, and Klauer, 2016 which showed that it fails in the case of the irrelevant conditionals and by Douven and Verbrugge, 2010 which shows that it is not supported in the case of the inductive conditionals. QAT performs poorly(Douven and Verbrugge, 2012) in comparison to the alternative theory proposed in Douven, 2008.

[1]Where "ac" indicates acceptability.

ST was suported by several empirical studies(e.g. Over, Hadjichristidis, Evans, Handley, and Sloman, 2007, Oberauer and Wilhelm, 2003). On the other hand it was shown, in Skovgaard-Olsen et al., 2016, that it fails in the cases of irrelevant conditionals. At the same time, as I will show, most of the experiments which confirmed ST do not use irrelevant conditionals and therefore fail to include a representative sample of conditionals. Such studies, at best, can support a version of the thesis restricted to the positively relevant conditionals. This seems to undermine the experiments which support ST and strongly suggests that overall it is not empirically adequate.

Then I will move to the theoretical considerations. I will discus some of the triviality proofs with a special attention dedicated to their assumptions. Then I will show that all three thesis are baseless. They are not implied by any semantic theory of conditionals.

I the last part of my talk I will discuss some alternative proposals. One of them will be Evidential Support Thesis(EST) proposed in Douven, 2008

"EST An indicative conditional "If A, B" is assertable/acceptable if and only if $Pr(B|A)$ is not only high but also higher than $Pr(B)$."

which is an alternative to QAT. Then, I will propose the alternative versions of ST and AT:

ST+/AT+ Probability/Acceptability of $A \to B$ equals:

a) $Pr(B/A)$ if $Pr(B/A) > Pr(B)$

b) 0 otherwise.

and show in what way they are superior and why they are not successful. Finally, I will try to point on more promising proposals.

2

References

Douven, I. (2008). The Evidential Support Theory of Conditionals. *Synthese*, *164*(1), 19–44.

Douven, I. & Verbrugge, S. (2010). The adams family. *Cognition*, *117*(3), 302–318.

Douven, I. & Verbrugge, S. (2012). Indicatives, concessives, and evidential support. *Thinking and reasoning*, *18*(4), 480–499.

Oberauer, K. & Wilhelm, O. (2003). The meaning (s) of conditionals: conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*, 680–693.

Over, D. E., Hadjichristidis, C., Evans, J. S. B., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, *54*(1), 62–97. doi:http://dx.doi.org/10.1016/j.cogpsych.2006.05.002

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition*, *150*, 26–36.

Does the sense of agency include a phenomenology of freedom?

The sense of agency is a catch-all term for the conscious experiences that are associated with acting. Under this term we may include such diverse phenomenological items as the sense of initiation, the sense of control, the experience of trying, the awareness of action, and many others. Philosophical research into the sense of agency is in part precipitated by studies in the neuroscience of voluntary movement which are mobilised in sceptical arguments against free will and moral responsibility. For this reason, it is strange that there is a lack of discussion in this literature on the question of whether there is a *sui generis* phenomenology of freedom that is to be included in the sense of agency. In this talk I answer that question in the positive.

The most serious objection to such a *sui generis* experience is what may be called the argument from content. It is gestured at by many authors but very rarely explicitly developed. E.g., Holton (2006, p. 2): "The real objection is to an account that invests vulgar experience with philosophical properties […]"; and Deery (2015, p. 325): "[…] how could anyone's experience have as content that her deciding to A is undetermined by her prior states together with the laws of nature?" I think that this argument is best schematised like so:

Argument C:

1. Experience has as its content particulars and low-level properties that are predicated of those particulars
2. The experience of freedom would have as part of its content modal properties.
3. Modal properties are not low-level properties
4. So we do not have an experience of freedom

In this paper I refute C by arguing against premise 3. I do so on the basis of some recent work in the philosophy of perception which includes "action-properties" amongst those properties which can be part of the content of perception. I argue for three claims: i) that action-properties are modal properties; ii) that action-properties are non-inferentially experienced (i.e., low-level); and iii) that they are of the kind that would be appropriate to an experience of freedom.

Action-properties are plausibly viewed as modal properties because they have to do with *possibilities for action* that exist for the agent. (i) should therefore be a relatively uncontroversial claim. Most of the work involves spelling out the nature of these properties in order to show the way in which they are modal. The weakest possible analysis is suggested by Nanay (2012) to be as involving non-impossibility: for some agent A, and some action x, it is not impossible for A to x (appropriately contextualised). On this weakest-possible analysis, action-properties are modal.

For (ii) I rely largely on Nanay's arguments in his (2012), but introduce some new considerations based on the comparator model of sense of agency. These arguments are based on both

phenomenological introspection as well as inference to the best explanation from empirical data. Here is one of these arguments. Imagine I am behind a sheet of plexiglass and I know I am behind plexiglass. A friend on the other side of the glass throws me a ball. I move to catch it. In order to try to catch that ball I had to represent it as catchable. But that property of catchability is in contradiction with other low-level (simple visual) properties from which it is putatively inferred (I can see the glass between me and the ball). So it could not have been inferred from those low-level properties.

There are two reasons that we may doubt that action-properties are metaphysically robust enough to be considered appropriate to an experience of freedom. Seeing off these challenges constitutes my defence of (iii). The first challenge is that we may doubt that there are truly alternative possibilities that we experience, since if we were to experience action-properties in such a simultaneous manner we would end up with a Buridan's-ass style scenario. To be more concrete, imagine that I am sitting at my desk and that I have in my conscious experience two simultaneous solicitations from action-properties in my environment: to answer my phone and to continue writing this abstract. I ignore my phone and continue working on this abstract. But if I was simultaneously solicited, then I should have been stuck between both of these possibilities for action. My solution to this problem is to introduce a role for motivation—just because we experience simultaneous solicitation this does not mean that action-properties solicit us with the same intensity.

The second reason for scepticism is that we may think that we require counterfactuals to be the content of an experience of free will, but that action-properties only yield subjunctive conditionals. Although counterfactuals are often analysed as subjunctive conditionals, this analysis is inapt: the antecedent of a counterfactual must be false, while the antecedent of a subjunctive conditional need not be. (E.g., its truth-value can be vague or undefined.) So action-properties are not robust enough to count as an experience of freedom.

Understanding the import of this objection requires us to look at the temporal character of the experience of freedom. Of course there is no way that we could have counterfactuals as the content of perception, since this would require us to experience false facts (the antecedents)—that is absurd. But I do not think it true that an experience of freedom would require us to experience counterfactuals. The experience of freedom is an experience of the *future* as being a certain way: the future appears to us as having a certain modal openness: an "unsettled" nature. I close the talk by proposing two separate metaphysical interpretations of the experience (epistemic and ontic) and argue that neither of these interpretations is obviously correct—they both raise philosophical problems.

References:

Deery, O. (2015). The fall from Eden: why libertarianism isn't justified by experience. *Australasian Journal of Philosophy*, 93, 319-334.

Holton, R. (2006). The act of choice. *Philosophers' Imprint*, 6(3).

Nanay, B. (2012). Action-oriented perception. *European Journal of Philosophy*, 20, 430-446.

Neurocognitive evidence in the semantics-pragmatics interface debate: the case of implicatures

The method of event-related brain potentials has received a wide attention from linguists since the discovery of a so-called N400 effect and its relation to semantics by (Kutas & Hillyard, 1980). They found that a meaning-related deviation is followed by a distinct event-related brain potential (ERP) component, which is negative and peaks around 400 ms post-stimulus onset. Since its discovery, the N400 has become considered an ERP signature of "semantic violation" in language even though very soon it became evident that the sensitivity of N400 is broader than the scope of what is considered "semantic" by linguists. Most recent views link the N400 to the degree of context-based expectancy/predictability of the linguistic input. Thus, the precise functional role of the N400 is still debated, especially its relevance for the theoretical and subtle questions in semantics, or those arising at the semantics-pragmatics resp. semantics-syntax interfaces. In spite of that over the last few decades we have observed a growth in interest in applying the ERP method to investigate theoretical questions in linguistics. In this paper I would like to review some of the challenges of this enterprise. I argue that (i) albeit the data from the ERP studies has to be interpreted carefully (ii) some questions can be definitely enlightened by such data. I will focus on the case of scalar implicatures which has been one of the challenges of semantics pragmatics interface. phenomena.

In the tradition stemming from Grice scalar implicatures are considered a pragmatic phenomenon (Grice, 1989). Thus, they are considered to result from a special sort of reasoning processes, where as input are taken not only literal linguistic meanings but also additional tacit principles assumed to govern any cooperative communication, in particular the *Maxim of Quantity*. Alternative views, e.g. (Chierchia et al., 2012) see scalar implicatures as part of the grammatical system, integrated in the compositional semantics. In my talk I will review the main empirical results that have been provided to enlighten the debate of the nature of scalar implicature. I put special attention to the ERP method (Noveck & Posada, 2003; Nieuwland et al., 2010; Politzer-Ahles et al., 2012; Hartshorne et al., 2015) including also original and novel ERP data from my own research on the processing of the scalar implicature of *some* as well as the scalar implicatures associated with bare numerals (*exactly* vs. *at least* interpretation of numerals such as *two*, *three*) (Spychalska et al., 2014)(Spychalska et al., 2015)(Spychalska et al., 2016). I will discuss what can be learned from the ERP results in relation to the philosophical controversies regarding SIs.

References

Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In Paul Portner, Claudia Maienborn & Klaus von Heusinger (eds.), *Semantics: An International Handbook of Natural Language Meaning*, vol. 3, 2297–2331. Mouton de Gruyter.

Grice, Herbert Paul. 1989. *Studies in the Ways of Words*. Cambridge MA: Harvard University Press.

Hartshorne, Joshua K., Jesse Snedeker, Stephanie Yen-Mun Liem Azar & Albert E. Kim. 2015. The neural computation of scalar implicature. *Language, Cognition and Neuroscience* 30(5). 620–634.

Kutas, Marta & Steven A. Hillyard. 1980. Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science* 207(4427). 203–205.

Nieuwland, Mante S., Tali Ditman & Gina R. Kuperberg. 2010. On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language* 63(3). 324–346.

Noveck, Ira A. & Andres Posada. 2003. Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language* 85(2). 203–210.

Politzer-Ahles, Stephen, Robert Fiorentino, Xiaoming Jiang & Xiaolin Zhou. 2012. Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain Research* 1490. 134–152.

Spychalska, Maria, Jarmo Kontinen, Ira A. Noveck, Ludmila Rösch & Markus Werning. 2015. Exploring the processing costs of the "exactly" and "at least" readings of bare numerals with event-related brain potentials. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings & P. P. Maglio (eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, Austin, TX: Cognitive Science Society.

Spychalska, Maria, Jarmo Kontinen & Markus Werning. 2014. Electrophysiology of Pragmatic Processing: Exploring the Processing Cost of the Scalar Implicature in the Truth-Value Judgment Task. In P. Bello, M. Guarini, M. McShane & B. Scassellati (eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, Austin, TX: Cognitive Science Society.

Spychalska, Maria, Jarmo Kontinen & Markus Werning. 2016. Investigating scalar implicatures in a truth-value judgment task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience* 31(6). 817–840.

Processing of affirmation and negation in contexts with unique and multiple alternatives: Evidence from event-related potentials

Negation is a core feature of every human language and is subject to recent debates in philosophy, psychology and linguistics. Two of the major questions focus on the mental representation of negation and negated concepts and on the integration of negation into the sentence meaning. In the psycholinguistic literature, it has been reported early on that in a truth-value judgment task negative sentences are associated with higher error rates and longer response times (Just & Carpenter, 1971, Clark & Chase, 1972) as well as longer reading times (Luedtke & Kaup, 2006) than affirmative sentences. Consequently, psycholinguistic theories seek to model negation processing in order to account for the cognitive costs related to the comprehension of negated sentences.

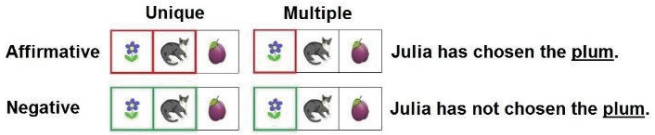
In a sentence-picture verification study, Kaup et al (2006) focused on sentences containing polar adjectives, e.g. *The door was open* and *The door was not open*, that were followed by matching or mismatching pictures after an inter-stimulus interval of either 750ms (short delay) or 1500ms (long delay). For the short delay condition subjects responded faster for affirmative sentences followed by matching pictures, whereas for negative sentences there was no effect. Yet, for the long delay the results were exactly opposite: there was no effect for the affirmative sentences, but the negative sentences were read faster when the pictures matched the described situation. Kaup et al (2006) concluded that negation processing requires two steps: At the first step the negated state of affairs is simulated (open door), whereas the actual state of affairs (not open = closed door) is simulated only in a second step.

The assumption that negation processing involves simulation processes, especially at the second stage, is faced with some difficulties: In the case of polar adjectives that have clear opposites, such as *closed* vs. *open*, the negative of the predicate (e.g. *not open*) can be simulated through its affirmative opposite (*closed*). In contrast, for predicates such as *blue*, that do not have unique opposites (e.g. *not blue* can be either *red*, or *green*, or *yellow*, etc.), it is not clear what sort of simulation is associated with the negative. In an Eye-Tracking-Study, Orenes et al (2014) investigated whether the presence of multiple alternatives in the context has an effect on the processing of negative and affirmative sentences. To this end, participants were presented with a visual context consisting of four figures of different colors. A context sentence indicated that either all objects (multary conditions) or only two objects (binary conditions) are possible choices. The target sentence made a statement about one figure, such as for example *The figure was red* or *The figure was not red*. Sentences were presented auditorily while they eye-movement to the figures on the screen were monitored. For negative sentences, subjects fixated on the target object (green figure) only in the binary condition, whereas in the multary condition they fixated on the object that had the mentioned, negated feature (red figure).

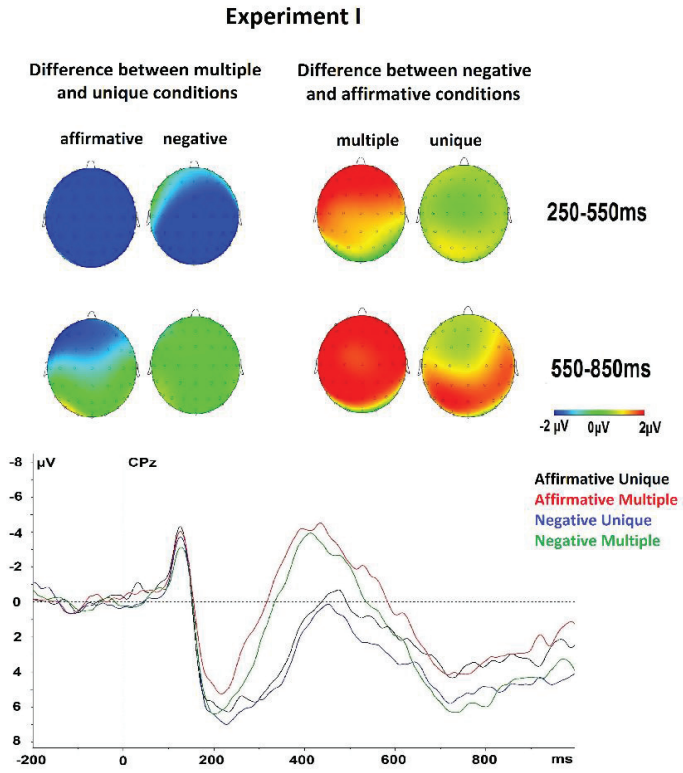
Whereas it is not surprising that in the case when the referent cannot be identified, the processing of a sentence differs relative to the case when the referent can be uniquely established in the context, the question arises whether this effect has anything specifically to do with negation. In our current experiment we aimed at directly comparing the processing of affirmative and negative sentences in situations when the context scenario provides multiple or unique potential referents. We ran our study using the method of event-related potentials, which has a high temporal resolution and therefore is frequently used for the investigation of sentence processing. The experiment had the form of a sentence-picture-verification paradigm. We used a 2 x 2 design with the factors: (i) **context model** (unique vs. multiple referent) and (ii) **polarity** of the target sentence (affirmative vs. negative). First the pictures were presented creating the context model and afterwards the target sentence was presented word-by-word on a screen while the EEG was recorded. The pictures depicted three different objects (all of same gender) out of which either one or two were then marked with a frame of red or green color. A green frame was used to indicate that an object was chosen and hence the unframed one(s) is/are unchosen by a virtual agent. A red frame was used to indicate that an object was not chosen and hence the unframed one(s) is/are chosen. Beforehand our subjects were informed that in the following they will observe a person's moves in a game where she or he will select or unselect objects. They were informed about the meaning of the different frames accordingly. The target sentence always stated which object was chosen

or unchosen and always referred to an **unframed** object.[1] Therefore, framing two out of three objects leaves only one possible and therefore **unique** referent (unframed picture) for the object named in the target sentence whereas framing only one object leaves two and hence **multiple** possible referents (unframed pictures) for the object named in the target sentence. The target sentence expressed which object was chosen (**affirmative**) or was not chosen (**negative**), e.g. *Julia hat nicht die Pflaume ausgewählt* (*Julia has not chosen the plum*).[2]

Table 1: Experimental conditions



The other two conditions accordingly used one (multiple) or two (unique) green frames and the negative version of the sentence, e.g. *Julia hat nicht die Pflaume ausgewählt* (*Julia has not chosen the plum*). All stimuli sentences gave true information with respect to the pictures. Subjects had to respond by clicking a button whether the sentence was true or false with respect to the four previously seen pictures. To balance out the material and the required responses we added false fillers.



The Picture shows the Grand Averages (N=24) at the position of the critical word for all four conditions. We observe a clear N400 effect for the multiple conditions compared to the unique ones which is independent of the sentence polarity. Assuming that the size of the N400 amplitude is a function of the context-based lexical prediction, we show that in the case of multiple predictions the size of the N400 is larger for each of the expected nouns relative to the case when the expectation is unique, thereby supporting prior observations by Spychalska et al., (2016). Thus, in the multiple case the prediction seems "weaker" for each of the predicted nouns relative to the unique case. Although it is well-established that the N400 is correlated with prediction, we show that the relative size of the N400 triggered by a noun directly depends on the number of equally predictable, scenario-based alternatives. This effect is similar for affirmative and negative sentences. Additionally, we observed a late positivity for the sentence critical nouns occurring after the negation, i.e. when the noun's expectancy was related to the prior use of a negation in the sentence. Under the assumption that the amplitude of the late positivity reflects integration of the lexical information into the semantic representation of the sentence (Brouwer & Hoeks, 2013), this effect can be taken to indicate that in the case of negated sentences the construction of the semantic representation was more effortful.

[1] In a second study we reversed the frames (meaning that the sentence always referred to a framed object) to test whether this affects the results. Everything else was kept identical. In the experiment using reversed frames we observe a very similar pattern. For brevity we only present results of the original framing in this abstract.
[2] The experiment was done in German.

References:

Brouwer, H., & Hoeks, C. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers of Human Neuroscience*, 7(758).

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. Cognitive *Psychology*, 3(3), 472-517.

Kaup, B., Lüdtke, J., Zwaan, R.A. (2006): Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed?, *Journal of Pragmatics* 38. 1033-1050

Luedtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In R. Sun (Ed.), *Proceedings of the 28th annual conference of the Cognitive Science Society* (p. 1735- 1740). Lawrence Erlbaum.

Orenes, Beltrán & Santamaría (2014): How negation is understood: evidence from the visual world paradigm. *Journal of Memory and Language*. 74. 36- 45.

Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgment task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience,* 31(6), 817840.

Uncertain Progress

This paper defends the thesis that doing something intentionally requires knowledge of doing it. The paper offers some novel reasons to accept this thesis, by drawing on considerations about control, and about a subject's knowledge of progressive truths. After this, the paper considers a type of case that has been thought to counterexample its thesis. This type of case involves someone doing something intentionally with no sensory, inferential, or testimonial way of coming to know that they are doing it. The paper argues that such cases can only seem to exemplify intentional action without knowledge to someone who tacitly accepts the following rationale:

It is possible for there to be a subject N such that

i. N is doing A intentionally.

ii. The only possible ways for N to know they are doing A is through perception, inference, or testimony, where this means knowing one is doing A in a sense not entailed by doing A.

iii. N is unable to know that they are doing A through perception, inference, or testimony, in this same sense.

The paper then presents one strategy for arguing that (i) is incompatible with (ii). The strategy is to argue that there is nothing obviously misguided about the idea that a subject knows of their intentional action merely in virtue of doing it; and that, moreover, there is little reason for saying that there are particular types of activity that can be done intentionally, but can't be known in this way.

The paper also presents a strategy for arguing that (i) is incompatible with (iii). The strategy is to argue for three claims: Doing A intentionally requires exercising control over the process called "A". But if one exercises control over A, one must at least have the capacity to gather information about the progress of A, through perception, inference, or testimony. Moreover, one must not be prevented, in the circumstances, from exercising this capacity, whether by paralysis, confusion, or something else. These claims entail that someone who conforms to (i) — someone who is doing A intentionally — can't conform to (iii).

The value of the paper lies in how it achieves two goals. First, it diagnoses an influential source of doubt about the connection between intentional action and knowledge. Second, it offers two strategies for overcoming such doubt, which are both novel, particularly in that neither rests on a tacit commitment to the thesis to be defended. The paper thereby reveals the weak foundations of the assumption that intentional actions can be unknown to their subjects.

Xxx Xxx

Modification: Pragmatics and Constraints

Prototype compositionality is a highly controversial subject. Modification of nouns is one aspect of this issue. While the selective modification model (SMM) by Smith et al. (1988) predicts that the modification doesn't affect other attributes, numerous experiments (e.g. Connolly et. al. 2007, Hampton et al. 2011, Gagné/Spalding 2014) proved that modification has an effect on the acceptance of statements: "Cotton shirts have buttons" is judged to be less likely than "Shirts have buttons". A further loss of is observed for non-typical modifications like "Itchy shirts have buttons". We propose a two-fold explanation of the effect. One part of the modification effect is explained by the pragmatics of the task and another part by subtle dependencies between the default and the modifier, which can be captured by constraints.

Constraints: An enrichment of SMM

Jönsson/Hampton (2012) observed that subjects justified their lower likelihood ratings of non-typical modifications by relations between the modifier and the default property. For example, in "Uncomfortable sofas are found in living rooms" there might be a negative relation between being uncomfortable and being placed in a living room. We use recursive attribute value structures with constraints (cf. Barsalou 1992) in order to model such dependencies and their influence on modification. In figure 1 a concept C with two attributes A1 und A2 is illustrated. The attributes have each a designated typical value (-W and F) and a non-typical one (W and -F). W and F are related by a constraint (cf. Barsalou 1992, p. 37ff), stating that a non-typical value for A1 indicates non-typicality for A2 as well. For example, birds have usually clawed feat and they mostly use flying as main locomotion. But webbed feet indicate that the bird rather moves by swimming than by flying.

According to the SSM, modification shifts all votes on the value of the modifier, but other attributes are not affected. We enrich this proposal by the constraint thesis: A modification also activates the constraints of the modifying value. As shown in figure 2 the default "flying" as main locomotion of birds is not inherited for "webbed-footed birds". It is given up in favour of another value. Note that in a probabilistic setting typical modifiers always enable default inheritance: A typical value of A1 cannot indicate an non-typical one for A2, because if P(W/C) and P(F/C) are high, P(-F/C&W) is low.

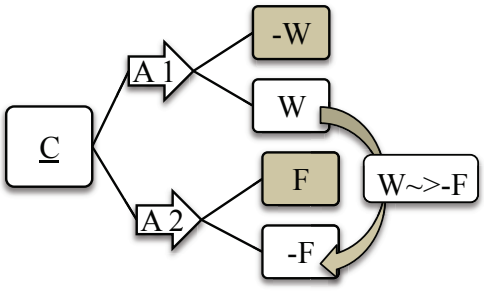


Figure 1: Frame with defaults and constraints

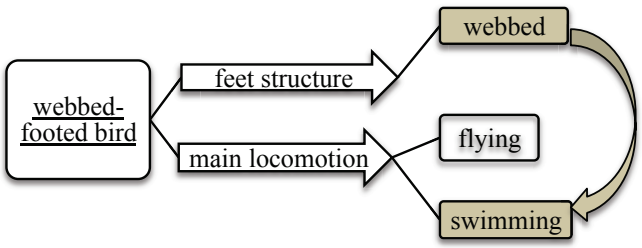


Figure 2: Example of a modification with constraint

Experiment

Material: We conducted an experiment with four different items from Connolly et al (2007). For two of them we suspected constraints to be relevant in the non-typical modification. The two other items were presumed to have no relevant constraint.

Design: Three groups of each 19 participants rated the unmodified, typically modified and non-typically modified conditions in three different ways. Group 1 rated the probability of the default property for all three conditions. All questions (together with fillers) were mingled on one questionnaire. Group 2 was presented with single case stories about individuals with or without the modifier information given between the lines. They rated the likelihood of the property for those individuals. Group 3 was explicitly asked whether they rated the three conditions differently before they answered. The rating was always given on a scale from 0 (very unlikely) to 10 (very likely)

Results: For each case we calculated the modification effects: the difference between the rating in the non-modified condition (A) and the typical condition (B) as B-A, and the difference between the rating in the non-modified condition (A) as the non-typical condition (C) as C-A, respectively. We distinguished items with a supposed constraint and items with no relevant constraints. Table 1 shows the mean values and the 0.95 confidence intervals of the modification change in the three groups. Typical and non-typical modification is represented in rows. Columns compare items without and with constraints. It was confirmed that the loss by non-typical modification is much more distinct for the items with a suspected relevant constraint. It also turned out that the single case story and the explicit question technique both moderated the possible loss by modification effect.

		Items without presumed constraint	Items with presumed constraint
Typical modification (B-A)	Group 1: Normal rating	-0.59 [-1.16, -0.03]	-0.47 [-1.07, 0.12]
	Group 2: Single case story	0.16 [-0.38, 0.70]	0.37 [-0.38, 1.11]
	Group 3: Difference rating	0.08 [-0.21, 0.37]	0.34 [-0.21, 0-89]
Non-typical modification (C-A)	Group 1: Normal rating	-0.89 [-1.60, -0.19]	-4.16 [-5.18, -3.13]
	Group 2: Single case story	-0.37 [-0.98, 0.24]	-1.16 [-1.92, -0.39]
	Group 3: Difference rating	-0.24 [-0.63, 0.16]	-2.47 [-3.33, -1.62]

Table 1: Mean change in likelihood, which was rated on a scale from 0 (very unlikely) to 10 (very likely)

Conclusion

Our experiment gave some preliminary support for our suspicions. The differences in the three groups support the assumption that pragmatic factors explain some part of the effect. Furthermore, the difference in relevant and non-relevant modification underpins our suspicion that constraints help to understand modification. New experiments with more items will be presented to further validate our model.

References

Barsalou, L. (1992). "Frames, concepts, and conceptual fields". A. Lehrer & E. Kittay (eds). *Frames, fields, and contrasts: New essays in semantic and lexical organization*, Hillsdale NJ: Lawrence Erlbaum Associates, 21-74.

Connolly, A., Fodor, J., Gleitman, L. & Gleitman, H. (2007). "Why stereotypes don't even make good defaults". *Cognition 103*, 1–22.

Gagné, C. & Spalding T. (2014). "Subcategorization, not uncertainty, drives the modification effect". *Language, Cognition, and Neuroscience, 29:10*, 1283-1294.

Hampton, J., Passanisi, A. & Jönsson, M. (2011). "The modifier effect and property mutability". *Journal of Memory and Language 64(3)*, 233-248.

Jönsson, M. & Hampton, J. (2012). "The modifier effect in within-category induction: Default inheritance in complex noun phrases". *Language and Cognitive Processes 27*, 90-116.

Smith, E., Osherson, D., Rips, L. & Keane, M. (1988). "Combining prototypes: A selective modification model", *Cognitive science 12(4)*, 485-527.

Social Externalism and Ordinary Speech

According to social externalism the contents of mental states and acts are determined not only by the intrinsic features of the subject of the state or act, but also by its social environment. In his famous argument for this view from partial understanding, Burge assumes that a sincere utterance may correctly be interpreted literally even if the subject of the utterance only partially understands one of the expressions she uses. He takes this assumption to be supported by our ordinary practice of attributing attitudes. In this talk, I take a wider look at utterances involving defective understanding and argue that Burge's appeal to our ordinary attitude attributions suffers from a diet of examples. I point to cases where ordinary speakers interpret an utterance literally even if the subject does not at least minimally understand an expression she uses. For instance, ordinary speakers interpret a subject who sincerely asks 'What is logorrhoea' as wondering what logorrhoea is even if they know that the subject does not even minimally understand 'logorrhoea'. This is surely incorrect. But if ordinary speakers do not even get such clear cases right, they are bad advisors in the more subtle cases of partial understanding. Hence, their attitude attributions are no good reason to interpret utterances involving partial understanding literally.

Furthermore, there are many cases of utterances involving partial understanding in which ordinary speakers opt *against* a literal interpretation and rather reinterpret. For instance, assume that someone mixes up the meaning of 'valid' with the meaning of 'sound', when applied to arguments. If we know of the confusion, we ordinarily reinterpret her utterances of 'valid' in terms of 'sound' rather than attribute the literally expressed attitude to her. We thereby deal with such utterances in the same way as with malapropism. And such malapropism might involve partial understanding too. For instance, we would ordinarily reinterpret Mrs Malaprops utterance 'lead the way and we'll precede' even it were no performance error but the result of confusing the meaning of 'precede' with the similar (relevant) meaning of 'proceed'. Why do ordinary speakers in some cases interpret utterances involving partial understanding literally but in other cases rather reinterpret them? One crucial difference between the cases is that in the first cases there is no alternative expression salient that correctly expresses how the subject interprets the incompletely understood expression while in the second cases there is. There are perfectly analogous cases of total lack of understanding where ordinary speakers similarly reinterpret when an apt alternative expression is salient, but falsely interpret literally when not. So in these cases, ordinary speakers misinterpret unless an apt alternative expression is salient. According to the most simple and straightforward account the same is true in cases of partial understanding. I argue that this account is far more plausible than any explanation of the data according to which the ordinary practice of interpreting utterances involving partial understanding is correct. In particular, there is no independently plausible motivation for such an explanation. Hence, from a more comprehensive perspective, ordinary attitude attributions give us in fact a reason against a literal interpretation of utterances involving partial understanding.

I will defend a reinterpretation of such utterances against some common objections and conclude with some remarks on the method of appealing to ordinary speech.

The Forms of Words

Philosophers often assume that languages are mind-external, public phenomena, perhaps depending on the existence of communities in which linguistic conventions are maintained. In contrast, linguists in the Chomskian tradition think of language as a computational capacity of an individual's mind/brain. In short, while it seems obvious that language is a cultural phenomenon, we have an empirically successful science which views language as a narrow aspect of human biology. Some see a tension here, but when we ask – in the most general terms – what language is, a full answer requires an approach which draws on the contrasting conceptions of language provided in linguistics and philosophy, and attends to language as *both* a biological *and* a cultural phenomenon. I assume a scientific realist attitude towards the posits of theoretical linguistics while insisting that public language is a genuine phenomenon of theoretical interest, one which ought to be a focus of inquiry in social ontology.

In this paper, I focus on public language words. This is partly because questions about the metaphysics and epistemology of words are a microcosm of these wider debates about language, and partly because the concept of a word is one which occurs both in our ordinary conception of language and in more theoretical contexts: teasing apart these differing conceptions and showing the relations between them, while respecting the different methodologies which ought to be employed in theoretical linguistics and in social ontology is a paradigmatically philosophical task. Finally, extant philosophical discussions of words have certain deficiencies which I would like to remedy.

According to many philosophers, words are kinds of utterances and inscriptions. Their instances are concrete, physical entities found in books, on billboards and in acoustic blasts emitted during speech. In contrast, linguists in the generativist tradition treat linguistic expressions as mental states, internal aspects of an individual's mind/brain. The contrast is not necessarily mutually exclusive: various hybrid or pluralistic views are possible. Still, we can describe these conceptions as respectively involving *externalist* and *internalist* conceptions of words. (This terminology is not intended to evoke Putnam-style content externalism: it reflects a theoretical focus on mind-external linguistic signals as opposed to mental states.)

From an externalist point of view, it makes sense to ask the following question: what makes an utterance an utterance of a given word? Assuming that words are significant units in natural languages, the question is about how to carve an aspect of linguistic reality at its joints. One answer to this question is what Kaplan (1990) called the *form-theoretic* account, FT. This approach seeks to explain sameness of word in terms of physical resemblance. It is something of a default view, sometimes tacitly assumed, sometimes explicitly defended. Adherents include Stebbing (1935), Hardie (1936), Lewis (1983:163), Davidson (1984:90), Devitt and Sterelny (1999), Cappelen (1999), Alward (2005), Predelli (2010), and, arguably, Searle (1995:83).

The work undertaken in this paper can be summarised in four parts:

(i) I present Cappelen's (1999) version of FT, and show how it is implemented within Searle's (1995) framework for social ontology, which involves his well-known XYC schema. On this view, words depend on conventions which are maintained within linguistic communities. These conventions involve collective acceptance that things of type X (things with a certain form) count as things of type Y (having linguistic properties such as being a noun, referring to cats, etc.). I argue that FT's appeal lies in a tempting but misleading conception of the processes involved in word recognition. The intuitive motivation for FT starts from the observation that competent language users are very good at working out what word types are tokened in the speech and writing to which they are exposed, and concludes that each word in the language must be associated with a distinctive acoustic (and orthographical) profile.

(ii) In the second part of this paper I make the case against FT. As well as the usual empirical arguments against FT (such as formal variation between tokens of a given word, and formal coincidence between tokens of different words), I explain the error behind the word recognition argument. Its mistake is to assume that word type recognition is entirely a matter of observing the forms of utterances and inscriptions. Against this *bottom-up* conception of speech perception – according to which word type must be detected on the basis of formal features of the linguistic signal before knowledge of syntax and semantics can be consulted – I argue that the forms of linguistic signals are highly unruly and not reliable indicators of word type; I present evidence from psycholinguistics which suggests that word recognition relies on individuals' specialised linguistic competence as well as background knowledge and guesswork. Word recognition is therefore not a discrete first step towards understanding what is said in a context, but is accomplished simultaneously with semantic interpretation. Thus, our epistemological success in word-type recognition does not require that all recognisable tokens of a word instantiate a specific form associated with that word. In addition, I raise a problem for the notion of form as it is usually employed in stating FT.

(iii) I go on to argue that the demise of FT throws up a problem for anyone hoping to use Searle's XYC schema to give a social ontological theory of words. Briefly, the XYC schema requires two ways of specifying the membership of a social kind. Cappelen assumes that the X slot can be filled in with a form-theoretic description of a certain kind of utterance, while the Y slot is filled in with syntactic and semantic properties. If, as I argue, no appropriate form-theoretic description is available, the XYC schema begins to look poorly suited to explaining the social construction of words. (Establishing this firmly would require ruling out other ways of filling in the X slot. In this paper I gesture to what these other ways might be, and express scepticism about their prospects.)

(iv) Finally, though having a certain form is neither necessary nor sufficient for being a token of a given word, a theory of words ought to say *something* about the formal features of utterances and inscriptions: Kaplan (1990) was wrong to think that form was *irrelevant* to word type. I present my preferred social ontological approach to words, drawing on Thomasson's (2014) work on public artefacts, and give an account according to which the forms of utterances are localised ways of signalling word-type without being in any sense constitutive of it.

Beyond the *What is Said* Debate: Towards a Psychologistic Theory of Successful Communication

In this paper, I begin from the dispute between semantic minimalism (SM) and radical contextualism (RC) about the reality of semantic underdetermination (Borg, 2004, 2012; Cappelen & Lepore, 2004; Carston, 2002; Recanati, 2004, 2006). I clarify this dispute by reviewing two arguments from the literature: the context-shifting and incompleteness arguments (Schoubye & Stokke, 2015). Then, I define the dispute as being about the question of whether or not *what is said* – truth-conditional content – is ever underdetermined by *linguistic meaning* – the content communicated by an utterance.

In the second part of this paper, I differentiate two distinct models of communicative interaction: the transmission model and the evidence model. My claim is that each of these models is (at least implicitly) endorsed in the literature and that each provides a different picture of the relation between truth-conditional content (*what is said*) and linguistic meaning. In the case of the transmission model, linguistic meanings are conceived as conduits that transmit truth-conditional content *simpliciter* from speaker to hearer (Cappelen & Lepore, 2004). In the case of the evidence model, linguistic meanings are conceived as coded "pieces of evidence" that a hearer uses to infer the truth-conditional content (*what is said*) expressed by a speaker (cf. Grice, 1957, 1987; Wilson & Sperber, 2012). I then claim that these models of communicative interaction satisfy widely different explanatory or descriptive purposes; e.g. to account for compositional semantics (transmission model) or to provide the basis for an account of metaphor and irony (evidence model). Furthermore, I claim that because we cannot objectively determine which model takes explanatory priority, we also cannot hope to establish a consensus about the reality – or not – of semantic underdetermination.

In part three, I provide the rationale for my constructive, "psychologistic" proposal (Crane, 2014). I begin from the claim that even if the two models of communicative interaction postulate a different relation between truth-conditional contents and linguistic meanings, both still aim to give an explanation of successful communication. Moreover, I argue that both models of communication can still agree that there must be one particular "psychological reality" that obtains between hearer and speaker when communication succeeds, and that this is true regardless of the reality – or not – of semantic underdetermination (Crane, 2013). Building on these claims, I propose that we should eschew debates about which model of communicative interaction to favour (and hence about minimalism vs. contextualism), and instead focus on developing a theory of successful communication that is consistent with both.

In part four, I set-out my "psychologistic" theory of successful communication, and I argue that this theory is consistent with both the transmission and evidence models of communicative interaction. I first assume that speakers and hearers possess semantically structured mental representations that aggregate information about objects – e.g. bananas – and can be used to make temporary "simulations" about the possible properties of those objects – e.g. bananas are yellow/purple etc. (Barsalou, 2007, 2009). Then, I argue that successful communication occurs when, as the result of communicative interaction, an isomorphism obtains between the "simulated" features of speaker's and hearer's mental representations. For example, I claim that in the case of a communicative interaction involving the utterance, "bananas are yellow," successful communication occurs if both a speaker and a hearer representationally "simulate" that bananas are yellow, and an isomorphism obtains between the features of both speaker's and hearer's "simulated" representations. Importantly, I argue that my theory of successful communication is compatible with both the transmission and evidence models of communicative interaction.

In conclusion, I argue that the debate between minimalists and contextualists about semantic underdetermination has reached an impasse, because they disagree about how best to model communicative interaction. To overcome this stalemate, I propose that we should "start from a theory of thought," whereby we seek to explain the success-conditions of communicative interaction and not the nature of communicative interaction itself (Recanati, 2010). By adopting this approach, I submit

that we can be quietist about the reality – or not – of semantic underdetermination, whilst still explaining the (psychological) state of affairs that obtains in the case of communicative success.

Bibliography

Barsalou, L. W. (2007). Grounded Cognition. *Annual Review of Psychology*, 59, 617–45.
— (2009). Simulation, Situated Conceptualization, and Prediction. *Phil. Trans. R. Soc. B*, 364, 1281–1289.
Borg, E. (2004). *Minimal Semantics*. Oxford: Oxford University Press.
— (2012). *Pursuing Meaning*. Oxford: Oxford University Press.
Cappelen, H. & Lepore, E. (2004). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Blackwell.
Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Wiley-Blackwell.
Crane, T. (2013). *The Objects of Thought*. Oxford: Oxford University Press.
— (2014). *Aspects of Psychologism*. Cambridge, MA: Harvard University Press.
Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3), 377–88.
— (1987). Logic and Conversation. In *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.
— (2006). Crazy Minimalism. *Mind and Language*, 21, 21–30.
— (2010). 'Singular Thought: In Defence of Acquaintance' in Jeshion (2010a), 141–90.
Schoubye, A. J. & Stokke, A. (2015). What is Said? *Noûs*, 49(4), 1-35.
Wilson, D. & Sperber, D. (2012). *Meaning and Relevance*. Cambridge: Cambridge University Press.

Lexical files as encyclopedic files

In this talk, I present a hierarchical interpretation of the theory of mental files (Recanati 2012; 2016) and adopt the notion of a lexical file introduced by Gasparri (2016). I defend the idea that lexical files are more than mere recognitional files identifying lexical items and recruiting the contextually relevant encyclopedic information. Lexical files must be construed as encyclopedic files, whose own stable "objects" are linguistic conventions externalistically conceived. On that basis, I argue for a distinction between lexical meaning and conventional meaning.

Mental files are handy metaphors to characterize the cognitive mechanisms, whatever they are, allowing subjects to keep track of objects and topics in thought (Schroeter 2008: 285). Mental files bind together information and attitudes which come to be treated, by the very operations of the filing system, as pertaining to the same topic. Recanati (2012; 2016) defines mental files as representational vehicles tokened in the presence of putative objects. A mental file consists of three components: the file itself (as some kind of mental label), the relations to the object the file represents, and the informational content of the file (Murez & Recanati 2016: 271). Mental files are supposed to play the two roles Frege assigned to senses: every mental file must determine (i) a referent and (ii) the mode of presentation of the referent. But mental files also differ importantly from Fregean senses. Mental files determine their referent relationally, through a specific epistemically rewarding (*ER*) relation between the subject and the object, not satisfactionally (by demanding that the object match with the file's content). The mode of presentation a file determines may thus be nondescriptive, depending on the ER relation with the object. Demonstrative and first-person files are nondescriptive, because their referent is fixed through acquaintance relations (perceptual or proprioceptive) and not through descriptions. The memory files derived from previous perceptual files (through a mechanism Recanati calls "conversion") are also nondescriptive. Arguably, there can be nondescriptive mental files for times, locations, events and situations. File types map to types of ER relation (Gasparri 2016: 467).

Crucially, the ER relations on which files are based must be understood as normative requirements, not as necessary conditions. The tokening of a mental file of a certain type comes with the presupposition that the corresponding ER relation holds with the object. But, of course, the filing system makes mistakes. Files can be opened for nonexisting objects, or files of the wrong type can be tokened, files can also fail to reidentify objects or confuse distinct objects. This accounts for all possible cases of hallucination, confusion and misidentification. So, mental files can fail to refer by failing to meet some of the requirements which come along with their tokening. They nevertheless determine a mode of presentation and allow subjects to store information about a putative topic from what they take to be a relevant informational channel. This way, mental files simultaneously account for subjects' rationality and for reference determination, while leaving room for mistakes.

An interesting idea of the mental file framework is that files are organized hierarchically according to their degree of stability and generality: first-order files are subsumed under higher-order files. Temporary perceptual files opened in the presence of an object can either vanish together with the acquaintance relation to the object or be converted into more stable memory files. Distinct memory files related to the same object form an even more stable recognitional file. Recognitional files abstract from the particular episodes of acquaintance they were built on. I might forget where, when or how many times I met somebody, and yet possess a recognitional file which allows me to identify that person and store new information about them. Multiple sources of information suitably related to one object are organized into higher-order "encyclopedic files", which are not based on specific ER relations but "may exploit a number of ER relations to the reference of the file, in an opportunistic manner, instead of being based on a single one" (Recanati 2012: 73). Encyclopedic files abstract from basic ER relations, even if they presuppose them (Recanati 2012: 75). One way to interpret this is to say that recognitial files (stable ways to recognize a given object through a particular ER relation) form a network constituting the content of an encyclopedic file. But the encyclopedic file itself does not depend on any particular one of its recognitional subfiles. All that is needed for an encyclopedic mental file is the presumption there exists some kind of relation to the object/topic the file is supposed to refer to, which is enough to activate the functional equivalent of a mental label (a mental file).

Once the mental file framework is adopted, it is natural to treat entries of the mental lexicon as mental files. Gasparri (2016) suggests that it is indeed possible to construe words of the mental lexicon ("M-words") as lexical files. Lexical files are stable recognitional files "whose function is to allow

speakers trained in a linguistic idiom to identify an auditory input as the presentation of a lexical element of that idiom, thereby tokening the conceptual-encyclopedic information required to pair it with an interpretation" (Gasparri 2016: 468). Gasparri draws on the distinction between file and file content to argue that lexical files are organized into form (phonological form) and content ("semantic core" and "combinational profile"). In context, the phonological recognition of a lexical form by the corresponding lexical file activates a set of instructions to open encyclopedic files containing the information which becomes, once pragmatically filtered, the semantic contribution of the lexical item. As Gasparri emphasizes, this view correlates with prodedural accounts of lexical semantics: semantic content is not provided *by* the lexicon but recovered *via* the lexicon.

I argue that it is possible to go further and construe lexical files as full-blown encyclopedic files, i. e., files based on a higher-order relation subsuming a variety of ER relations. First, using the notion of a hierarchy of mental files, I show that equating the lexical file itself (the mental label) with its phonological form rests on an undue conflation between the proto-lexical (recognitional) subfiles and the lexical (encyclopedic) file proper. A lexical file subsumes different recognitional subfiles tracking the different morphological realizations of the same word, and, in a written culture, the different media through which the word can be identified. Furthermore, I argue that a lexical file's complete reference includes the word's history (through phonological drift and language shift). Consequently, I construe word forms and word history as part of the rich lexical file content, not as the file itself. The file qua mental label abstracts from its various morpho-phonological properties.

In a second move, I argue that the other encyclopedic files activated by lexical files should also be construed as part of the file content, instead of a procedural semantic core. A rich network of encyclopedic information (arrangement of higher-order files) forms, together with the various abilities to track concrete physical realizations of the word, what I call the *texture* of the lexical file. The "semantic core" of the lexical file, I claim, is nothing more and nothing less than the complete semantic and expressive potential the lexical texture offers. On particular activations, it is pragmatic factors that filter the information and, if necessary, extend the encyclopedic network to deliver the context-specific semantic contributions of the lexical item. So construed, the lexical file content depends on each individual's filing system and is not reducible to a set of instructions shared by competent speakers, even if, in the normal case, a considerable overlap is to be expected between competent speakers in the way their lexical filing systems are organized and thus in the interactions between lexical and other encyclopedic files (as predicted by procedural accounts of lexical meaning).

Finally, I argue that lexical files are also based, among others, on a specific context-invariant relation to their "object", namely, a *semantic-deferential* relation to the conventional meaning the lexical item represented by the file is supposed to have. Following Millikan (1984, 2005), I conceive of linguistic conventions as natural objects, parts of the speakers' environment. Millikanian linguistic conventions are reproduced patterns owing their reproduction and evolutionary success to the performance of their "proper function": the reason they survived in the repertoire, their historical survival value. The core meaning of an expression is based on its history, and is not inferrable from averaging speakers' conceptions. Individually, speakers do not epitomize linguistic conventions but defer semantically to conventions, which are externalized. At the mental level, I argue that the default semantic-deferential attitude – which accounts for the implicit assumption of the existence of a public language – can be represented by stipulating the existence of a deferential relation as an ER relation specific to lexical encyclopedic files, whose objects are linguistic conventions themselves. In Recanati's terms (2013: 221), lexical files are all of the same "abstract indexical" type: they refer to whatever the lexical item means in the relevant communal language. I conclude that my account rests on a distinction between lexical meaning (the default structure of a lexical file) and conventional meaning (the core functions the lexical file is supposed to refer to) and supposes a commitment to the existence of communal languages.

References

Gasparri, L. (2016). Mental Files and the Lexicon. *Review of Philosophy and Psychology* 7, 2. | **Millikan**, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press. | **Millikan**, R. G. (2005). *Language: A Biological Model*. Clarendon Press. | **Murez**, M., **Recanati**, F. (2016). Mental Files: an Introduction. *Review of Philosophy and Psychology* 7, 2. | **Recanati**, F. (2012). *Mental File*s. Oxford University Press. | **Recanati**, F. (2013). Mental files: Replies to my critics. *Disputatio* 5, 36. | **Recanati**, F. (2016). *Mental Files in Flux*. Oxford University Press. | **Schroeter**, L. (2008). Are Concepts Creatures of Darkness? *Analytic Philosophy* 54, 2.

Self-Ascription and Agency of the First-Person

Debates and discussions on reflexive pronouns, the mechanisms of self-knowledge and the nature of the first-person perspective in contemporary philosophy of language and mind have led us not only to recognize the irreplaceable semantic valence of reflexive pronouns, it reignited the classical philosophical task of investigating the apriority of self-knowledge and the functional relevance of self-ascription for rational agency. Whereas this trend is often associated with 'rationalism' or 'neo-rationalism' (rightly or wrongly), there have been two major rationalist approaches propounded by Tyler Burge and Christopher Peacocke that are nevertheless opposed in significant respects. Whereas both philosophers take the agential aspects of subjects and the apriority of self-knowledge to be central in any positive account of the first person notion, Peacocke emphasizes on the non-conceptual dimension of first person content and characterizes his agency-involving account as 'metaphysics-first'. In comparison, Burge's view can be called 'epistemology-first' insofar as his take on the first person notion is more conceptual and explains the first person concept appealing to its role in rational critical reasoning. It is the aim of this essay to critically examine Peacocke's metaphysics-first view and then suggest that self-ascription is not grafted upon or derivative from a non-self-ascriptive ontology of agents (as Peacocke would have it); but rather, self-ascription should be incorporated into the core of the metaphysics of subjects as a theoretical requirement.

Burge identifies two sources of the epistemic entitlement to making a knowledgeable transition from first-order perceptual judgments to second-order self-ascriptive judgments. He says, in "Our Entitlement to Self-Knowledge":

> One is the role of the relevant judgments in critical reasoning. The other is a constitutive relation between the judgments and their subject matter - or between the judgments about one's thoughts and the judgments' being true. Understanding and making such judgments is constitutively associated both with being reasonable and with getting them right. (Burge 2013, 73)

Although these ideas are not entirely absent from Peacocke's account (for example 'the truth conduciveness thesis' of his rationalism), he offers a somewhat more complicated account of a priori self-knowledge. Whereas Burge takes self-ascription to be a central performance that is constitutive of the first person as critical reasoner, Peacocke decentralizes the notion of self-ascription within the explanatory order of his account of the first person. In his article "First Person Illusions: Are They Descartes', or Kant's?", he distinguishes an 'ownership-first view' and an 'ascription-first view' latent in Strawson's neo-Kantian account of self-consciousness. The following passage is distinctive of his critical response to what he takes to be the Kantian 'ascription-first view' of the first person:

> The fact that, for such subjects, when *s* owns *e*, it is possible for *s* to self-ascribe *e* does not imply that that fact is what makes *e* one of *s*'s experiences. For there may be some philosophical explanation of why that possibility always holds that traces the possibility to a quite different constitutive origin. That seems to me to be the case here. *e* as a conscious event is in an element of its subject's total subjective state. Thereby *e* can make rational for

> the subject various judgements, including a judgement of it, demonstratively given, to the effect 'That's mine', or 'I am having that experience of such-and-such kind'. It is entirely consistent with this non-Strawsonian, non-Kantian direction of constitutive explanation that whenever there is ownership of a conscious event, there is the possibility of self-ascription. But this 'whenever' claim is true not because ownership consists in the possibility of such self-ascription. Rather, it is ownership that grounds the possibility of the self-ascription. In short, in the order of philosophical explanation, Strawson has not established that an 'ascription-first' view of these matters is correct, rather than an 'ownership-first' view. (Peacocke 2012, 269)

Upon the ownership-first view, self-ascription is not a *constitutive* (the ascription-first view is hence also called 'constitutivism' in the literature) element of the first person notion. Self-ascription is not the genuine ground that explains *s*'s *having e*, and that means it is possibly the case that the first person remains what it is without having to self-ascribe any experience *e*. This negative account is supported by a positive account (which Peacocke later in another essay describes as 'agency-involving') of the nature of an integrating apparatus that is identical over time and possesses what Peacocke calls a 'self-file'. He gives four theses in the above-mentioned article (more detailed and systematic illustration can be found in *The Mirror of the World*). In summary, they state a fundamental reference rule that individuates the first person, a possession condition of *de se* content in general, the mental self-file as a possession condition of the first person concept, and the identity of an integrating apparatus as a metaphysical condition for the identity of conscious subject over time. His metaphysics-first account receives an agency-involving extension (I take Peacocke to be developing one single consistent theory) in his recent essay "Philosophical Reflections on the First Person, the Body, and Agency" where the reference of the concept *I* has to be the agent or the producer of a range of actions. I will argue that Peacocke's ownership-first, agency-involving account cannot explain and establish non-conceptual first person content without any appeal to self-ascription and hence self-ascription needs not and should not be excluded from the metaphysics of subjects.

The I cannot be identified as the producer of a range of actions without the determination of the possible actions being ascribed to the I. The representations of actions that are to be included in the mental self-file have to represent the actions as the subject's own actions, and as a result the non-inferential tracking of the mental file operates not without self-ascription. Peacocke's notion of an integrating apparatus that grounds the identity of conscious subject is prima facie analogous (even though not identical) to Kant's notion of the transcendental apperception. Taken in the transcendental sense, self-ascription does not necessarily proceed in an introspective manner as to generate an occurrent self-awareness, but is nevertheless an implicit thought. My critical examination of Peacocke's account is not construed as anti-metaphysical nor does it deny the possibility of non-conceptual first person content in general. Rather, it questions the division and opposition of ascription-involving and agency-involving views; they are mutually constitutive instead of mutually exclusive.

2

A Mechanistic Take on the Theory Theory of Concepts

Concepts play an essential role in the philosophy of language and mind. This paper addresses one of three main approaches that pertain to the structure and processing of *concepts*, the *theory theory*, and proposes a fresh take on it. In particular, I shall argue for the use of *mechanisms* instead – as currently discussed in the philosophy of science [7, 2, 1] – for the explication of the theory theory approach rather than a theory based (e.g., [5]) or causal graph based take (e.g., [4, 6]). This gives, as I shall argue, a fresh and fruitful perspective on the structure of concepts and reasoning processes involved. Importantly, it allows to subscribe to one of the original tenets that ordinary reasoning mimics scientific reasoning in essential ways (rather than general probabilistic reasoning), a point that seems to have been weakened in recent iterations of the theory theory, including versions that are based on causal graphs rather than theories (e.g., [4, 6]).

Let me first focus on the general status of the *theory theory* of concepts. The *theory theory* competes with two main rivals, the *classical* and the *prototype* approach to concepts [8]. The classical theory determines concepts in terms of necessary and conjointly sufficient conditions. For example, the concept BACHELOR is purported to have the necessary and jointly sufficient constituents UNMARRIED and MAN. However, there is strong empirical evidence that – abstracting from examples such as BACHELOR – the classical account does not describe reasoning with and about concepts adequately [9]. As an alternative, prototype accounts were proposed. These accounts take the processing of concepts to be guided by prototypes according to which class membership is, for example, judged based on the degree to which they resemble prototypical instances. Prototypical instances are thereby either directly described by (a) prototypical instances – exemplars – or (b) a general structure that specifies prototypes in terms of the properties these prototypical instances have to satisfy [8].

There is, however, also evidence that prototype theory cannot fully explain some empirical findings such the effect of background knowledge or the absence of general prototype effects on expert judgments [9, Ch. 2 and 6]. The theory theory [5] provides an explanation for these observations. It presupposes that naive reasoning – exhibited by children and naive adults alike – is based on the same cognitive capacities on which scientific reasoning is based [5, 8], where in the original proposal [5] scientific reasoning is purported to be done by theories (hence the name 'theory theory') – which often relies on background assumptions or background knowledge. Applied to the processing of concepts, this suggests that naive adults and children have their own *general knowledge* theory and use those theories to draw inferences about properties of individuals and their categorization in principally the same way scientists do.
[5] are explicit in that their theory theory account follows the then dominant view of scientific reasoning and change in philosophy of science. However, since then, there has been ample evidence that the theory view of scientific reasoning and change is in general inadequate. Whereas the theory approach might still have merit for fundamental physics, this is not the case for the biological sciences [1, 2, 7, 11]. Rather, abstract structures, called 'mechanisms', have been introduced that are better at describing scientific progress and reasoning in biology and seem to do that in a much more fruitful way [1, 2, 7]. By the principal motivation of the theory theory, this should make a mechanism-based approach a prime candidate for a new 'theory theory' approach to concepts.

To be fair, recent iterations of the theory theory (e.g., [4, 6]) seem to go into the same direction. For example, [4, 6] use *causal graphs* [10], which take causal structures as fundamental units. (On the other hand, [3, p. 75] seems to give up the theory theory altogether.) However, there are, as I shall argue, substantial differences between mechanisms, as employed in philosophy of biology, on the one hand and causal graphs on the other. I shall, furthermore, argue that these differences allow us to add flexibility that is instrumental for the success of mechanisms in philosophy of biology and that this flexibility is helpful for a *theory theory* approach to concepts – a point that is missing from an alternative, causal graphs based account.

Let me first go into the advantages of a mechanism-based account of a *theory theor*y of concepts and then argue why it might in a certain respect be more fruitful than a causal graphs based one. An example for a mechanism in biology is the chemical information transmission at the synaptic cleft of neurons. The entities (of the mechanisms) are (1) the axon, (2) dendrite, (3) calcium ions, (4) neuro transmitters, (5) receptors for neuro transmitters at the dendrite, etc. The activities include (A)

the release of neuro transmitter into the synaptic cleft and (B) the docking of the transmitters at the receptors as well as (C) an action potential that precedes the release of neuro transmitters and (D) which in turn is typically followed by the intake of calcium ions in the axon (cf. [2]). The resulting mechanism is in a representational format that can be reconstructed from scientific practice in biology and explain observed phenomena, including the different ways poisons, such as Curare and Strychnine, interfere with this mechanism (cf. [2, 1]).

Interestingly, it is not difficult to reconstruct naive reasoning about concepts in terms of mechanisms. Artificial kinds are designed in such a way that they afford certain properties based on causal mechanisms (e.g., microwaves heat food), whereas living natural kinds are governed by mechanisms that pertain, for example, to ways of locomotion (e.g., birds' flying ability). Whereas a layperson's understanding of these mechanisms is often limited, they can arguably identify some entities and causal activities that are relevant for the respective properties.
Furthermore, the purported similarity between some of the activities and entities might can play an important role in categorization and at least in part explain the effects of background knowledge in the processing of concepts. For example, one reason why we categories both birds' and insects' wings as WINGS is that they have a functional role in locomotion that is based on highly similar mechanisms underwriting them. Thus, although similarity might play an important part in categorization, it does so only, as I shall argue, based on the individuation of entities and the causal activities, a point that is missing from alternative accounts, including standard prototype approaches [cf. 8, 9].

Let me describe why causal graphs are at a disadvantage in that context and do not lend themselves as easily into representing mechanisms in the above sense. (None of the approaches in [1, 2, 7] relies on *causal graphs* but only on the alternative notion of *causal mechanisms*.)
Firstly, the arrows in a causal graph indicate only that there is some (direct causal) influence between specific variables, but do not specify what sort of causal influence it is (in terms of what activities exert the causal influence, e.g. if it is excitatory or inhibitory) [10]. Rather, the type of activities that govern causal influence have to be encoded in variables and values (e.g., by use of a variable 'release of neuro transmitters' and values '0 (yes)' and '1 (no)'; see also [4]). Secondly, entities cannot be directly represented but have to encoded *again* in terms of the variables (e.g., 'release of neuro transmitters *by the axon'*). Thirdly, there are further specific issues such as (1) the satisfaction of the Markov (and the faithfulness) condition that might not be always be satisfied by naive mechanisms and (2) the treatment of feedback loops due to the fact that causal graphs are naively described as circular, whereas causal graphs are defined to be directed *acyclical* graphs [10, 4], allowing for no simple mapping of such loops (which are essential for comprehensive account of mechanisms).

References

[1] Bechtel, W. and Richardson, R (2010). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, Mass: MIT Press.
[2] Craver, C. (2007). *Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press.
[3] Gopnik, A. and Bonawitz, E. (2015). Bayesian Models of Child Development. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 75–86.
[4] Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., and Danks, D. (2004). A Theory of Causal Learning in Children. Causal Maps and Bayes Nets. *Psychological Review*, *111*, 3–32.
[5] Gopnik, A. and Meltzoff. A. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.
[6] Gopnik, A. and Tenenbaum, J. (2007). Bayesian Networks, Bayesian Learning and Cognitive Development. *Developmental Science*, 10, 281–287.
[7] Machamer, P., Darden, L., and Craver, C. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67, 1–25.
[8] Margolis, E. and Laurence, S. (2011). Concepts. In E. Zalta, *Stanford Encyclopedia of Philosophy* (Spring 2014 Edition).
[9] Murphy, G. (2002). *The Big Book of Concepts*. Cambridge, MA: MITPress.
[10] Pearl. J. (2009). *Causality* (2nd. Ed.). Cambridge University Press, Cambridge.
[11] Weber, M. (2004). *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.

Conditionals as representative inferences

All approaches to the meaning of conditionals are driven by the idea that unless explicitly signaled, the truth or acceptability of a conditional relies on a dependence of the consequent of the antecedent. In probabilistic approaches this dependence is expressed in terms of the conditional probability of the consequent on the antecedent. Inferentialists [**3**] propose that we have to be able to infer the consequent from the antecedent. Recently we have seen many causal approaches to conditionals (cf. [**5**]). One shortcoming they share for their application to indicative conditionals, is that they have a hard time explaining what is wrong about a conditional like **(1)** *If it is sunny today, Real Madrid won the Champions League in 2017*, especially given that Real Madrid just won the Champions League yesterday.[1] There is general agreement as to what the problem is: there is no dependence in this case. Still, for these theories, the truth of the consequent is sufficient to warrant the truth/acceptability of the conditional.

In this talk we will build on insights from associative psychology and probabilistic analyses of causal learning to account for the dependence between antecedent and consequent. The analysis builds on the notion of *contingency*. The notion of contingency has been introduced in learning psychology to measure the learnability of a dependence between the two features. We will propose that for the appropriateness of a conditional, the conditional probability of the consequent on the antecedent has to be weighted by their contingency (**proposal**). This will be shown to solve the "missing link" problem discussed above (**application 1**). Furthermore,— drawn on experimental results on learning—, we will motivate an extension of contingency to what we will call 'representativenes' of y for x. This extension will allow us to account for biscuit conditionals (**application 2**) and conditional threats and promises (**application 3**). We will consider how far this proposal can serve as a general analysis of the meaning of conditional sentences. The talk will end with pointing out the close link the present analysis draws between conditionals and generics (**application 4**).

Within classical learning-by-conditioning psychology, learning a dependency between two events f and g is measured in terms of the contingency ΔP_g^f of one event on the other: $\Delta P_g^f = P(f|g) - P(f|\neg g)$, where P measures frequencies. Contingency not simply measures whether the probability of f given g is high, but whether it is high compared to the probability of f given all other (contextually relevant) cases than g ($\neg g$ abbreviates $\bigcup Alt(g)$). Thus, it is measured how representative or typical f is for g. [**6**] showed that rats learn a $\langle tone, shock\rangle$ association if the frequency of shocks immediately after the tone is higher than the frequency of shocks undergone otherwise, even if shocks occur only in, say, 12% of the trials in which a tone is present.[2]

Experiments in the aversive (i.e. fear) conditioning paradigms show that the speed of acquisition and the strength of the association in rats increases with the *intensity* of the shock. [**11**] show, similarly, that people build stronger associations related to events with high emotional impact. To capture this we introduce a new measure, the representativeness ∇P_g^f, defined as in (2), where $V(x)$ measures the absolute value (or intensity) of x.

$$(RES)\quad \nabla P_g^f = P(f|g) \times V(f|g) - P(f|\neg g) \times V(f|\neg g)$$

In the special case where the $Value$ only depends on f, as in the given example, representativeness comes down to $\Delta P_g^f \times V(f)$ and V measures the emotional intensity of f.

[1] Inferentialists claim that they can account for the observation, but we disagree.

[2] [**9**] shows that ΔP_G is the asymptotic value of the weight given to G when the learning task is modeled with a linear associator trained using the Rescorla-Wagner [**7**] learning rule, the most influential learning rule in associative learning and equivalent to the delta rule used in connectionists models.

Proposal. ΔP_g^f is a measure of the probabilistic dependence between f and g. To overcome the missing link problem of approaches to conditionals, one might therefore suggest to use ΔP_g^f or (RES) to check the assertability of a conditional sentence. However, [**10**] show that although $\Delta P_C^A > 0$ is a necessary condition for acceptability of indicative conditionals, it is not a sufficient one: it is also demanded that $P(C/A)$ is high. Therefore, we propose the following condition:

$$(CON) \quad A > C \text{ is assertable iff } \nabla P_A^C \times P(C/A) \text{ is high.}$$

Notice that if $Value$ is irrelevant, for acceptability it is a *necessary* condition that $\Delta P_A^C > 0$.

Application 1. Already contingency accounts for conditionals like (1), where $Value$ doesn't seem to count. If antecedent and consequent are probabilistically independent, we get $\Delta P_C^A = P(C|A) - P(C|\neg A) = 0$. If $Value$ doesn't count, $\Delta P_A^C = \nabla P_A^C$ and $\nabla P_A^C \times P(C/A) = 0$ as well. Hence, we predict that even in case $P(C) = 1$ and, therefore, $P(C|A) = 1$, the conditional is not assertable. But this doesn't hold for all conditionals with true consequents.

Application 2. An exception are **biscuit conditionals** like **(2)** *If you're hungry, there are biscuits on the table.* The biscuits are on the table, independently of whether you are hungry or not. Thus $\Delta P_A^C = P(C/A) - P(C/\neg A) = 0$. For these conditionals the $Value$ in the definition of representativeness ∇P_A^C matters. In (2) the value of the consequent is only high if the antecedent is true. Hence, $V(C|A) \gg V(C|\neg A) = 0$. As a result, ∇P_A^C will be high, and this explains the appropriateness of the conditional. A similar analysis will be proposed for **(2.ii)** *If you want to go to Harlem, you (must) take the A-train.*

Application 3. Our analysis works also for **conditional threats and promises** like **(3.i)** *If you won't give me your wallet, I will kill you* and **(3.ii)** *If you give me 10.000 euro, I will destroy the (for you hazardous) tapes.* What needs to be explained for such conditionals is that we accept them, although these threats and promises are not very credible (cf. [**8**]): $P(C|A)$ is typically not very high. In these conditionals the emotional impact of the consequent is independently of the antecedent. Thus, representativeness reduces to $\Delta P_A^C \times V(C)$. Notice that in **(3.i)** and **(3.ii)**, $P(C|A) - P(C|\neg A) > 0$, because the credibility of the consequent given $\neg A$ will certainly not be higher than given A. Given that in these cases $V(C)$ is extremely high for the hearer, it follows that ∇P_A^C will be high, even if $P(C|A)$ is not very high. Thus, these conditional threats/promises are accepted, as long as the stakes communicated in the consequent are high enough.

Application 4: Generics and conditionals are much alike. They both have at least the following purposes: (i) to state (inductive) *generalizations* ('Tigers are striped', 'If you push this button, the lamp will light'); (ii) to express (perhaps desired) *norms* ('Boys don't cry', 'If you see a general, you solute him'), and (iii) to express threatening cases like (iii) 'Pit bulls are dangerous dogs' and 'If you don't give me your wallet, I will kill you'. This suggests that they should be given very similar analyses. In this talk we will show that, indeed, a **uniform analysis** can be given to all types of examples, including both conditionals and generics (building on [**2**] and [**4**])!

Representative references: [**1**] Adams (1965), 'A logic of conditionals', *Inquiry*; [**2**] Cohen (1999), *Think generic!*; [**3**] Krzyzanowska et al (2013), 'Inferential Conditionals and Evidentiality', *JOLLI*; [**4**] Leslie (2008), 'Generics: Cognition and acquisition', *The Philosophical Review*; [**5**] Pearl (2000), *Causality: Models, Reasoning and Inference*; [**6**] Rescorla (1968), 'Probability of shock in the presence and absence of CS in fear conditioning', *Journal of Comparative and Physiological Psychology*; [**7**] Rescorla and Wagner (1972), 'A theory of Pavlovian conditioning' in *Classical Conditioning II*; [**8**] Schelling (1981), *The Strategy of Conflict*; [**9**] Shanks (1995), *The Psychology of Associative Learning*; [**10**] Skovgaart-Olsen et al. (2016), 'The relevance effect and conditionals', *Cognition*; [**11**] Slovic et al. (2004), 'Risk as analysis and risk as feelings', *Risk Analysis.*

The Role of Perceiver in the Predictive Processing Framework

As a perceiving agent, I enter a perceptual experience in two different ways: In perceiving myself in a mirror, I enter the perceptual experience as the *object* identified (based on visual information), and as the *subject* seeing the mirror and myself in it. This portrays a question that remains open within the predictive processing framework: what is the role of the perceiver, or how is the *self* represented?

In this presentation, I expound the answer attempted by Hohwy and Michael in a forthcoming paper. Their central idea is that through action, understood as active inference, we position our bodies and eventually ourselves in the realm of the hidden causes of sensory input. Given that the brain is in the business of inferring causes from input, it is (somehow) prompted to construct a self-model to capture the hidden causes that are generated through action. Hohwy and Michael rely on Metzinger's notion of a "self-model" but identify it primarily as a "full body model" which is a hierarchical structure of hidden, endogenous causes that will eventually be identified as the self.

At closer inspection, Hohwy and Michael's suggestion accounts for the subject-as-object only, which seems straightforward since in the object-sense I can be treated just like any other object in the world, in terms of a self-model that falls under the scope of prediction error minimization. The organism as object of experience is firmly embedded in the world and capable of interacting with it in various sensorimotor and cognitive ways. But the organism as subject of experience is not represented in the contents of experience. Thus, as far as it is represented, it is always only represented as object. This leaves open the question of how we could capture the two dimensions of the agent in the predictive coding framework? In this presentation, I explore several options to determine the place of the dimension of self-as-subject in the predictive coding framework:

I propose to take seriously the theoretical option of identifying the self with the organism and to analyze active inference in terms of the sensorimotor contingencies between action performed by the cognitive agent and changes in the immediate environment. This route may be the most promising story of how the predictive coding framework connects with the debate of perceptual experience.

Philosophy of Perception as Model-building

~

The problem of the contents of perceptual experience is one of the most debated issues in contemporary philosophy of perception. However, what is exactly the contribution made by theories of perceptual content in an age of science? On the one hand, some philosophers assume that theories of perceptual content will eventually lead to a better understanding of the underlying neural correlates (Siegel 2010). Yet, this assumption is often not supported by meta-theoretical considerations about the status of theories of perceptual content. On the other hand, some philosophers deny or downplay the role of philosophy of mind and advocate for a thoroughly scientific approach to the study of the mind (Chemero & Silberstein 2008). The research project is meant to articulate an answer to the presented issue. My claim is that the philosophical contributions to our understanding of visual perceptual content is better understood in terms of *phenomenological models*.

Recently, philosophers have showed a renewed interest for the role of models in science (Godfrey-Smith 2009; Weisberg 2013); whilst others argue for a model-building approach to metaphysics (Godfrey-Smith 2006; Paul 2012), and epistemology (Williamson *forth*.). Roughly speaking, a model is an *interpreted structure* that can be used to study more complex, real-world target phenomena (Godfrey-Smith 2006). In contrast with these works, my proposal has an element of originality. Whereas Weisberg and others have investigated the role of models in science, and Godfrey-Smith and others have put forward a meta-theoretical interpretation of the role of metaphysics, my claim is that philosophical modeling can be used within the sciences of the mind, and thus represent a contribution to the larger project of explanatory integration in cognitive science (Danks 2014; Miłkowski *forth*.).

Scientific models can be concrete, mathematical, or computational (Weisberg 2013). Philosophical models have a *propositional character* that can either take a verbal or symbolic form, e.g. when characterizing the phenomenon by means of logical symbolism. In this sense, Russellian or Fregean theories of content are models that seek to describe the phenomenon of visual content. Such models may also be mutually inconsistent, as they provide different characterizations of the target phenomenon. However, they all lack explanatory power. This comes as no surprise, as philosophical models do *not* provide scientific explanations. Instead, these models are mere descriptions of the observable properties and their clustering of a target phenomenon; they do not provide a scientific explanation about why or in virtue of what the phenomenon occurs. In this sense, they are *phenomenological models* (Craver 2006; Frigg & Hartmann 2009; Hochstein 2013).

Although phenomenological models are not explanatory, there is a widespread consensus that they nonetheless play an important role in science (Batterman 2002; Bogen 2005; Craver 2006; Hochstein 2013). I argue that philosophical theories of perceptual content—or *MoPs* (from 'Models of Perceptual content')—provide a *preliminary decomposition* of the target phenomenon. In this sense, philosophical theories of perceptual content may represent a complement to the methodology of data-analysis in psychology (Haig 2014). Segmenting different, but closely interwoven phenomena, more specifically, can be achieved by means of the following sub-tasks: *a*. the identification of similarities among contents (color properties, form properties, visual objects, etc.); *b*. in identifying regularities in perceptual content (e.g. color and form properties as 'dependent-properties', Stumpf 1873; Gestalt laws, etc.); *c*. constraining possible theories of the underlying perceptual mechanisms (Bechtel & Richardson 2010). Since psychological phenomena are conceived as the explananda of scientific theories (Bogen & Woodward 1988; Feest *forth*.), the identification or at least a preliminary decomposition of the phenomena of visual perceptual experience can represent can pave the way to a fruitful interplay between thoroughly scientific and philosophical theories.

MoPs must conform a criterion of phenomenal adequacy, i.e. they must provide an adequate characterization of the phenomenon that they are about. As our scientific understanding of the phenomenon improves, MoPs may be revised or rejected. Scientific discoveries can show the inconsistency or implausibility of some models, thereby confirming some other models, or opening the logical space for new models. MoPs can also be put within a hierarchy or family—for example, the family of structural propositional MoPs—, and accordingly, some scientific achievements might, more realistically, lead to adjust or reject some specific models that belong to a broader family.

The proposed interpretation does not aim at exhaustiveness in a twofold respect: firstly, philosophy of mind and perception is not exhaustively described by MoPs, some philosophical questions may eschew the model-building purpose; secondly, the positive role of MoPs in scientific research may actually turn out to be far richer than described above. The research project is meant as a contribution not only to the self-understanding of philosophical works in the philosophy of mind and cognitive science, but also strives to make philosophy "count" on the face of scientific advancements in our study of the mind.

Short bibliography:

— Batterman (2002). "Asymptotics and the Role of Minimal Models" *British Journal for the Philosophy of Science* 51.
— Bechtel & Richardson (2010). *Discovering Complexity*. MIT Press.
— Bogen (2005). "Regularity and Causality: Generalizations and Causal Explanations." *Studies in the Histiory and Philsophy of Biology and Biomedical Science* 36.
— Bogen & Woodward (1988) Saving the Phenomena. *The Philosophical Review* 97.
— Chemero & Silberstein (2008). "After the Philosophy of Mind" *Philosophy of Science* 80.
— Craver (2006). "When Mechanistic Models Explain." *Synthese*153.
— Danks (2014). *Unifying the Mind*. MIT Press.
— Feest (forth.). "Phenomena and Objects of Research in the Cognitive and Behavioral Sciences" *Philosophy of Science*.
— Frigg & Hartmann (2009). "Models in Science" in *Stanford Encyclopedia of Philosophy*.
— Godfrey-Smith (2006). "Theories and Models in Metaphysics." *Harvard Review of Philosophy* 14.
— Godfrey-Smith (2009). "Models and Fictions in Science." *Philosophical Studies* 143.
— Hochstein (2013). "Intentional Models as Essential Scientific Tools." *International Studies in the Philosophy of Science* 27.
— Miłkowski (forth). "Unification Strategies in Cognitive Science."
— Paul (2012). "Metaphysics as modeling: the handmaiden's tale" *Philosophical Studies* 160.
— Siegel (2010). *The Contents of Visual Experience*. Oxford University Press.
— Stumpf (1873). *Über den psychologischen Ursprung der Raumvorstellung*. Leipzig: Hirzel.
— Weisberg (2013). *Simulation and Similarity*. MIT Press.
— Williamson (forth.) "Model-Building in Philosophy." In Blackford & Broderick (Eds.) *Philosophy's Future.* Oxford: Wiley.

Referential Intuitions are Still Problematic

The seminal paper "Semantics, Cross-Cultural Style" by Machery, Mallon, Nichols and Stich (2004) – MMNS – has provoked a large and intense debate among philosophers about the role of referential intuitions in theory of reference. Martí (2009) put forward one of the most radical objections to MMNS's experiment. Martí argued that the kind of metalinguistic intuitions tested by MMNS, i.e. referential intuitions, is evidentially irrelevant for individuating the correct theory of reference. MMNS's experiment prompts participants to think about what a fictional speaker refers to with a proper name in a hypothetical scenario. However, philosophers of language are not interested in what people think about how proper names refer. Rather, they want to know how proper names refer.

In order to respond to Martí's objection and to maintain that referential intuitions are a reliable source of evidence for the theory of reference, Machery *et al*. (2009) – MOD – conducted an empirical research by testing truth-value judgments. MOD's experiment aimed at showing that people's referential intuitions are congruent with the way they use proper names. We will discuss a conceptual limitation of MOD's experiment on truth-value judgments. Then, we will present the data of an empirical survey that shows that people's truth-value judgments are not congruent with their use of proper names. We will explain why the results of our empirical research refute the conclusions of MOD's experiment on truth-value judgments. Truth-value judgments do not prove what MOD take them to prove. We will conclude that referential intuitions are still problematic.

MOD presented a hypothetical scenario in which Ivy, a fictional character, has been told that Tsu Ch'ung Chih was the man who first determined the precise time of the summer and winter solstices and this is the only thing Ivy has heard about Tsu Ch'ung Chih. However, Tsu Ch'ung Chih did not really make this discovery. He stole it from Ye Chan, an astronomer who died soon after making the discovery. The theft remained entirely undetected and Tsu Ch'ung Chih became famous for the discovery of the precise times of the solstices. MOD asked the participants to tell whether they think that Ivy's claim, when she says "Tsu Ch'ung Chih was a great astronomer", is true or false.

Let X be the man that Ivy was told to have discovered the solstice times and Y the astronomer who really made the discovery. Since the participants know from the vignette that Y and not X made the discovery of the solstice times, the participants, who respond "true" to MOD's question task, understand Ivy to use the name "Tsu Ch'ung Chih" to refer to Y and not to X. MOD claim that asking people to assess the truth-value of a sentence relates, though it is not the same thing, to their willingness to assert it. MOD claim that typically if one says that someone else's utterance is true, one is disposed to assert it oneself. Thus, in MOD's view, if a participant answers "true", this is evidence that the participant is disposed to assert "Tsu Ch'ung Chih was a great astronomer". Given that the participant learns from the vignette that Y is the man who discovered the solstice times, this is evidence that the participant uses the name "Tsu Ch'ung Chih" to refer to Y. Since Y is the satisfier of the description "the man who discovered the solstice time", this is evidence that the participant uses the name "Tsu Ch'ung Chih" according to the convention that its reference is determined by that description.

MOD found out that about a third of participants chose the answer "true". This result confirmed a substantial within-culture variation in truth-value judgments. MOD concluded that people's referential intuitions are congruent with the way people use proper names and, therefore, referential intuitions are a reliable source of evidence in theory of reference.

We will argue that the methodology of MOD's experiment is flawed. We agree that the participants, who answer "true", understand Ivy to refer to Y with the name "Tsu Ch'ung Chih" and are disposed to assert the proposition that Y is a great astronomer. The point is that these data give us no clue on whether the participants are disposed to use the name "Tsu Ch'ung Chih" to refer semantically to Y. The mere fact that a participant understands Ivy to have expressed a proposition about Y is not evidence that the participant is disposed to use the proper name "Tsu Ch'ung Chih" according to the semantic convention that the reference of "Tsu Ch'ung Chih" is determined by the description "the person who discovered the solstice times".

We will explain the conceptual design and present the data of an experiment in which a truth-value judgment task similar to the one used in MOD's experiment is followed first by a comprehension task and then by a forced-choice task.

TASK 1 (truth-value judgment)
when Ivy says: "Tsu Ch'ung Chih was a great astronomer", do you think that her claim is:
(A) True
(B) False
TASK 2 (comprehension)
If Ivy says to you: "Write Tsu Ch'ung Chih's biography", whose would you write the biography?
(i) The person who really discovered the solstice times
(ii) The person who stole the discovery
TASK 3 (forced-choice)
Suppose you wrote the biography and you sent it to the Editor for printing. Which title would you choose for the biography?
(iii) "The biography of Tsu Ch'ung Chih"
(iv) "The biography of Ye Chan"

We used the task 1 in order to reproduce MOD's experiment. The data we collected from task 1 replicate the data of MOD's experiment: 34,95% of participants chose the answer "true" and 65,05% chose the answer "false". In order to investigate whether MOD are right that the answer "true" to the task 1 means that the participant is disposed to use the proper name "Tsu Ch'ung Chih" according to the convention that its reference is the satisfier of the description "the man who discovered the solstice times", we added the task 2 and the task 3. The rationale behind the task 2 and the task 3 is the following. The task 2 drives the participant to hypothesize a concrete and not convoluted linguistic exchange with Ivy and tests how the participant understands Ivy's use of the name "Tsu Ch'ung Chih". The purpose of the task 3 is to check how the participant is disposed to use the name "Tsu Ch'ung Chih" independently from the hypothetical linguistic interaction with Ivy. We take it to be a safe assumption that the choice of the title of the biography results from the coincidence between (a) the participant's general communicative intention of using the proper name that occurs in the title with its conventional meaning and (b) the participant's specific communicative intention of referring to the man who is the subject of the biography. Therefore if, for instance, a participant chooses the title "The biography of Tsu Ch'ung Chih", this is evidence that the participant uses the name "Tsu Ch'ung Chih" to refer semantically to the man who is the subject of the biography.

If MOD were right that the answer "true" to task 1 is evidence that the participant is disposed to use the name "Tsu Ch'ung Chih" according to the convention that its reference is the satisfier of the description "the man who discovered the solstice times", then we would expect the answer "true" to task 1 to be mainly followed by the combination of responses (i)/(iii). In fact, this would be evidence not only that the participant understands Ivy to refer to the man who discovered the solstice times with the name "Tsu Ch'ung Chih", but also that the participant is disposed to use that name to refer semantically to him. The data collected in our experiment falsify this prediction. We found out that the answer "true" is mainly not followed by the combination of responses (i)/(iii).

References

Machery, E., Mallon, R., Nichols, S. and Stich, S. P. 2004. Semantics, Cross-cultural Style. *Cognition* 92: B1–B12.
Machery, E., Olivola, C. Y. and de Blanc, M. 2009. Linguistic and Metalinguistic Intuitions in the Philosophy of Language. *Analysis* 69: 689–694.
Martí, G. 2009. Against Semantic Multi-culturalism. *Analysis* 69: 42–48.

De Jure Co-Reference: In Defence of Pointer Relations

This paper offers an account of same-thinking. Same-thinking occurs whenever two thoughts concern the same referent. There are two different ways in which this can occur. First, there are cases in which co-reference is manifest to the subject. In such cases the thinker is warranted in trading on identity and thereby inferentially exploit the sameness of reference (c.f. Campbell 1987, 1994, 2002). Two thoughts that are related in this way are *de jure* co-referential. Second, there are cases in which two thoughts concern the same referent, but where the sameness of reference is not manifest to the subject. In such cases the subject is not warranted in trading on identity. Such thoughts are *de facto* co-referential. This means that two pairs of thoughts that are referentially equivalent may nonetheless play different roles in cognition, depending on whether they are *de jure* or *de facto* co-referential. The question of how to account for the difference between *de jure* and *de facto* co-reference has been at the center of many debates within philosophy of language and mind since Frege (1892). The account of same-thinking I offer in this paper claims that the difference between *de jure* and *de facto* co-reference is to be understood in terms of a particular kind of irreducible relation – *the pointer relation* – that holds at the level of representational vehicles.

We may distinguish between two classes of views about the difference between *de jure* and *de facto* co-reference: On the one hand, there are those who hold that the difference is to be accounted for in terms of intrinsic representational properties, i.e. properties that do not concern relations to other representations. I will call such views *intrinsicalist* views:

> **Intrinsicalism:** The difference between *de jure* co-reference and *de facto* co-reference is to be accounted for in terms of intrinsic representational properties.

On the other hand, there are those who hold that same-thinking cannot be accounted for purely in terms of intrinsic representational features (c.f. Fine 2007, Heck 2012, Pinillos 2011). According to such views, the difference between *de jure* and *de facto* co-reference is essentially a matter of relational representational features. Whether or not two thoughts are *de jure* co-referential depends on how the thoughts are related, and such relations cannot be reduced to sameness or difference of intrinsic representational features. Such views are often labelled *relationist* views:

> **Relationism:** The difference between *de jure* co-reference and *de facto* co-reference is to be accounted for in terms of irreducible relational representational features.

I will motivate a version of relationism by criticising a particular strand of intrinsicalist views, according to which same-thinking is to be accounted for in terms of co-location of information within mental files (c.f. Perry 1980, Recanati 2012). I argue that if we want to give a unified account of same-thinking, such views collapse into relationism. I take this observation as an outset and suggest a new account of same-thinking.

According to the mental file theorist, two pieces of information are *de jure* co-referential iff they are stored within the same mental file. If two pieces of information regard the same individual, but are stored in distinct files, the two pieces of information are *de facto* co-referential. I argue that the mental file theorist faces problems in accounting for *de jure* co-reference of relational predicates in this way. The case of relational predicates cannot be accounted for purely in terms of co-location of information within a single file, since relational predicates involve two (or more) distinct files. I consider two strategies available to the mental file theorist:

> **The Duplication Strategy:** The relational predicates are duplicated into all the relevant files.

Appealing to Pointer Relations: Introducing the notion of a *pointer* that relates information in one file to information in the other.

I argue that the first option fails: The duplication strategy does not add anything of explanatory interest to the original picture. I then show that the remaining option for the mental file theorist is to postulate a further theoretical notion – the notion of a *pointer relation* – in addition to the notion of a file. Pointer relations account for how *de jure* co-reference may obtain between pieces of information across distinct files. I then argue that the explanation of *de jure* co-reference in terms of pointer relations can be generalized to cases of monadic predicates. Finally I argue that it is the notion of a pointer rather than the notion of a file that is of explanatory interests in the case of same-thinking. As a result we should abandon talk of mental files in favour of unified account of *de jure* co-reference in terms of pointer relations. The suggested view is a version of relationism. The pointer relations are what account for the difference between *de jure* and *de facto* co-reference and such pointer relations cannot be reduced to sameness of intrinsic representational features.

The positive view has several virtues. Pointer relations hold at the level of representational vehicles rather than mental content, and appealing to pointers thus allows us to account for the Fregean data (c.f. Frege 1892) within a Millian/Russellian framework. The notion of a pointer is a minimal addition to the classical Millian/Russellian picture, but it is explanatorily powerful. Appealing to pointer relations gives us a straight-forward account of how *de jure* co-reference can fail to be transitive (c.f. Pinillos 2011). Since pointer relations do not hold in virtue of intrinsic representational features, there is no reason to think that *de jure* co-reference is limited to cases in which the intrinsic representational features are the same. The pointer picture thus provides a unified account of same-thinking without having to invoke a distinction between weak and strong *de jure* co-reference (c.f. Recanati 2016). I conclude that the suggested framework is preferable to the mental file framework since it allows us to give a unified account of same-thinking, and also that it is preferable to other versions of relationism (such as Fine's (2007) *semantic relationism*) in that it is compatible with classical (i.e. non-relational) Millian/Russellianism.

References:

Campbell, J. 1987. Is Sense Transparent? *Proceedings of the Aristotelian Society* 88: 273—92.

_____. 1994. *Past, Space and Self.* Cambridge, Mass.: MIT Press.

_____. 2002. *Reference and Consciousness,* Oxford: Oxford University Press.

Fine, K. 2007. *Semantic Relationism.* Oxford: Blackwell.

Frege, G. 1892. "On Sense and Reference", in P. Geach & M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege* (1960), Oxford: Blackwell.

Heck, R. G., Jr. (2012), Solving Frege's puzzle, The Journal of Philosophy, 109, 132—174.

Perry, J. 1980. A Problem about Continued Belief. *Pacific Philosophical Quarterly* 61: 317—32.

Pinillos, A. 2011. Coreference and Meaning. *Philosophical Studies* 154: 301—24.

Recanati, F. 2012. . *Mental Files.* Oxford: Oxford University Press.

_____. 2016. *Mental Files in Flux.* Oxford: Oxford University Press.

The Development of Concepts of Mind and Mental States

An important issue underlying the question of how and when children come to have a concept of mind is what it even means to have a concept of this. Presupposing Peacocke's (1992) epistemic theory of concepts, I argue that we need to distinguish between having a concept *that* other people have minds (concept-that) and having concepts of *what* kinds of mental states they have (concept-what). Moreover, it is important to note that these two issues can come apart to the extent that it could be possible to understand that other people have minds while not understanding that they can have particular kinds of mental states (e.g. beliefs). This is especially apparent from the empirical literature on the false belief task, where children's failure in the false belief task before the age of 4 is often taken to suggest that the children fail to understand that other people have (false) beliefs (lacking concept-what), but not that they fail to understand that other people have minds (not lacking concept-that). Conversely, sensitivity to animacy may be a criterion for having a concept of mind (concept-that), while not indicating concepts of particular mental states (concept-what).

This distinction between the concept-that and concept-what question is crucial for examining the empirical literature as well as evaluating theories of how we come to understand other minds. Posing the two distinct questions presents us with a framework in which to address the recent debate as to whether Interactionist Theories (IT) can be seen as a genuine alternative theory for how we come to understand other minds to the traditional Theory Theory (TT) and Simulation Theory (ST) accounts. IT criticises TT's and ST's focus on third person observation and the assumption that minds are hidden. Instead, IT argues for a direct experience of other minds and emphasises the importance of interaction. I argue that the crucial contribution of IT is with regards to the concept-that question, that is to say how we come to think of others as minded in the first place, which both ST and TT struggle to account for (see also Reddy, 2008). It is through interaction that we experience others as people and not as objects and it is in virtue of this that we come to think of them as minded beings. When it comes to the concept-what question, which is what is at issue in the false belief task literature, we must go beyond mere interaction. Comparing evidence of normally developing children and children with autism suggests, however, that the earlier understanding of concept-that through interaction facilitates the development of concept-what. The task, therefore, is to provide an account of how an initial interaction based understanding that other people have minds develops into the more complex appreciation of other people's mental states, in particular beliefs.